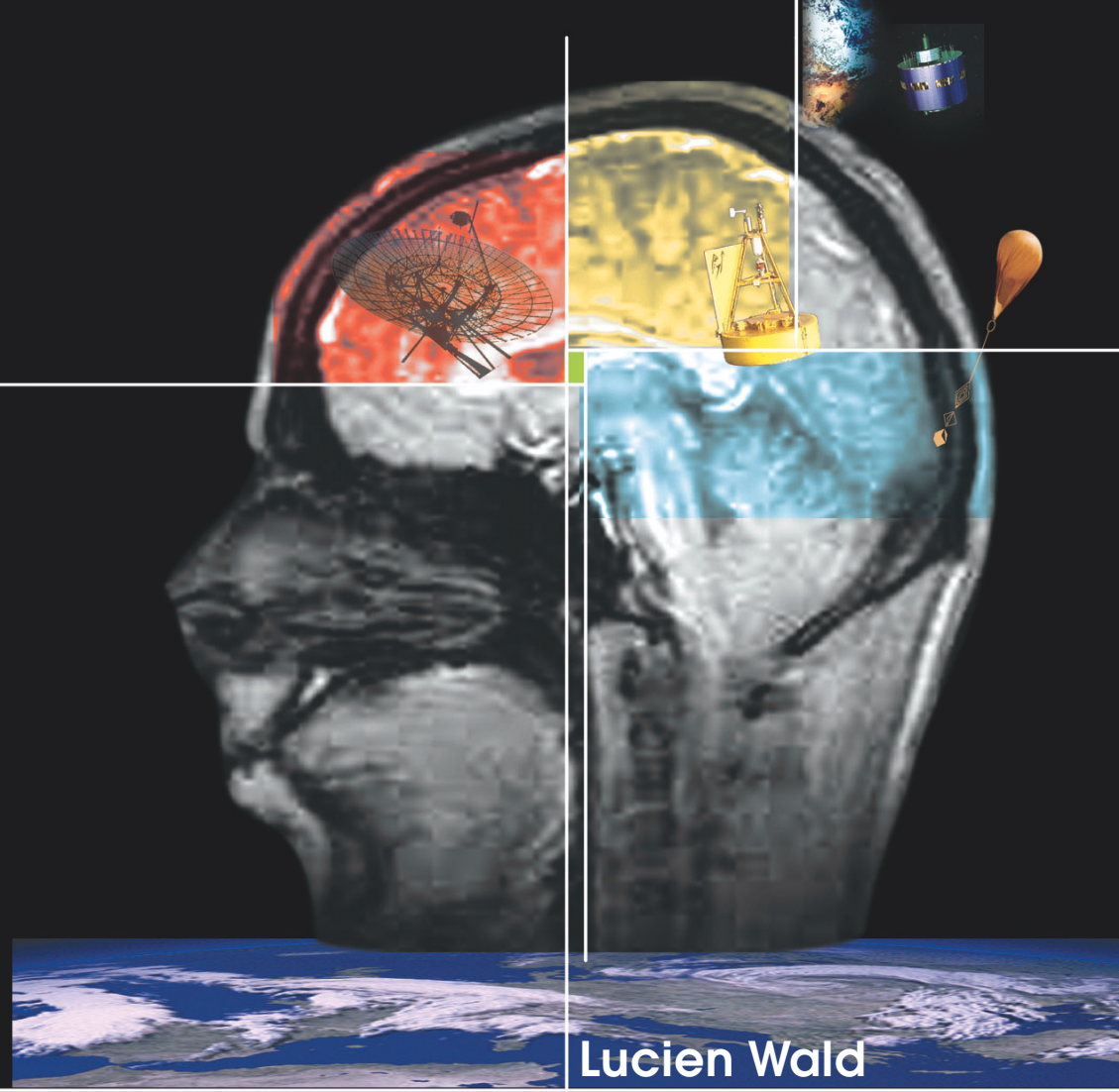
Data Fusion. Definitions and Architectures - Fusion of Images of Different Spatial Resolutions

Lucien Wald



Lucien Wald

DATA FUSION

DEFINITIONS AND ARCHITECTURES

Fusion of images of different spatial resolutions

ECOLE DES MINES DE PARIS
Les Presses

DATA FUSION

DEFINITIONS AND ARCHITECTURES

FUSION OF IMAGES OF DIFFERENT SPATIAL RESOLUTIONS

LUCIEN WALD
Professor, École des Mines de Paris

Les Presses de l'École des Mines
Paris, 2002

60 boulevard Saint Michel, 75006 PARIS, FRANCE

http://www.ensmp.fr/Presses

catherine.delamare@ensmp.fr

ISBN: 2 - 911762 – 38 – X

Legal deposit: 2002

Printed in 2002 (Grou Radenez, Paris)

Table of Contents

ACKNOWLEDGMENTS

This work has been made thanks to fruitful discussions with a large number of researchers and engineers, and with the many participants to the EARSeL - SEE working group "data fusion". I am particularly indebted to Thierry Ranchin; together, we realized several studies forming the core of the second part of the book. I also thank Michel Albuisson, Manfred Buchroithner, Luce Castagnas, Isabelle Couloigner, Alfonso Farina, Douglas J. Kewley, Marc Mangolini, Louis-François Pau, Stelios Thomopoulos and Robin Vaughan for their comments and assistance.

FOREWORD – *AVANT-PROPOS*

The first objective of this book is to clarify the concept of data fusion, or information fusion. By this document, the author hopes that data fusion will be better understood, accepted and used more efficiently. Presently, for most of its users, conscious or not, data fusion is more an ensemble of techniques and methods than a formal framework.

The second and last objective of the book is the detailed description of techniques in image fusion, without pretending to completeness. The techniques are dealing mostly with the fusion of measurements with the pixel as a support of information. The synthesis of images of various modalities at the best spatial resolution available in the original sets of images is of major concern in this second part of the book. The assessment of the quality of the fused product is also an important topic.

- o0o -

Ce livre a pour premier objet de clarifier le concept de la fusion de données, ou fusion d'informations, et, par conséquent, de mieux le faire comprendre et accepter. Actuellement, pour la plupart de ses utilisateurs, conscients ou non, la fusion de données représente plus un ensemble de techniques et méthodes qu'un cadre formel.

Le deuxième objet du livre est la présentation détaillée et pratique des techniques de fusion d'images, sans pour autant prétendre à l'exhaustivité. Les techniques présentées sont essentiellement limitées à la fusion de mesures, en utilisant le pixel comme support d'information. La synthèse d'images multi-modales à la meilleure résolution spatiale disponible dans le jeu d'images originales, ainsi que l'évaluation de la qualité des produits de fusion occupent une part très importante de cette deuxième partie.

1. INTRODUCTION

Data fusion is a recent term. It means an approach to information extraction spontaneously adopted in several domains before this was expressed as "data fusion". This approach is based upon the exploitation of the synergy offered by the information originating from various sources. Here, data is a generic term and is equivalent to information. Combination of additional independent and/or redundant data usually results into an improvement of the results. The example of human vision is often given to illustrate the advantages and benefits of data fusion. The two eyes of a man have slightly different viewing angle, making possible stereo vision and depth perception. Hence having two eyes extends the capability of a single eye. Another advantage of having two eyes is redundancy; if one is disabled, vision is still possible, though in a degraded mode.

Data fusion research and development was conducted under a wide variety of systems, methods and names. Using recent words such as "data fusion", or "information fusion" translates the recent understanding that whatever the application domain, these synergistic approaches share common problems and common properties. Let take a very simple example. Actually, an addition is a fusion process. It may appear curious to claim that a formal framework is really needed for such a simple operation. However those who have taken high-level classes in mathematics know how much theory is behind the addition of two numbers. The others know quite well that addition can only be performed on quantities that belong to the same space. We all know that we cannot add US dollars and euros without converting them to a common currency. Physicists know that they cannot add temperatures of objects, but they can do with heat quantities. Statisticians know that the standard deviations do not add, while variances do. These are simple examples, which, though not illustrating the complexity of data fusion, show that these problems share similar concerns, which are named under a single name in data fusion: the alignment. This property is part of the data fusion framework, together with many other elements dealing with methods, architectures, system design, etc.

These common problems and common properties form a paradigm. Research in data fusion aims at exploring this paradigm. It expresses and clarifies the concept of data fusion and its properties. Definitions and terms of reference can be established that permits better co-operation between various domains because they share a common language. Research reveals the fundamentals in data fusion with respect to the fundamentals of the related sciences, e.g., physics, mathematics... It also expresses the properties of the data / information to be fused, of the methods for fusion, of the

architectures, thus permitting better design, implementation and analysis of fusion processes. It is then easier to develop the most appropriate methods and algorithms, to monitor the quality throughout a process etc.

There are many advantages in using data fusion[1]:

- robustness and reliability. The system is operational even if one or several sources of information are missing or malfunctioning,
- extended coverage in space and time,
- increased dimensionality of the data space. It increase the quality of the deduced information; it also reduces the vulnerability of the system,
- reduced ambiguity. More complete information provides better discrimination between available hypotheses,
- providing a solution to the explosion of the information that is available today.

Data fusion is exploited by a large number of biological systems. An illustration is given by the human system, which calls upon its different senses to perceive its environment (Fig. 1.1).

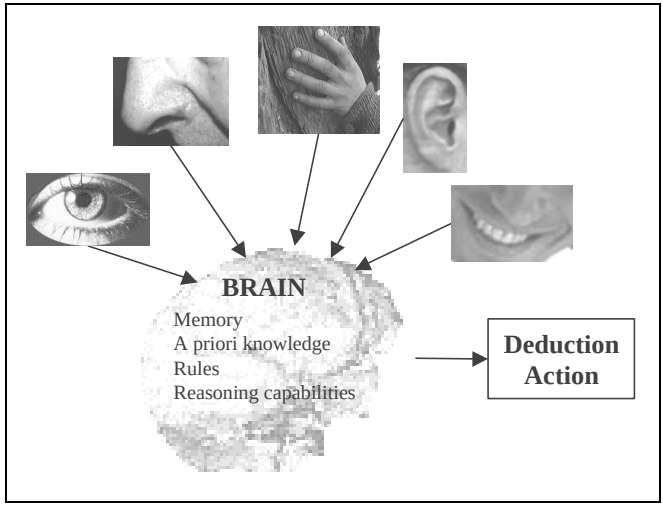


Figure 1.1. The human brain and perception system as an example of fusion process

[1] E. Waltz and J. Llinas. *Multisensor data fusion*. Artech House, 1990.

Human sensors acquire information on sight, smell, touch, hearing, and taste. The acquired data are processed within the brain. To do so, the brain will use other sources of information: its memory, its experience and its *a priori* knowledge. Calling upon its reasoning capabilities, the brain "fuses" all this available information to perform deductions, to produce a representation of the environment and to order action.

This example also illustrates how much data fusion is at the crossings of several scientific domains. Here neuroscience, sciences of cognition, and medicine are at stake.

Data fusion is not limited to biology. It originates in Defense activities, and such applications are still very vivid. Almost half of the scientific literature is devoted to defense systems[2]. Nevertheless, fusion applies to many other domains. Examples are numerous in transportation, and especially in civil aviation (aids in aircraft, air traffic control, landing aids) and motorways management. Large research efforts are devoted to intelligent car traffic, where each car embark a set of sensors and fusion capabilities, in order to best co-operate with other vehicles and the environment itself. Navigation / positioning is a service routinely offered today. An efficient service calls upon the fusion (often called hybridization in this domain) of several sources: fleets of orbiting satellites and ground systems. Telephone is another example, where several resources must be used through complex fusion systems to make a phone call: transponders aboard geostationnary or low Earth orbiting satellites and terrestrial networks. Robotics calls upon data fusion for 3-D vision and displacement in hostile environment, monitoring, inspection and maintenance of pieces of equipment.

The exploitation of satellite images and more generally of observations of the Earth and our environment is presently one of the most productive in data fusion. Observation of the Earth is performed by means of satellites, planes, ships, and ground-based instruments. It results into a great variety of measurements, partly redundant, partly complementary. There are very few domains, where such a diversity is present and this makes Earth observation so fascinating. The availability of so many types of information constitutes a tremendous field of investigation for mathematicians. This interest is enhanced by the challenge of correctly modeling natural landscapes and outdoor scenes, which are usually more difficult than indoor scenes. The research in this field is backed up by the present political interest in environment and global changes.

---

[2] L. Valet, G. Mauris and Ph. Bolon. *A statistical overview of recent literature in information fusion*. In Proceedings 3rd International Conference Fusion 2000, Paris, July 2000, pp. MOC3-22-29, IEEE catalog 00EX438, ISBN-2-7257-00001-9, 2000.

These measurements in Earth observation may be punctual and time-integrated, bi-dimensional and instantaneous (images), vertical profiles with time-integration or not, three-dimensional information (oceanic / atmospheric profiler / sounder at ground level, or satellite-borne, or ship-borne). Adding the large amount of archives and numerical models representing the geophysical / biological processes, one should conclude that the quantity of information available to describe and model the Earth and our environment increases rapidly. Data fusion is a subject becoming increasingly relevant because it efficiently helps scientists to extract increasingly precise and relevant knowledge from the available information.

The set of sensors for Earth observation is extremely various. The spectrum of their characteristics is very large, with respect to spatial and temporal scales, spatial and temporal sampling and means of acquisition. Such diversity is a tremendous source of practical problems, whose resolutions lie upon a good understanding and modeling of more fundamental questions. For example, what are the links between temperature measurements made at ground level using a thermometer and integrated over an hour, and the instantaneous measurements of the same temperature but made using a satellite-borne radiometer sensing the radiation emitted by a surface of several square kilometers? Data fusion is here at the crossings of the physics of the measurements, Earth sciences and sciences of information and communication. These crossings offer many opportunities and benefits to the progresses in data fusion.

Weather forecasting fully illustrates data fusion in environment (Fig. 1.2). It is one of the most sophisticated fusion systems nowadays and is performed several times a day for the whole planet. It calls upon sensors, signal processing, artificial intelligence and complex modeling of physics and chemistry and the atmosphere, oceans and land. There are processing issues, topological issues (the distribution of sensors in 3-D space and time) and communication challenges.

Meteorological satellites are orbiting the Earth, in a geostationnary orbit or in a near-polar one. They are equipped with sensors providing sets of measurements on the 3-D properties of the atmosphere and on the characteristics of the surface of the ground and the ocean. Balloons and planes operate at lower altitudes. Tens of thousands of ground stations are distributed in the world. They measure the basic weather parameters, such as air temperature and pressure, wind, cloudiness, rainfall, and more for some of them. Ground radars follow storms and rain cells. At sea, ships and automated buoys provide weather parameters and measurements of the sea surface, such as temperature and waves.

All these measurements are processed to extract geophysical parameters of interest, and transmitted by means of specialized communication networks.

Then in numerical weather prediction centers, numerical models through data assimilation techniques ingest this wealth of information, together with weather prediction of the previous instants. They produce weather forecast that are used by professionals and are also presented on TV news and other media.

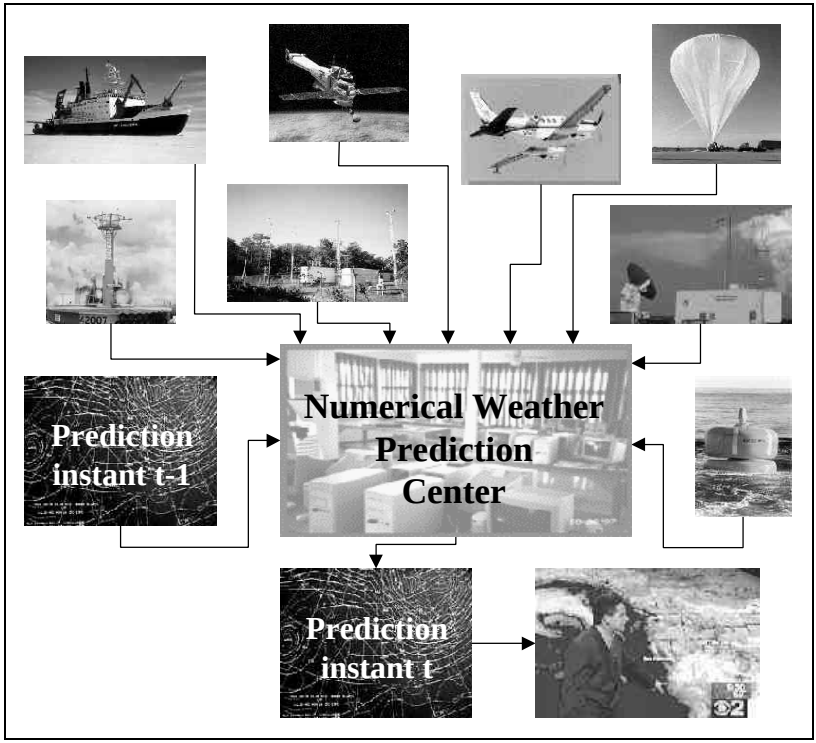


Figure 1.2. Weather forecasting is an excellent example of a fusion system in environment

The operation of data fusion by itself is not new in environment or in any domain of application. For example, meteorologists predict weather for several tens of years. In remote sensing (i.e. Earth observation from spacecraft or aircraft), classification procedures are performed since long and are obviously relevant to data fusion.

Data fusion allows formalizing the combination of these measurements, as well as to monitor the quality of information in the course of the fusion process. The formal framework for data fusion provides a better understanding of data fusion fundamentals and of its properties. Once established, such a framework permits a better description and formalization

of the potentials of synergy between the available sources of information, and accordingly, a better exploitation of the data.

This book intends to foster the understanding of data fusion by a wider community, for which data fusion denotes presently more a set of techniques than a framework for these techniques and more. Its content originates from lectures given to students of master degree level or higher, engineers and researchers in information, or computer sciences or environment sciences. This book should be valuable to any engineer, scientist and practitioner interested in fundamentals in data fusion.

It does not reproduce or mimic the well-known books written on data fusion several years ago. These books are full of methods and other technological considerations on systems, architectures, communications etc. Though some of the latter are obsolete nowadays, these books are a good baseline to apprehend more sophisticated methods and technologies. However, they offer little space to the formal framework of data fusion and to the specifics of fusion of images. One of the scopes of the present book is to fill the gap in both aspects, though not pretending to be exhaustive.

The first part of this book presents the concept of data fusion. Several definitions and terms of reference are analyzed in Chapter 3. The properties of the data to be fused are presented. A typology of the class of problems in data fusion is discussed. This book seeks to clearly establish the fundamentals and help setting the foundations for subsequent growth. Such a discussion on fundamentals is seldom found in the literature. This part of the book should fill the gap, especially with regards to education and training. This part is fully general and applies to all fields. It can be read with profit by anybody interested in data fusion, whatever his domain of expertise. Chapter 4 comprises a discussion on the representation of a fusion process and on architectures.

The second part of the book is devoted to techniques and procedures for the fusion of images and assessment of the quality of the resulting products. It intends to be problem solving. It offers an in-depth presentation of standard and advanced methods for the fusion of multi-modality images, which is of interest to the full spectrum of the community dealing with imagery. The emphasis is on images having different spatial resolutions, though the book is not limited to this case. Given several sets of images acquired by disparate sensors, the problems treated in this book are to create new sets of images of reduced dimensionality, in order to either better visualize the original sets of images as a comprehensive ensemble of information, or to synthesize images with a better spatial resolution.

Imaging sensors are more and more present today: medicine, industrial processes, traffic management, verification of treaties, crisis management, ... Even geographical data may be considered to a certain extent as acquired

by a sensor when they are digitized (rasterized maps). In some way, they may be assimilated to images. This is the case of many digitized maps, available on the Worl Wide Web, and asembled for the studies and analyses of global change. These maps are made of measurements, e.g., air temperature, or categorical data, e.g., type of vegetation.

Plenty of techniques for data fusion already exist; Hall[3] described many of them. A few of them apply to images and imaging sensors. Given the importance of the images and the liveliness of the technical developments in this domain, the second part of the book specializes in the fusion of images. Nowadays, satellite images are intensively used in various applications, especially since their spatial resolution reaches 1 meter or better, thus offering a great deal of details. Many examples given in this part deal with such images.

Several mathematical tools are presented in Chapter 5; they form the bases of many popular or advanced techniques for the fusion of images. Advices and details are given for the practical implementation of these tools. Then these tools are used for image fusion.

Chapter 6 focuses on data fusion as a means for a better visual analysis of several sets of images. The very popular intensity-hue-saturation technique is extensively discussed. Together with the principal components analysis technique, they are appropriate to visualize images acquired by multiple sensors, disparate or not, commensurate or not. Another technique is presented, which treats non-commensurate sensors by the means of a fusion of representations. A composite scene is created by encrustation to display most of the information of interest.

A sub-domain of image fusion is explored in great details in Chapter 7; it deals with the synthesis of images having different spatial resolutions at the best resolution available within the sets of images. These synthesized images are close to what should be observed by the corresponding sensor, if existent. Several observing systems acquire images, B_{il}, in different modalities or spectral bands i (or wavelength ranges) with a spatial resolution l, and images A_{jh} in bands j, with a better spatial resolution h. An example in Earth observation is the Ikonos system, which acquires one image in four bands: red, green, blue and near-infrared with a spatial resolution of 4 m, and at the same time, one panchromatic image with a resolution of 1 m. Synthesizing these multispectral images at a resolution of 1 m permits a better mapping of our environment, and especially of cities. Another example is given by the industrial systems that acquire images in

[3] D. Hall. *Mathematical techniques in multisensor data fusion*. Artech House, Boston, London, 1992.

different modalities (e.g., X-rays, electron microscope, infrared, etc.), each modality having different horizontal and vertical resolutions. Methods have been designed to use one or more modalities to increase the spatial resolution of other modalities in a very realistic way.

Several authors have stressed that large benefits are expected from having synthesized B^*_{ih} images that are close to reality. This sub-domain of the fusion of imaging sensors is getting more and more interest. Makers of instruments are now integrating the capabilities of fusion techniques within the processing software (at ground level for spaceborne systems), and are consequently designing lighter and cheaper observation systems.

Quality is an important topic, especially when industrial systems are under concern. An important part of this book is devoted to the quality assessment of images resulting from fusion process and to the comparison of fusion methods for the synthesis of images.

Chapter 8 deals with the assessment of the quality of the synthetic images produced by methods, such as those described in Chapter 7. How to assess the benefits of the fused products to the visual analysis is described. The means to check whether the fused products meet the theoretical properties of the synthetic products are discussed. A protocol of validation of fused products is presented. Some criteria for a global assessment of the quality are analyzed.

Chapter 9 illustrates both Chapters 7 and 8. Using the protocol for quality assessment, a comparison is performed between the fused products resulting from several methods presented in Chapter 7.

Chapter 2 is written in French and summarizes the content of the book to the attention of the French-reading persons.

2. RÉSUMÉ A L'ATTENTION DU LECTEUR FRANCOPHONE

1. INTRODUCTION

La fusion de données est un terme plutôt récent. Elle traduit une approche du traitement de l'information, adoptée spontanément dans plusieurs domaines, et ce, bien avant que le terme existe. Cette approche est fondée sur l'utilisation de la synergie offerte par les données de sources diverses. L'exploitation conjointe de sources indépendantes et/ou redondantes est connue comme fournissant de meilleurs résultats que l'exploitation des sources individuelles.

On utilise souvent la vision humaine pour illustrer les bénéfices de la fusion de données. Les deux yeux d'un homme observent les objets sous des angles légèrement différents, et cette stéréovision permet la perception du relief. Exploiter conjointement deux yeux étend donc les capacités d'avoir deux fois un œil. Un autre avantage est la redondance. Si l'un des yeux est défaillant, la vision est encore possible, quoiqu'en mode dégradé.

Les recherches et développements en fusion de données ont lieu depuis bien longtemps sous des noms différents. L'utilisation de ce vocable nouveau, "fusion d'informations", "fusion de données", traduit la prise de conscience que, quel que soit le domaine, on retrouve les mêmes problèmes fondamentaux. L'objet de la fusion de données est d'exprimer formellement ces problèmes, notamment en relation avec les fondements des sciences sur laquelle elle s'appuie, comme les mathématiques, la physique ... La recherche en fusion de données vise à exprimer le concept de fusion de données et ses propriétés, et aide à sa mise en œuvre pratique dans différentes applications. On peut par conséquent établir des définitions et un lexique commun à toutes les applications de la fusion de données, permettant ainsi un partage des connaissances plus efficace. On peut mieux décrire les propriétés des données et leurs interactions, développer des méthodes plus appropriées, mieux suivre la qualité de l'information tout au long du processus, mieux concevoir et réaliser des systèmes de fusion de données et les analyser sous différents aspects.

La fusion de données offre de nombreux avantages :

- robustesse et fiabilité ; le système est opérationnel même si l'une ou plusieurs sources d'information sont défectueuses ;
- augmentation de la couverture spatiale et temporelle de l'information et des déductions ;

- accroissement du nombre de dimensions de l'espace des observations, menant à un accroissement de la qualité des déductions, et à une réduction de la vulnérabilité du système ;
- réduction de l'ambiguïté des déductions ; des informations plus complètes ou plus précises permettent un meilleur choix entre les différentes hypothèses ;
- apport d'une solution à l'explosion de la quantité d'informations disponible aujourd'hui.

Les systèmes biologiques exploitent la fusion de données. Une illustration en est donnée par le système humain, qui utilise ses cinq sens pour percevoir son environnement (au sens très large). Les capteurs de notre corps acquièrent des informations par la vue, l'odorat, le toucher, l'ouïe et le goût (fig. 1.1, chapitre 1 "Introduction"). Les données acquises sont traitées par le cerveau. Pour ce faire, le cerveau va utiliser d'autres sources d'information : sa mémoire, son expérience, et ses connaissances *a priori*. En faisant appel à ses capacités de raisonnement, le cerveau "fusionne" toutes les informations, et effectue des déductions afin de produire éventuellement une représentation de cet environnement et ordonner des actions. Cet exemple montre également que la fusion de données est à l'intersection de plusieurs domaines scientifiques, ici les neurosciences, les sciences cognitives et la médecine.

La fusion de données ne se limite pas aux processus biologiques. A l'origine militaire, elle touche énormément de domaines. L'un des plus actifs est celui de l'exploitation des images de satellites, et par extension, de l'observation de la Terre, c'est-à-dire l'exploitation de toutes les observations et mesures concernant la géosphère et la biosphère. L'observation de la Terre est effectuée au moyen de satellites, d'avions, de bateaux et de stations de mesure au sol. Cet ensemble fournit des mesures variées, partiellement redondantes, partiellement complémentaires, qui peuvent être très localisées et intégrées dans le temps, ou bi-dimensionnelles et instantanées (images), ou des profils verticaux, intégrés dans le temps ou non, ou encore des informations tridimensionnelles (profileurs imageurs de l'atmosphère ou de l'océan, portés par satellite ou bateau ou encore opérant depuis le sol). Si l'on considère également la grande quantité d'archives de mesures, la somme d'informations disponibles pour décrire l'environnement croît rapidement. La fusion de données est un sujet de plus en plus actuel, car susceptible d'aider efficacement les scientifiques à extraire des informations de plus en plus pertinentes et précises de toutes ces mesures.

L'ensemble des capteurs pour l'observation de la Terre est extrêmement diversifié. Le spectre de leurs caractéristiques est très large, en termes d'échelles et d'échantillonnage dans l'espace et le temps, et de modalités d'observation. Cette diversité est une source formidable de questions

pratiques, dont la résolution repose sur une bonne compréhension et modélisation de problèmes plus fondamentaux. Par exemple, quels sont les liens entre des mesures de température du sol effectuées à l'aide d'un thermomètre au sol et intégrées sur une heure, et les mesures du même phénomène, mais effectuées depuis l'espace avec un radiomètre mesurant le rayonnement émis, de manière instantanée mais intégrées sur une surface de quelques kilomètres carrés ? La fusion de données est ici à l'intersection de la physique des processus de l'environnement, des sciences de la Terre, des sciences de l'information et des communications, et de la physique de la mesure.

La prédiction du temps est un exemple de fusion de données dans ce domaine (fig. 1.2, chapitre 1). Les satellites météorologiques fournissent des mesures sur l'état tridimensionnel de l'atmosphère et sur les propriétés de surface du sol et de l'océan. Des avions et des ballons opèrent à des altitudes moins élevées. Des dizaines de milliers de stations au sol sont réparties irrégulièrement dans le monde. Elles mesurent les paramètres météorologiques, comme la température, le vent, la pression, etc. Les radars au sol suivent les orages et les cellules de pluie. En mer, des bateaux et des bouées automatiques mesurent également les paramètres météorologiques, ainsi que la houle. Toutes ces mesures sont traitées pour extraire les paramètres géophysiques pertinents, puis transmises par des réseaux de communication spécialisés. Cette somme d'informations, ainsi que les prédictions faites aux instants précédents, sont ingérées par des modèles numériques, au moyen de techniques d'assimilation de données, dans les centres de prédiction du temps. Ces modèles fournissent des prévisions, qui sont utilisées par les professionnels, et sont également diffusées par les *media*.

2. OBJET DE CE LIVRE. SA STRUCTURE

Ce livre a pour premier objet de clarifier le concept de la fusion de données, ou fusion d'informations, et, par conséquent, de mieux le faire comprendre et accepter. Actuellement, pour la plupart de ses utilisateurs, conscients ou non, la fusion de données représente plus un ensemble de techniques et méthodes qu'un cadre formel.

Les cours donnés par l'auteur à des étudiants de troisième cycle universitaire français ou en école d'ingénieurs, ou encore à des ingénieurs confirmés, sont à l'origine de cet ouvrage. Ce livre devrait donc être d'un apport certain aux scientifiques, ingénieurs et autres praticiens intéressés par la fusion de données.

Cet ouvrage se démarque des livres précédents et bien connus[1] [2], en offrant d'une part une discussion approfondie sur les aspects fondamentaux et le concept de la fusion de données, et, d'autre part, un guide, tant théorique que pratique, de certaines techniques de fusion d'images et de l'évaluation des résultats, sans pour autant prétendre à l'exhaustivité.

Ce livre est écrit en anglais. Le présent chapitre a pour objet d'en présenter le contenu de manière synthétique à l'attention des lecteurs francophones.

La première partie du livre présente le concept de la fusion de données. Après avoir discuté l'état actuel des connaissances, elle définit de manière précise ce concept et établit l'essentiel des définitions nécessaires au domaine (chapitre 3). Dans ce même chapitre, sont discutées les propriétés des données devant entrer dans un processus de fusion. Une typologie des problèmes que doit résoudre la fusion de données est également présentée. La représentation d'un processus de fusion fait l'objet du chapitre 4, qui traite aussi des différentes architectures.

La deuxième partie concerne la fusion d'images et la qualité des produits de fusion. Cette partie propose des solutions pratiques. Auparavant, elle offre une présentation approfondie des méthodes les plus usitées ainsi que d'autres plus évoluées. Cette partie s'adresse à un large public s'intéressant à l'imagerie. Il faut toutefois préciser, pour éviter tout désappointement, que, hormis quelques exemples, les techniques présentées sont limitées à la fusion de mesures, en utilisant le pixel comme support d'information. L'accent est mis sur les images de résolution spatiale différente.

L'évaluation de la qualité des produits occupe une part importante de cette deuxième partie. Les protocoles et critères d'évaluation développés sont appliqués à différentes techniques de fusion d'images, afin de les comparer d'une part, et d'illustrer l'évaluation de la qualité, d'autre part.

L'imagerie est de plus en plus utilisée de nos jours : en médecine, processus industriels, gestion du trafic automobile urbain, vérification des traités, gestion de crise naturelle ou politique, etc. Dans une certaine mesure, les informations présentées sous forme de grille, où chaque cellule peut être assimilée à un pixel, peuvent être considérées comme des images acquises par un capteur. C'est le cas notamment de toutes les cartes disponibles sur Internet relatives aux études et analyses du changement global. Les informations peuvent être de type mesures, par exemple, la température de l'air, ou de type catégorie, par exemple, le type de végétation

---

[1] E. Waltz and J. Llinas. *Multisensor data fusion*. Artech House, 1990.

[2] D. Hall. *Mathematical techniques in multisensor data fusion*. Artech House, Boston, London, 1992.

De nombreuses méthodes et techniques existent en fusion de données. La plupart d'entre elles ont été décrites dans des ouvrages de référence. Cependant, peu d'entre elles concernent les images et les capteurs imageurs. Étant donné l'importance de l'imagerie et des développements méthodologiques y afférents, nous avons consacré la seconde partie de cet ouvrage à la fusion d'images. Les images acquises par satellite à haute ou très haute résolution spatiale, offrent une vue inédite de notre environnement avec de nombreux détails. Elles servent souvent d'illustrations dans cette deuxième partie.

L'objet de la fusion d'images se réduit dans ce livre aux problèmes liés aux jeux d'images de résolution spatiale différente et de modalité différente. Il s'agit alors de la création de jeux d'images de dimension réduite, afin de mieux visualiser l'ensemble des informations ou afin d'effectuer la synthèse d'images à meilleure résolution spatiale.

Plusieurs outils mathématiques sont décrits dans le chapitre 5. Ils forment la base de nombreuses méthodes usitées en fusion d'images.

Le chapitre 6 traite de la fusion de données comme un moyen d'analyse détaillée et complète de plusieurs jeux d'images. Les techniques populaires "intensité - teinte - saturation" et "analyse en composantes principales" y sont décrites, avec d'autres.

Le chapitre 7 présente de manière très détaillée les techniques usitées et les plus récentes, afin d'effectuer la synthèse d'images à la meilleure résolution spatiale disponible dans l'ensemble des jeux d'images. Ces images synthétisées doivent être aussi proches que possible des images qui seraient observées dans la même modalité si elle existait avec cette résolution spatiale.

C'est un problème que l'on trouve fréquemment, aussi bien en observation de la Terre, avec des capteurs qui observent dans les bandes bleue, verte, rouge et infrarouge à 4 m de résolution ainsi qu'en mode panchromatique à 1 m de résolution, ou encore dans des systèmes industriels utilisant les rayons-X et les microscopes électroniques pour analyser le même échantillon avec des résolutions horizontales et verticales différentes. Des méthodes ont été et sont élaborées pour accroître la résolution spatiale d'une ou plusieurs modalités de basse résolution spatiale, en utilisant une ou plusieurs modalités de meilleure résolution spatiale.

La qualité est un sujet important, surtout lorsque des systèmes industriels ou opérationnels sont concernés. Une part importante du livre est dévolue à l'évaluation de la qualité des images résultant de processus de fusion et à la comparaison des méthodes de synthèse des images, discutés au chapitre 7.

Le chapitre 8 pose le problème de l'évaluation de la qualité et propose une généralisation de plusieurs protocoles déjà publiés et relatifs aux évaluations tant visuelles que numériques des produits de fusion.

Le chapitre 9 est une illustration des chapitres 7 et 8. D'une part, il montre les résultats des méthodes discutées au chapitre 7 et, d'autre part, il met en œuvre le protocole d'évaluation pour comparer les différentes méthodes.

3. DÉFINITIONS

Si le concept de la fusion de données est facile à comprendre, il est difficile d'en trouver une définition, qui rende compte de ce cadre formel et des multiples facettes de la fusion de données. Une définition doit, de plus, être consensuelle pour être utilisable.

Le chapitre 3 montre la pauvreté des définitions jusqu'alors proposées. Elles sont souvent restreintes à un ensemble d'outils ou de méthodes, voire à un ensemble d'informations. Il est rarement question de qualité et la notion de concept est totalement exclue.

La définition proposée par le Joint Directors of Laboratories (JDL), du ministère de la défense aux États-Unis d'Amérique, est un cas à part[3]. Elle est appelée "le modèle JDL" et a fait l'objet de nombreuses études. Ce modèle fonctionnel met en avant les fonctions principales, les informations pertinentes et les interconnexions, rencontrées dans la fusion de données. Ce modèle donne une définition de la fusion de données comme étant un processus multi-niveaux et à facettes multiples (*sic*) ayant pour objet la détection automatique, l'association, la corrélation, l'estimation et la combinaison d'informations de sources singulières et plurielles. Ce modèle est décrit dans le chapitre 3 en détail.

Le modèle JDL est extrêmement populaire dans le domaine militaire. En fait, il contient une certaine perversité, due à l'association étroite d'une définition, d'un modèle fonctionnel et de quatre niveaux hiérarchiques de traitement liés au modèle fonctionnel. Cette association est tellement étroite qu'aucun de ces éléments ne peut être dissocié des autres. Ceci entraîne une confusion de ce qu'est effectivement la fusion de données, confusion perceptible dans la littérature associée. En particulier, de nombreux articles confondent les niveaux de traitement avec des niveaux sémantiques, et ont tendance à les séparer, contrairement aux intentions des auteurs et des textes initiaux.

---

[3] U.S. Department of Defense, *Data fusion lexicon*, Data Fusion Subpanel of the Joint Directors of Laboratories, Technical Panel for C3, 1991.

Malgré sa popularité et son importance pour le développement de la fusion de données, ce modèle ne constitue pas pour autant, une définition de la fusion de données, et, en aucun cas, ne fait référence à un cadre conceptuel.

Le besoin d'une définition plus appropriée a entraîné la création d'un groupe de travail européen en 1996, sous les auspices de la SEE (société d'électricité et d'électronique), la branche française de l'Institute of Electric and Electronics Engineers (IEEE), et de EARSeL (European Association of Remote Sensing Laboratories), la branche européenne de l'International Society for Photogrammetry and Remote Sensing (ISPRS). Ce groupe a proposé la définition suivante : *la fusion de données constitue un cadre formel dans lequel s'expriment les moyens et techniques permettant l'alliance des données provenant de sources diverses.* Cette définition met clairement l'accent sur le concept et non plus sur les méthodes, techniques ou stratégies.

La définition ajoute que la fusion de données vise à l'obtention d'information de plus grande qualité ; la définition exacte de «plus grande qualité» dépendra de l'application. La qualité est un mot générique indiquant que le résultat de la fusion est plus satisfaisant pour l'usager que l'ensemble de l'information originale. Une meilleure qualité peut signifier une plus grande précision sur une valeur ou l'estimation d'une classe, mais également un meilleur usage des ressources disponibles pour un même résultat.

Au-delà de la définition de la fusion de données, c'est tout un ensemble de termes qui doit être défini. Grâce à l'utilisation des mêmes mots ayant la même signification pour tous, les scientifiques peuvent mieux échanger leurs idées et leurs expériences. Ces connaissances peuvent être mieux diffusées auprès des communautés utilisatrices du savoir scientifique. Partager le même lexique, accepté et connu par tous, permet une profonde irrigation des sociétés par le savoir.

Ce chapitre contient ainsi une liste de termes de référence. Le groupe de travail a préféré adopter des termes déjà usités et bien compris dans d'autres domaines ou présents dans des normes, comme ISO ou CEN. Les principaux termes de référence définis sont:

- combinaison, intégration et assimilation ;
- mesure, signal et observation ;
- objet, attribut et vecteur d'état ;
- règle, décision et représentation.

Ce chapitre discute aussi du problème de l'alignement. Les informations entrant dans un processus de fusion doivent être alignées. Il faut par conséquent définir une représentation commune de toutes ces informations.

Cette opération d'alignement est extrêmement importante. Elle doit être faite avec soin, car elle conditionne les résultats du processus de fusion.

L'alignement, ou conditionnement, ou encore parfois, harmonisation, consiste à définir un espace commun, dans lequel les informations vont être projetées afin d'y être comparables. Lorsque l'on parle d'images, on a souvent affaire à un problème d'alignement géométrique ou de géocodage. Il faut projeter les images dans un même référentiel d'espace. On peut aussi avoir besoin d'un même référentiel de temps, ou encore harmoniser des nomenclatures de classes etc. L'alignement représente un grand ensemble de problèmes, souvent complexes, liés à l'observation (instrumentation et physique de la mesure et des objets observés) et au traitement de l'information. L'alignement fait partie du processus de fusion dans la mesure où cette opération est effectuée afin de satisfaire des contraintes imposées par le processus choisi. Cependant, de plus en plus de fournisseurs d'informations délivrent des informations déjà alignées et prêtes à entrer les processus de fusion les plus courants.

L'association est l'union des différentes représentations issues des informations multi-sources. Ces informations sont alignées. L'association requiert que les représentations se réfèrent au même objet. Il n'y a aucun intérêt à essayer de fusionner des informations, quelles qu'elles soient, ne décrivant pas le même objet ou phénomène. L'association est aussi appelée concaténation, car elle entraîne une augmentation de la taille du vecteur d'état de l'objet considéré. Elle est indépendante du niveau sémantique des informations et s'effectue au moyen d'une analyse du niveau de corrélation et de relation entre les informations à fusionner et l'objet à représenter. L'association peut avoir pour objet la sélection de sous-ensemble de capteurs, qui sont les plus appropriés pour un problème donné.

Un système de fusion est généralement composé de sources d'information, de moyens d'acquisition d'information, de moyens de communications et de capacités de traiter l'information. Il peut être par conséquent très complexe. Il est fréquent et pratique lors de l'étude ou la représentation d'un système, de séparer les aspects topologiques et les aspects de traitement d'information, même s'il existe des interconnexions. Plusieurs taxonomies d'algorithmes de traitement ont été proposées dans la littérature et sont brièvement présentées dans cette partie du chapitre 3.

Enfin, ce chapitre se termine sur une présentation d'une typologie des problèmes de fusion de données. Cette typologie a une influence importante sur le choix de l'architecture du système de fusion, sur les choix d'outils et méthodes de traitement et de communications. Les typologies usuelles, comme "fusion de mesures", "fusion d'attributs" et "fusion de décisions", ou encore "fusion de bas et haut niveau" sont tout d'abord discutées. Elles peuvent parfois prêter à confusion et leur usage devrait être limité. Une

autre typologie est présentée en détails. Elle comprend la fusion d'attributs, la fusion d'analyses et la fusion de représentations.

4. REPRÉSENTATION D'UNE OPÉRATION DE FUSION. ARCHITECTURES

Il est important de pouvoir représenter un processus de fusion de manière simple et schématique. Une telle présentation simple mais précise de la fusion est utile en enseignement et formation des personnes, mais supporte également l'analyse d'un système à un plus haut niveau d'abstraction. L'adoption d'un schéma commun offre de nombreux avantages. Ce schéma doit être indépendant des applications, du type des informations utilisées et du type d'information résultante.

Le modèle JDL est un exemple de schéma. Il est bien sûr parfaitement adapté aux besoins des militaires et est aisément extensible à tout problème de gestion de crises, militaire ou non. Cependant, pour la plupart des autres problèmes, il s'avère plutôt inadapté. En effet, le modèle, qui est assez complexe, ne s'applique que de manière partielle à la plupart des applications.

Ce livre adopte un autre schéma, beaucoup plus simple, tiré de la littérature (fig. 4.1). Ce schéma permet de décrire aussi bien des opérations élémentaires que des opérations complexes. Étant modulaire, il peut être combiné de façon à représenter des systèmes faisant appel à plusieurs processus de fusion. Ce schéma est illustré par de nombreux exemples d'application.

Trois types d'information forment les entrées de la cellule de fusion : les sources d'information à fusionner, les informations auxiliaires et les connaissances externes. Les sources d'information doivent être alignées. Elles peuvent être constituées des sorties de capteurs, et, plus généralement, de mesures, ou d'attributs ou encore de décisions. Les informations auxiliaires apportent des informations supplémentaires, résultant, par exemple, d'un traitement particulier d'une source spécifique, ou d'une autre opération de fusion. Dans le cas de processus itératifs, incluant des opérations dépendantes du temps, les résultats de l'itération précédente deviennent des entrées de l'itération courante. Ils sont considérés comme des informations auxiliaires, car ne provenant pas des sources originales. Les connaissances externes forment aussi une information additionnelle, dont l'objectif est principalement de contraindre ou guider le processus de fusion, par exemple, en imposant des règles *a priori*. *A priori* signifie que la connaissance est disponible avant que la fusion n'ait lieu.

Les architectures de fusion décrivent l'ensemble des sources, la manière dont elles sont assemblées et les techniques mathématiques pour le

traitement. Ce chapitre donne les bases pour comprendre et concevoir des architectures. Il n'est pas un guide de mise en œuvre. En effet, la variété des applications de fusion est telle qu'il est impossible de fournir de tels guides pratiques applicables à tous les cas.

Trois types d'architectures sont définis : centralisée, décentralisée (parfois appelée autonome) et hybride. L'architecture centralisée exploite en un seul lieu, simultanément ou non, l'ensemble des informations disponibles (fig. 4.8). L'avantage théorique de la fusion centralisée est qu'elle fournit le meilleur résultat possible puisque la décision est prise en considérant toute la connaissance disponible. Si une source est très bruitée, cet avantage peut devenir un défaut car cette source peut contaminer l'ensemble de l'information et entraîner une diminution de la qualité du résultat. L'architecture centralisée requiert la disponibilité de toutes les informations en un même lieu, ce qui implique en particulier, des moyens de communication appropriés. Elle impose également une charge de calcul importante. À chaque changement d'entrée, l'ensemble des calculs doit être fait.

L'architecture décentralisée offre une grande flexibilité et modularité (fig. 4.10). La fusion de données est effectuée en plusieurs opérations s'effectuant, pour les premières, sur chaque source ou sous-ensemble de sources. Les résultats sont ensuite les entrées d'un processus de fusion final. Cette architecture est recommandée dans les domaines risqués, par exemple, lorsque les communications ne sont pas fiables ou lorsque les modes opératoires des capteurs sont soumis à de forts aléas. Un autre avantage de cette architecture réside dans la faible charge de calcul. Dans le cas notamment des capteurs asynchrones, les calculs sont actualisés au rythme d'acquisition de chaque capteur et non au rythme le plus rapide, comme dans le cas d'une architecture centralisée.

D'autres architectures peuvent être conçues, à partir d'un mélange des architectures centralisée et décentralisée. Elles sont appelées hybrides. Selon leur conception, elles combinent les avantages et inconvénients de l'une ou l'autre architecture de base.

Le choix d'une architecture n'est pas toujours chose aisée. L'architecture centralisée doit être préférée dès que possible, car elle produit les meilleurs résultats. Cependant, chaque architecture a ses propriétés et il convient de les analyser avant de se décider pour une architecture centralisée ou décentralisée, voire hybride afin de tirer le meilleur parti des propriétés de chacune. Des compromis sont souvent nécessaires en fonction des bandes passantes pour les communications et de leur fiabilité et sécurité, des types d'information à fusionner, des types de capteurs, de l'application elle-même, des méthodes mathématiques mises en jeu, de la mise en œuvre du système complet, de son implantation physique, etc.

5. QUELQUES OUTILS MATHÉMATIQUES POUR LA FUSION D'IMAGES

Le chapitre 5 est consacré à la présentation de quelques outils mathématiques, formant la base de nombreuses méthodes usitées en fusion d'images. Le parti pris de ce chapitre est de se limiter à quatre outils, utilisés dans la suite du livre, de les détailler et de fournir les bases algorithmiques permettant leur mise en œuvre numérique.

Le premier outil est relatif à l'espace des couleurs. Cet espace à trois dimensions est souvent représenté, notamment dans le monde de l'éclairage et de l'électronique, par trois composantes : teinte, saturation, brillance. La teinte distingue les couleurs : rouge, jaune, bleu, etc. La saturation se réfère à la pureté, c'est-à-dire comment la couleur est diluée par la lumière blanche. Elle permet de distinguer par exemple, le bleu marine du bleu ciel. La brillance est équivalente à l'intensité de la lumière achromatique. Trois couleurs primaires ont été définies par la Commission Internationale pour l'Eclairage (CIE). Combinées, elles permettent de retrouver toutes les couleurs possibles. A partir de ce standard, d'autres standards ont été développés pour répondre à certaines applications. Parmi ceux-ci, le standard dit RGB selon les initiales en anglais de Rouge, Vert, Bleu, a été défini pour les besoins de la télévision et de l'affichage numérique d'images. La conversion entre le standard RGB et le système (teinte, saturation, brillance) n'est pas triviale. D'ailleurs, on utilise plus souvent le système TSI (teinte, saturation, intensité) pour modéliser cette conversion. L'intensité est, à quelques nuances près selon les modèles, la moyenne des couleurs Rouge, Vert et Bleu. Le chapitre 5 présente deux modèles de conversion RGB - TSI (IHS en anglais) et la réciproque TSI - RGB.

D'un point de vue mathématique, cette conversion RGB - TSI s'apparente à un problème de projection d'un repère dans un autre. L'intérêt de la projection réside dans le fait que certaines opérations sont plus aisées dans le deuxième repère.

Une autre technique de projection est l'analyse en composantes principales, connue aussi sous le nom de la transformation de Karhunen-Loeve. Soit un ensemble de N images. L'analyse en composantes principales fournira un ensemble de N nouvelles images, appelées composantes. La caractéristique de ces composantes est qu'elles sont orthogonales, c'est-à-dire décorrélées. Les composantes sont ordonnées par décroissance de la variance. Le calcul des composantes principales s'effectue par diagonalisation de la matrice de variance - covariance, ou encore de la matrice de corrélation, des N images d'origine.

Outre ces deux outils de projection, le chapitre 5 présente des outils d'analyse spatiale de l'image : la transformée en ondelettes et l'analyse multirésolution. Si la transformée de Fourier est un excellent outil pour l'analyse du domaine fréquentiel (plus exactement, des vecteurs d'onde)

d'une image, la transformée en ondelettes permet d'observer à la fois le signal et ses fréquences. C'est une transformée temps-fréquence. Quant à l'analyse multirésolution, c'est un moyen de décrire et modéliser de manière exacte et inversible un signal et ses fréquences. La combinaison de l'analyse multirésolution et de la transformée en ondelettes forme un outil performant et pratique pour décrire, analyser et modéliser le contenu spatial d'une image.

Le chapitre 5 présente ces deux outils dans leurs principes. Il propose ensuite des éléments pour une mise en œuvre aisée de deux algorithmes. L'un est l'algorithme de Mallat, combiné ici avec une transformée en ondelettes de Daubechies. Cet algorithme est dit pyramidal et comprend une décimation des images au fur et à mesure de l'analyse. L'autre est l'algorithme dit "à trous". Il ne comprend pas de décimation. Leurs propriétés respectives sont discutées.

6. FUSION D'IMAGES

L'approche générale en fusion d'images est de créer un nouvel ensemble d'images *I*, généralement de dimension réduite, à partir de l'ensemble original d'images *A, B, C,...* :

$$I = f(A, B, C, D,...)$$

Un exemple classique de la fusion d'images est la classification. Ce chapitre présente brièvement la classification, l'identification et la reconnaissance de formes en tant que processus de fusion.

Le chapitre 6 a pour objet de décrire complètement quelques techniques populaires utilisées en fusion d'images pour une analyse visuelle de l'ensemble des images disponibles, qu'elles soient de modalités différentes ou multi-temporelles, ou une combinaison des deux. La technique d'incrustation est un moyen efficace de fusionner des observations non-commensurables.

Une condition nécessaire à la fusion d'images est très souvent l'alignement géométrique des images. Il s'agit d'une des opérations les plus critiques de l'alignement. Il est aussi appelé co-enregistrement, superposition, correction géométrique, géocodage ou navigation. Ce chapitre traite de ce problème de manière détaillée. L'alignement peut être effectué de manière absolue, c'est-à-dire par rapport à un repère non entièrement lié au problème courant. Un exemple d'un tel référentiel est le système canonique en latitude - longitude. L'alignement peut aussi être effectué de manière relative. Une image, ou de manière générale, une source, est sélectionnée qui sert de référence. L'alignement géométrique est décrit par un modèle, parfois analytique, souvent obtenu par ajustement à l'aide de points similaires observés dans les images, permettant de convertir une géométrie en une autre.

Assez souvent, l'application du modèle géométrique s'accompagne d'un ré-échantillonnage des images. La commodité résultant de l'obtention d'un jeu d'images totalement homogène d'un point de vue géométrique et taille de pixel, en est la raison majeure. Ce ré-échantillonnage est également une opération critique puisqu'il va transformer le contenu des images originales.

Dans les méthodes de projection - substitution abordées dans ce chapitre (IHS, PCA), l'alignement de la dynamique du signal est nécessaire. Les observations de certaines sources et combinaisons de sources doivent être converties, souvent par des fonctions affines, afin d'être similaires. La similarité est souvent représentée par les premiers moments statistiques : moyenne, variance. Dans la mesure où aucune loi de la physique n'est requise dans ce type d'approche, elle peut être utilisée pour fusionner des informations à des fins de visualisation et d'analyse de sources hétérogènes ou homogènes.

La combinaison colorée est un moyen très usité pour visualiser un ensemble d'images. Soit un triplet d'images. A chaque image est allouée une voie (voie rouge, verte et bleue). La combinaison de ces trois voies produit une couleur, fonction des valeurs originales dans le triplet. S'il n'y a pas exactement trois sources à l'origine, le triplet est construit par sélection arbitraire des sources ou par combinaison des sources. La couleur peut être projetée dans le système (teinte, saturation, intensité). Sachant que l'intensité lumineuse porte l'information structurant l'ensemble des images, on comprend que l'on peut moduler / transformer / substituer les hautes fréquences d'ensemble originales à l'aide d'une autre information non prise en compte dans la combinaison colorée. Un exemple est donné, concernant la création d'une interface plus conviviale pour l'exploitation de données de type géographique.

Deux techniques sont principalement utilisées : la technique IHS (utilisation de la conversion RGB - IHS) et la technique PCA (principal component analysis). Dans cette dernière, la projection s'effectue par analyse en composantes principales. La première composante joue le rôle de l'intensité dans la méthode IHS. C'est elle qui sera modifiée par l'information à fusionner. Enfin, une projection inverse est effectuée, pour revenir au référentiel original des sources. Cette projection inverse n'est généralement pas effectuée si l'application ne concerne qu'une analyse visuelle des combinaisons colorées. D'autres techniques peuvent être conçues en appliquant ce principe à l'aide d'autres transformées, orthogonales ou non.

L'incrustation est une forme triviale de fusion. Il s'agit d'incruster dans des images ou combinaisons d'images, des éléments provenant d'autres sources. Ces éléments peuvent être des observations ou des attributs. C'est un moyen efficace de fusionner des observations non-commensurables, permettant d'augmenter et affiner la perception et l'analyse visuelle de ces éléments,

notamment, en créant des images composites mettant en avant l'information essentielle pour l'application. Plusieurs techniques sont disponibles. Ce chapitre en traite une particulière, à l'aide d'un exemple.

7. FUSION POUR LA SYNTHÈSE D'IMAGES À MEILLEURE RÉSOLUTION SPATIALE

Les chapitres 7, 8 et 9 traitent d'un problème de fusion de données particulier : étant donné un ensemble d'images multi-modalités possédant des résolutions spatiales différentes, le but du processus de fusion est d'effectuer la synthèse de certaines de ces images à la meilleure résolution spatiale disponible dans l'ensemble original. Ces images synthétisées doivent être aussi proches que possible des images qui seraient observées dans la même modalité si elle existait avec cette résolution spatiale.

De nombreux travaux ont démontré l'intérêt de telles images synthétiques et ce domaine de recherche reçoit de plus en plus d'attention. L'intégration des capacités de telles techniques dans les systèmes d'observation peut aussi mener à des instruments et systèmes aussi performants mais moins complexes, plus robustes et moins chers.

Soit B_l, les images de basse résolution spatiale l et A_h les images de plus haute résolution spatiale h. Chaque ensemble d'images a été acquis par plusieurs modalités. Il est possible d'étendre le problème à plusieurs résolutions spatiales. Le problème général est la construction d'un nouvel ensemble d'images B^* :

$$B^* = f(A, B)$$

Ces images synthétiques B^*_h doivent être proches de la réalité et respecter les trois propriétés suivantes.

Première propriété. Toute image synthétique B^*_h ramenée à la résolution originale l, doit être identique à l'image originale B_l.

Deuxième propriété. Toute image synthétique B^*_{kh} dans une modalité donnée k doit être identique à l'image B_{kh} qui serait observée dans la même modalité si elle existait avec cette résolution spatiale.

Troisième propriété. L'ensemble multi-modalités synthétique B^*_h doit être identique à l'ensemble multi-modalité B_h qui serait observé avec les mêmes modalités si elles existaient avec cette résolution spatiale.

De nombreuses méthodes ont été publiées. Elles diffèrent essentiellement par la manière dont elles respectent ces trois propriétés. On distingue trois groupes de méthodes. Ces trois groupes sont discutés en détail dans ce chapitre. Les propriétés, avantages et inconvénients, de ces méthodes, que l'on peut déduire de l'analyse de leurs équations, sont mises en avant. Ce

chapitre traite également des aspects pratiques de mise en œuvre des méthodes présentées dans chacun des groupes.

Ces groupes sont :

- projection et substitution : ces méthodes sont présentées dans le chapitre 6. Quelques variantes existent, mais, dans l'ensemble, les méthodes IHS et PCA sont de loin les plus usitées ;
- *relative spectral contribution* : ces méthodes exploitent des relations qui pourraient exister entre les différentes modalités si elles étaient ramenées à la même résolution spatiale, soit en fonction des instruments d'observation eux-mêmes, soit en fonction de la nature des objets observés. La méthode P+XS de l'agence française spatiale (CNES), la transformée de Brovey et la méthode de "couleur normalisée" sont les plus connues de ce groupe ;
- concept ARSIS : ce concept (amélioration de la résolution spatiale par injection de structures) utilise des techniques de multirésolution, ou multi-échelle, et de filtrage sélectif de fréquences afin d'injecter dans les images à basse résolution les hautes fréquences à la plus haute résolution. De nombreuses méthodes et variantes ont été développées ces dernières années sur ce concept. Le concept ARSIS se démarque des deux autres groupes par, d'une part, la prise en compte de la première propriété lors de la construction des méthodes, et, d'autre part, par une séparation explicite du modèle d'analyse et de synthèse de l'information fréquentielle, du modèle de conversion de l'information entre les modalités et du modèle de transformation de ce modèle de conversion lors du changement de résolution.

8. ÉVALUATION DE LA QUALITÉ DES IMAGES SYNTHÉTISÉES

La qualité des méthodes et des images synthétisées fait l'objet du chapitre 8. Il s'agit d'un sujet important ayant un impact fort sur la mise en œuvre industrielle de telles méthodes et sur l'acceptation de ces synthèses par leur public.

Le problème de l'évaluation de la qualité des synthèses est posé dans ce chapitre. On propose un nouveau protocole, qui est une généralisation de plusieurs protocoles déjà publiés. La standardisation des protocoles contribue à une meilleure acceptation des méthodes par l'industrie et des produits par leurs clients. Ce protocole exploite les trois propriétés définies au chapitre précédent et comprend des évaluations tant visuelles que numériques des produits de fusion.

Ce protocole fait appel à une référence qui fait souvent défaut. Ce chapitre explique comment pallier ce manque. Il passe en revue certaines approches proposées et leurs avantages et inconvénients. On décrit l'influence de

certaines hypothèses sur les résultats. Lors de la mise au point des méthodes, on souligne l'importance du choix des scènes observées. Elles doivent comprendre un contenu important en hautes fréquences spatiales, lié à une forte variation du signal inter-modalités.

La comparaison objective de la qualité d'images multi-modalités est une tâche difficile et fastidieuse. Le système visuel humain diffère d'un individu à l'autre, et, pour un même individu, ne réagit pas de manière égale à diverses distorsions visuelles. La qualité perçue par un observateur dépend ainsi fortement de l'observateur et de l'application. Un panel d'analystes est formé, qui va évaluer les produits de fusion au regard de critères bien définis. Un score moyen est établi à partir des notations individuelles. Ce chapitre donne un exemple de critères utilisés par le ministère de la défense des États-Unis d'Amérique. Lors de l'analyse visuelle, la notion de résolution d'image est importante envers l'interprétabilité de l'image. Un modèle est proposé, permettant de prédire la résolution effective de l'image en fonction des résolutions *l* et *h*.

Par ailleurs, des calculs sont effectués sur les produits de fusion par comparaison avec les références. Des critères numériques sont proposés, de manière à quantifier objectivement les différents aspects des produits de fusion. Il s'agit souvent de quantités statistiques résumant les similitudes et différences entre les références et les produits de fusion, au regard des trois propriétés énoncées. Ces mesures de performance offrent l'avantage d'être faciles à mettre en œuvre et d'être automatisables, par exemple, au sein d'une ligne de production.

Le besoin d'une quantité simple exprimant de manière globale mais représentative, la qualité d'un produit a déjà été exprimée. On montre qu'une telle quantité doit remplir trois contraintes : indépendance vis-à-vis des unités des mesures, des étalonnages des instruments et de leurs gains, du nombre de modalités et des résolutions *l* et *h*. Plusieurs quantités sont passées en revue vis-à-vis de ces contraintes. Une analyse bibliographique suggère que la quantité, appelée ERGAS (erreur relative globale adimensionnelle de synthèse), soit un bon candidat. Un seuil de 3 semble séparer des produits satisfaisants des produits insatisfaisants. Plus la quantité ERGAS est faible, plus faible est l'erreur de manière globale, meilleure est la qualité.

9. ANALYSE ET COMPARAISON DE DIFFÉRENTES MÉTHODES

Les méthodes discutées au chapitre 7 sont comparées dans ce chapitre à l'aide de cas concrets. Pour cette comparaison, le protocole présenté au chapitre 8 est mis en œuvre. Le présent chapitre repose sur un large ensemble de comparaisons entre différentes méthodes, soit publiées, soit réalisées au sein de l'École des Mines de Paris. Seules ont été prises en

compte les comparaisons effectuées selon le protocole décrit précédemment. Les critères sélectionnés dans les publications attachent beaucoup d'importance à la synthèse du signal pour chaque modalité et pour l'ensemble multi-modalités, soit de manière globale (par exemple, respect des moyenne et variance), soit au niveau du pixel. Il s'agit des aspects les plus importants pour l'application ultérieure d'algorithmes de classification multi-modalités aux images synthétiques.

A cette comparaison, s'ajoute une discussion détaillée sur l'influence de l'intervalle de temps séparant les instants d'acquisition des différentes modalités et la manière dont les différentes méthodes prennent en compte cet intervalle. Une évaluation analytique de cette influence montre la forte relation existant entre elle et les performances vis-à-vis de la première propriété "toute image synthétique $B*_h$ ramenée à la résolution originale l, doit être identique à l'image originale B_l". Plus cette propriété est respectée, plus l'influence de l'intervalle de temps sera faible. Exceptée la méthode HPF, les méthodes du groupe du concept ARSIS sont ainsi nettement plus insensibles que les autres à l'intervalle temporel d'acquisition. Cette analyse est illustrée au moyen d'un cas très spectaculaire sur le gigantesque barrage des Trois Gorges en Chine. Les géologues chinois ont dans ce cas, constitué le panel d'analystes pour l'évaluation visuelle de la qualité. De même, leurs outils de classification et de détection de failles ont été utilisés pour quantifier leur degré de satisfaction.

En conclusion de ce chapitre, les méthodes sont ordonnées en fonction des performances atteintes. Ces performances sont des moyennes ; il est possible d'observer des fluctuations dans ce classement.

La **transformée de Brovey** et la méthode "**couleur normalisée**". Ces deux méthodes comportent un fort biais dans leur construction et produisent des erreurs importantes. On observe également une forte distorsion du contenu multi-modalités.

La **méthode HPF**. En tant que réalisation possible du concept ARSIS, de meilleurs résultats étaient attendus. On observe un renforcement très excessif des structures. La synthèse de l'ensemble multi-modalité est généralement mauvaise.

Les méthodes de projection-substitution : **IHS** et **PCA**. Ces méthodes produisent des synthèses de qualité variable et souvent mauvaise. On observe une distorsion du contenu multi-modalités. La méthode PCA doit être préférée à la méthode IHS, de manière générale.

L'**interpolation**. L'interpolation n'est pas une méthode de fusion, bien entendu. Ses résultats indiquent le possible bénéfice d'une méthode de fusion. Ainsi, pour la classification multi-modalités, il vaut mieux effectuer une interpolation que l'une des méthodes citées au-dessus.

La **méthode P+XS**. Les méthodes faisant appel à la contribution spectrale relative généralisée donnent des résultats meilleurs que le groupe projection - substitution, sans être toutefois satisfaisants. Les contours sont trop renforcés et la synthèse du contenu multi-modalités comporte de nombreuses erreurs. Cette méthode est très sensible à l'intervalle de temps séparant les acquisitions des différentes modalités.

Les méthodes **Model 1, Model 2** et **RWM**. Ces trois méthodes sont des réalisations du concept ARSIS. Les deux méthodes Model 2 et RWM offrent les meilleurs résultats. La qualité est généralement constante. Ces méthodes sont également plutôt insensibles à l'intervalle de temps séparant les acquisitions des différentes modalités.

La conclusion générale est que très peu de méthodes aboutissent à des résultats satisfaisants. Même si les méthodes Model 2 et RWM donnent très souvent satisfaction, des améliorations restent à apporter. Elles portent tant sur le modèle de représentation de l'information spatiale que sur le modèle de conversion de l'information fréquentielle ou contextuelle d'une modalité à l'autre. Un fort accroissement de la qualité est attendu d'une amélioration significative de la modélisation inter-modalités. Actuellement cette modélisation relève souvent d'un ajustement de dynamique soit au niveau de l'image originale, soit au niveau d'une fenêtre de fréquences. Une meilleure prise en compte des lois de la physique devrait améliorer les résultats.

PART 1.

CONCEPT OF DATA FUSION

DEFINITIONS

ARCHITECTURES

3. DEFINITIONS

THE QUEST FOR AN APPROPRIATE DEFINITION OF DATA FUSION

The concept of data fusion is easy to understand. As explained before, we are all performing data fusion without naming it. Thus speaking of data fusion, of its properties, its fundamentals should be easy. Not at all! Several years ago, as mathematicians told the author that he was doing data fusion, he asked for a definition. Then the author was stunned by the poverty of the few definitions, the lack of clarity and consensus, and by the battle of words. The exact meaning of data fusion varied from one scientist to another. Several words appeared, such as merging, combination, synergy, integration, ... Some scientists said that merging was not fusion, or that fusion was more than merging; others argued that data fusion was no more than optimal control etc. All these words and expressions appealed more or less to the same concept but without expressing it, and were however felt differently. Data fusion also became fashionable, making things less easy. Several times, the term data fusion was used while classification would have been more appropriate, given the contents of the publication. Another striking aspect was that data fusion was often referred to as a collection of methods and techniques.

Actually data fusion means a very wide domain. It is multidisciplinary by essence and is at the crossing of several sciences. It gathers together a large number of methods and mathematical tools, ranging from spectral analysis to plausibility theory. Fusion is not specific to a theme or an application. On the contrary, the tools used in a fusion process for a specific application may be tailored to that specific case.

Data fusion should be seen as a concept, not merely as a collection of tools and means. A formal framework permits a better understanding of the fundamentals and properties of data fusion. It offers the advantages of a better description and formalization of the potentials of synergy between the various sources of information, and accordingly, a better exploitation of this information. It helps in organizing the richness of this domain in order to extract more benefit. It increases understanding between the various sciences; it brings mutual enrichment by sharing knowledge in fundamentals as well as in techniques and solutions.

Expressing the concept of data fusion requires establishing terms of reference. Such terms allow the scientific community to express the same ideas using the same words and also to disseminate their knowledge towards the industry and 'customers' communities. Moreover it is a *sine qua*

non condition to set up clearly the concept of data fusion and the associated formal framework.

Eventually the author discovered that introducing the concept of data fusion strongly increases the awareness of the scientific community on the whole chain of acquisition and processing of the information, ranging from the sensor to the decision, including the management, assessment and control of the quality of the information.

Surprisingly, it is very difficult to provide a precise definition of data fusion. This large domain cannot be simply defined by restricting it, for example, to specific modalities, or specific wavelengths, or specific acquisition means, or specific applications. Fusion process may call upon so many different mathematical tools that it is impossible to define fusion by these tools.

A few definitions can be found in the literature, apart that of the JDL discussed later. In geography, including images from airborne or spaceborne instruments and analysis of collected intelligence, the documents of the Open GIS consortium[1] define fusion as « the process of organizing, merging and linking disparate information elements (e.g., map features, images, text reports, video, etc.) to produce a consistent and understandable representation of an actual or hypothetical set of objects and/or events in space and time ». In these documents, fusion is clearly a set of algorithms, techniques and operators. Fusion is conceived mostly as an analyst-driven process. They further define merging and integration as « the process of physically merging two data sets into a common, or fused, representation ».

In Earth observation from space, Pohl, Van Genderen[2] proposed « image fusion is the combination of two or more different images to form a new image by using a certain algorithm », which is restricted to images. Mangolini[3] extended data fusion to information in general and added a reference to quality. He defined data fusion as a « set of methods, tools and means using data coming from various sources of different nature, in order

---

[1] *Geospatial fusion services testbed*. The Open GIS Consortium (OGC), Wayland, Ma, USA, 2000.

[2] C. Pohl, and J. L. van Genderen. *Multisensor image fusion in remote sensing: concepts, methods and applications*. International Journal of Remote Sensing, vol. 19, n° 5, pp. 823-854, 1998.

[3] M. Mangolini. *Apport de la fusion d'images satellitaires multicapteurs au niveau pixel en télédétection et photo-interprétation*. Thèse de Doctorat, Université Nice - Sophia Antipolis, France, 174 p., 1994.

to increase the quality (in a broad sense) of the requested information ». These definitions put the accent on the methods. They contain the large diversity of tools, but are restricted to these.

In applied mathematics and image processing, the definition proposed by Hall, Llinas[4] also refers to information quality and details the purposes of the data fusion. But it still focus on the methods: « data fusion techniques combine data from multiple sensors, and related information from associated databases, to achieve improved accuracy and more specific inferences that could be achieved by the use of a single sensor alone ». Li *et al.*[5] wrote « fusion refers to the combination of a group of sensors with the objective of producing a single signal of greater quality and reliability ». Quality and reliability are referred to, but there is no reference to concepts. Furthermore it is restricted to sensors and signal.

Indeed most of these definitions are focusing too many on methods though paying some attention to quality. As a whole, there is no reference to concept in these definitions while the need for a conceptual framework was clearly expressed by the scientists as well as practitioners.

THE JDL DEFINITION

Special consideration should be devoted to the works performed by the Department of Defense of the United States of America, and especially by the Data Fusion Subpanel of the Technology Panel for C3 (command, control, communications) of the Joint Directors of Laboratories (JDL).

The JDL developed a functional model that illustrates the primary functions, relevant information and databases, and interconnectivity to perform data fusion. The JDL also gave a definition of data fusion[6], which was further refined[7] as a « multilevel, multifaceted process dealing with the automatic detection, association, correlation, estimation, and combination of

---

4 D. L. Hall, and J. Llinas. *An introduction to multisensor data fusion*. In Proceedings of the IEEE, vol. 85, n° 1, pp. 6-23, 1997.

5 H. Li, B. S. Manjunath, and S. K. Mitra. *Multisensor image fusion using the wavelet transform*. Computer Vision, Graphics, and Image Processing: Graphical Models and Image Processing, vol. 57, pp. 235-245, 1993.

6 U.S. Department of Defense, *Data fusion lexicon*, Data Fusion Subpanel of the Joint Directors of Laboratories, Technical Panel for C3, 1991.

7 DSTO (Defence Science and Technology Organization) Data Fusion Special Interest Group, *Data fusion lexicon*. Department of Defence, Australia, 7 p., 21 September 1994.

data and information from single and multiple sources ». This definition is more general than the previous ones with respect to the types of information than can be combined. It is very popular in the military community.

This definition cannot stand alone. The word "multilevel" refers to the four levels of the functional model, *i.e.* how the processing is organized. Consequently the description of the functional model should accompany the definition.

Figure 3.1 displays this model, revised by the Australian Department of Defense (DSTO). Refinements have been made to this model since then, with especially the introduction of a Level 0 "preprocessing" operating at the sensor level, but they do not impact on the following discussion.

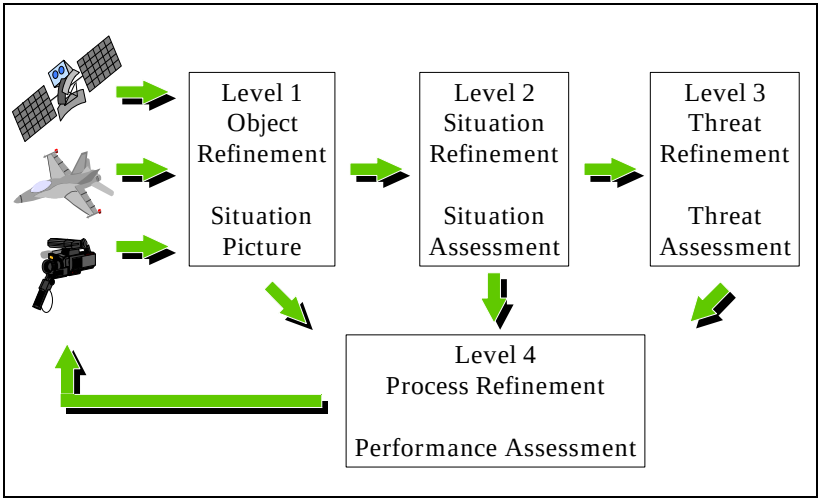


Figure 3.1. The data fusion model of the Australian Department of Defense (courtesy D. Kewley)

The model comprises four levels, noted levels 1 to 4. They form a hierarchy of processing.

In Level 1, is performed "object refinement". This is an iterative process of fusing data to determine the identity and other attributes of entities and also to build tracks to represent their behavior. The term *entity* refers here to a distinct object. A track is usually directly based on detections of an entity, but can also be indirectly based on detecting its actions. The product from this level is called the situation picture. That is, Level 1 tries to determine what it is (*i.e.* identification) and where it is and when (*i.e.* tracking).

Level 1 is usually partitioned into four functions[8]: data alignment, association, tracking and identification. Data alignment functions project data into a common reference frame. Association tackles the problem of sorting or correlating observations into groups, with each group representing data related to a single entity. Tracking refers to the estimation of the position and velocity of the entity. Identification seeks to better identify / describe the entity.

Level 2 performs "situation refinement", which is an iterative process of fusing the spatial and temporal relationships between entities to group them together and form an abstracted interpretation of the patterns in the order of battle data. The product from this level is called the situation assessment[9].

Level 3 performs "threat refinement", which is an iterative process of fusing the combined activity and capability of enemy forces to infer their intentions and assess the threat that they pose. The product from this level is called the threat assessment.

Level 4 performs "process refinement", which is an ongoing monitoring and assessment of the fusion process to refine the process itself and to regulate the acquisition of data to achieve optimal results[10]. Level 4 interacts with each of the other levels.

Notwithstanding the large use of the functional model, the JDL definition is not suitable for the concept of data fusion, since it includes its functionality, as well as the processing levels. Its generalities as a definition for the concept are reduced.

In the literature, especially that devoted to defense systems, this definition is necessarily associated with a description of the four hierarchical levels. Contrary to the intention of the authors of the JDL definition, Level 1 is very often said "low-level" processing, and the others are said "high-level" processing. The association of the definition, the functional model and the way of presenting the levels create confusion in the use of terms. In particular, many documents refer to the "low-level" and "high-level" fusion. However, the concept of fusion does not call upon such levels and such a hierarchy in processes. Often documents contradict themselves by saying that levels are linked in an ascending mode or in a descending one. At times, it is even written that discrimination between levels is impossible.

---

8 D. Hall. *Mathematical techniques in multisensor data fusion*. Artech House, Boston, London, 1992.

9 DSTO. *Op. cit.*

10 L. A. Klein. *Sensor and data fusion concepts and applications.* Tutorial texts, vol. TT 14, SPIE Optical Engineering Press, USA, 131 p., 1993.

The confusion is further enforced by references to the level of semantic content and to the level of inference. A measurement has a lower semantic content than the attributes that are deduced. In turn, the attributes have a lower semantic level that the decisions that are taken, which have a lower level than the meta-decisions etc. The qualification of a semantic level or an inference as "low level" or "high level" depends upon the context and is not absolute.

Data fusion applies to all semantic levels[11], and this at all functional level defined by the JDL model. Therefore, it is impossible to establish a hierarchy in data fusion, which is general and can be always applied. Accordingly, such a hierarchy does not exist from a conceptual point of view, and should not be evidenced in the definition.

Additional questions arise. How can the JDL model be applied outside the military domain? As seen before, weather forecasting is a perfect example of a system calling upon data fusion. If one may find the equivalent of Level 1 "situation picture" and Level 4 "performance assessment" in weather forecasting, the analogy with the two others is far from obvious. They may simply not exist in this case and in others.

In our modern world, information is sold by specialized companies or institutes (e.g., geographical databases). The sensor / data acquisition systems are outside the control of the fusion process (Level 4). The fusion process is open-looped and optimal acquisition of data cannot be realized. Furthermore, the process of alignment has been already performed by the provider, and is further excluded from the subsequent fusion operations made by the customers. These are a few examples showing the difficulties in applying the JDL model.

Actually, the definition of the JDL is not suitable for defining the concept of data fusion. Nevertheless, the influence of the JDL functional model on the development of data fusion has been, and still is, instrumental. Though not presenting a real formal framework, this hierarchy of processing levels has permitted practical implementations and to develop several projects of importance, contributing to a better understanding of the principles. A work is currently under way to extend the model in the framework of the UML language (universal modeling language)[12].

---

11 L. F. Pau. *Sensor data fusion*. Journal of Intelligent and Robotics Systems, vol. 1, pp. 103-116, 1988.

12 C. Kobryn. *UML 2001: a standardisation odyssey*. Communications of the ACM, 42, 10, 1999.

A NEW DEFINITION IN DATA FUSION

In data fusion, information may be of various kinds, ranging from measurements to verbal reports. Some data cannot be quantified; their accuracy and reliability may be difficult to assess. In mapping activities, one often uses some features held in a geographical information system to help in classifying multispectral images provided by several sensors airborne or spaceborne. In this particular case, some data are measurements of electromagnetic energy, and others may be symbols.

Accordingly the definition for data fusion should not be restricted to data output from sensors (signal). Opposite to most of the published definitions, it should not be restricted to methods and techniques or refer to functional models or architectures of systems.

Considering the lack of appropriate definition, a European working group was formed in 1996[13], under the auspices of the SEE, the French affiliate of the Institute of Electric and Electronics Engineers (IEEE), and the EARSeL, the European affiliate of the International Society for Photogrammetry and Remote Sensing (ISPRS). During several meetings, the debate focused on the formalization of the data fusion in remote sensing. The main outcomes of the debate were on definitions and terms of reference. The following definition was finally agreed upon in January 1998[14].

Data fusion: *data fusion is a formal framework in which are expressed the means and tools for the alliance of data originating from different sources*. (In French: la fusion de données constitue un cadre formel dans lequel s'expriment les moyens et techniques permettant l'alliance des données provenant de sources diverses.)

This definition is clearly putting an emphasis on the framework and on the fundamentals underlying data fusion instead of on the tools and means themselves, as is done usually. The latter have obviously strong importance but they are only means not principles.

Note that the word "data" in data fusion is taken in a broad sense. It may be replaced by information fusion.

---

13 L. Wald. *The present achievements of the EARSeL - SIG "data fusion"*. In : Proceedings, EARSeL Symposium 2000 "a decade of trans-European remote sensing cooperation", Dresden, Germany, Buchroithner M. ed., Balkema, Rotterdam, pp 263-266.

14 L. Wald. *Some terms of reference in data fusion*. IEEE Transactions on Geosciences and Remote Sensing, 37, 3, 1190-1193, 1999.

Note also that in this definition, the different observation modalities of one sensor (e.g., multispectral channels) are to be considered as different sources, as well as observations taken at different times by the same sensor.

The definition adds that data fusion *aims at obtaining information of greater quality; the exact definition of 'greater quality' will depend upon the application*. Here quality does not have a very specific meaning. It is a generic word denoting that the resulting information is more satisfactory for the "customer" when performing the fusion process than that available without the fusion process. For example, better quality may be an increase in accuracy of a geophysical parameter or of a classification. It may also be related to the production of more relevant information of increased utility, or to the robustness in operational procedures. Fused information represents an entity in greater detail and with less uncertainty than what is obtainable from any of the individual sources. The fusion process can also extract higher order spatial, temporal and behavioral relationships between those entities. Greater quality may also mean a better coverage of the area of interest, or a better use of financial or human resources allotted to a project. In some cases, quality may be replaced by efficiency.

If compared to the JDL work, this definition does not propose any functional model or architecture. As we will see further, the architecture that can be drawn is very open, and consequently is of less practical value than the JDL model.

One immediate outcome of the definition is that now some aspects of data fusion, such as alignment or association are not all considered as stages of processing. For example, alignment becomes a property of the data to fuse, while association remains a processing issue. This is discussed in a further section.

TERMS OF REFERENCE

Other terms of reference are required to describe data fusion. Most of them exist in other domains and are of widely accepted use, or are the subject of standards, such as ISO, CEN, FGDC or OpenGIS. Examples of such terms are measurements, features, symbols, etc. The above-mentioned working group recommended their adoption in data fusion to avoid confusion and for the sake of the simplicity.

MERGING, COMBINATION, INTEGRATION, ASSIMILATION

The terms *merging* and *combination*, are used in a much broader sense than fusion, with combination being even broader than merging. These two terms define any process that implies a mathematical operation performed on at least two sets of information. These definitions are intentionally loose and offer space for various interpretations. Merging and combination are not

defined with an opposition to fusion. They are simply more general, also because we often need such terms to describe processes and methods in a general way, without entering details. *Integration* may play a similar role especially in system aspects; in information, it implicitly refers more to concatenation (*i.e.*, increasing the state vector) than to extraction of relevant information.

Another domain pertains to data fusion - *data assimilation* or optimal control. Data assimilation deals with the inclusion, or ingestion, of measured data into numerical models for the forecasting or analysis of the behavior of a system. A well-known example of a mathematical technique used in data assimilation is Kalman filtering. It is a technique that produces estimates of the state vector of the observed system. It is characterized by recursive evaluation, a model of the dynamics of the system and dynamic weighting of incoming observations. Data assimilation is daily used for weather forecasting.

MEASUREMENTS, SIGNAL, OBSERVATION

Terms such as measurements, attributes, rules or decisions, are often used in data fusion. These terms as well as others related to information are defined in the following. These definitions are those used in information sciences and optronics.

Measurements are primarily the output of a sensor. This is also called a *signal*, or image in the 2-D case. The elementary support of the measurement is a *pixel* in the case of an image, a *voxel* in the case of 3-D measurement and is called a *sample* in the general case. Bijective functions are often applied to a signal as a pre-processing prior to a fusion process. The result is usually considered as a signal. For example, the measurements made by optical sensors are digital numbers that can be converted into radiances once the calibration operations are performed. If corrections for the spectral irradiance of the illuminating source are applied, reflectances are obtained, which are still considered as a signal.

Commensurate sensors observe the same physical manifestation of an entity. Stereo-imagery uses a couple of similar sensors delivering commensurate data. Infrared sensors and UV sensors are usually not commensurate. An image of the roughness of the surface of a piece of metal acquired by a laser is commensurate to that acquired by radar. The same data are not commensurate to that acquired by an X-ray sensor because the latter image is depth-integrated.

Observation is a general word that denotes the raw information. For example, a verbal report is a piece of raw information, and may be considered as an observation. Measurements are observations. In information theory, an observation is also called signal when it triggers a

process, whatever it is. For example, when one sees an acquaintance, the sight of him is a signal that triggers recognition.

Terms relating to electronic imagery may be found in ISO documents[15]. In the case of image, the measurements are often called *gray levels*, whatever the type in computer encoding (character, integer, float, etc.). An image is called *multi-modality* if it is composed of several images acquired by a different modality. It is also called multi-channel, or multi-band, or multispectral in the case of optical sensors, or multi-frequencies in the case of microwave sensors. By extension, one may denote image any information that is presented under raster format, *i.e.* on a regular grid. Grid cell is equivalent to pixel. The information is often called *gridded information* (e.g., gridded temperature) or rasterized information. Actually, the term "gridded information" may denote information having more than 2 dimensions in space, contrary to images.

In this book, the term frequency is used indifferently for frequencies in time dimension (expressed in Hz) or for wavenumbers or wavevectors (expressed in m^{-1}), which belong to the space domain.

OBJECT, ATTRIBUTE, STATE VECTOR

An *object*, or *entity*, is defined by its properties, e.g., its color, its materials, its shapes, its neighborhood, etc. It can be a natural object (e.g., tree, mountain), a natural phenomenon (e.g., a cyclone), or a man-made object (e.g., engine, road) etc.

By extension, the support of a signal (e.g., a pixel) may be considered as an object. Some of its properties are defined by the set of observations that are attached to this support. For example, if a classification has been performed onto a multispectral image, the pixels belonging to the same class can be spatially aggregated. This results in a map of objects having a spatial extension of several pixels. Note that geographers limit objects as being single phenomena existing in the real world (e.g., river, street) [16] [17].

An *attribute* is a property of an object, which describes geometrical, topological, thematic or other characteristics. For example, the position and the velocity are two attributes of a vehicle. According to the problem, we

---

[15] ISO/FDIS 12651. *Electronic imaging - vocabulary*. 1999.

[16] CEN/TC 287. *Geographic information - Vocabulary*. CR 13346:1998, Comité Européen de Normalisation (CEN), 1998.

[17] FGDC. *Content standard for digital geospatial metadata. Annex A - glossary*. Federal Geographic Data Committee, c/o US Geological Survey, Reston, Va, USA. FGDC-STD-001-1998, 1998.

may add other attributes of this vehicle: its color, shape, sizes, number of seats, fuel consumption etc. *Feature* is equivalent to attribute. It should be noted that some standards in geography[18] [19] use features to denote an abstraction of the real world phenomena.

Attributes may be of measurements type or not. Examples of non-continuous attributes of a vehicle are mode of propulsion and type of fuel. Such information is non-continuous and is called *class*, *label*, *category*, *categorical data* or *taxon*. In another example, classification of multi-modalities images is often used as a fusion process. Outputs are classes and are attributes of the pixel. Another well-known example is the *spatial context* of a pixel, computed by local variance, or structure function or any spatial operator. This operation can be extended to *time context* in the case of time-series of measurements. Equivalent terms are local variability, local fluctuations, spatial or time texture, or pattern.

By extension, any information extracted from an image or any signal is an attribute for the object. The aggregation of measurements made for each of the elements of the object (for example, the pixels or samples constituting the object), such as the mean value, is an attribute. Some authors call *mathematical attribute* such attribute deriving from statistical operations on measurements.

The properties of an object constitute the *state vector* of this object. This state vector describes the object, preferably in a unique way. The state vector is also called feature vector, or attribute vector. The common property of the elements of the state vector is that they all describe the same object. If the object is a sample (e.g., a pixel), the state vector may contain the measurements as well as the attributes extracted from the processing of the measurements.

The sometimes-called positional state vector contains the position (measurements) and the velocity of a moving object. The velocity is either a measurement (e.g., by Doppler effect) or an attribute (e.g., first derivative of consecutive observations of position).

Another example of state vector is the color image, which is composed of three images (measurements): red, green, and blue. A more complex example is the meteorological state vector usually composed of assessments of the fields of pressure, temperature, humidity and wind at various

---

18 ISO/TC 211. *Geographic information / geomatics. Definitions.* ISO/TC 211 N038, 1996.

19 OpenGIS. *OpenGIS abstract specification*, The Open GIS Consortium (OGC), Wayland, Ma, USA, 1999.

altitudes. These assessments (attributes) derive from sophisticated processing of measurements.

RULES, DECISIONS, REPRESENTATION

Works in pattern recognition have drawn an analogy with the syntax of a language. Terms of higher semantic content have been defined, such as rules and decisions. *Rules*, like the syntax rules in language, define relationships between objects and their state vectors, and also between attributes of a same state vector. Rules may be state equations, or mathematical operations, or methods (that is a suite of operations, *i.e.* of elementary rules). They may be expressed in elaborated language. Examples of such rules are those used in artificial intelligence and expert systems. *Decisions* result from the application of rules on a set of rules, objects and state vectors.

A *representation* of the entity / object is the set of measurements, or attributes, or rules describing the object, completely or not. In principle, a representation consists in all the knowledge available about this object. A representation includes the state vector of the object together with the relevant rules.

For example, the representation of a fighter aircraft at instant *t* will comprise its position, velocity, past and possible trajectory, and additional knowledge not necessarily derived from the set of sensors observing at instant *t*, such as its type, typical mission, range of action, maximum speed, maneuverability, weapons, ammunitions, etc. In the case of forest fire fighting, the state vector will include the location of the fire front, its velocity, past and possible trajectory and intensity. The representation will also comprise information about the terrain (e.g., slopes, vegetation inflammability, and accessibility) and resources for fire combating.

SUB-DOMAINS IN DATA FUSION

Data fusion may be sub-divided into many domains. The sub-division may be made by functionalities or objectives of the fusion (e.g., tracking), by theme (e.g., medicine), by type of inputs to the fusion process (e.g., attributes), by class of architectures, by class of algorithm or mathematical tools (e.g., wavelet transform), etc.

For example, the military community uses the term *positional fusion* to denote aspects relevant to the assessment of the state vector or *identity fusion* when establishing the identity of the entities is at stake.

If observations are provided by sensors and only by sensors, one will use the term *sensor fusion* or *multisensor fusion*. *Image fusion* is a sub-class of sensor fusion; here the observations are images. If the support of the information is always a pixel, one may speak of *pixel fusion*. Other terms

easily understandable are *measurement fusion*, *signal fusion*, *features fusion*, and *decision fusion*. They mean that the fusion process deal only with respectively, measurements, signals, features, and decision.

Evidential fusion means that the algorithms behind call upon the evidence theory, *fuzzy fusion* denotes processes and algorithms using fuzzy logic, etc.

ALIGNMENT

The information entering a fusion process should be aligned. The alignment of the sources defines a common representation (X_S) on the basis of the measurements $(z_S)^t$, and the representations $(X_S)^t$ at instant *t*.

Differently said, a common co-ordinate system (e.g., geographical space and time) should be found wherein the sources data as well as the global knowledge can be represented. The data are said aligned, and the relevant operations are called alignment operations or alignment processes. This is called alignment, or conditioning, or sometimes harmonization.

For example, the geocoding of airborne or space-borne images is part of the alignment operation. Geocoding aims at providing an assessment of the absolute (or relative) geographical location of a pixel. Similarly mathematical techniques exist, which render two images of the same object superimposable, including a resampling for harmonizing the pixel sizes. The specific case of the geometrical alignment of images is discussed in Chapter 6.

Alignment provides a general frame of referencing that can applied to homogeneous (commensurate) as well as heterogeneous (non-commensurate) data. This is a difficult problem, and there is no general theory.

For example, assume a parallelepiped made of metal, observed by a laser and by X-rays. The common reference space of these two non-commensurate sources has three spatial dimensions. However the depth perception is ill defined because of the depth integration performed by the X-rays sensor and its sensitivity to heterogeneity in the piece of metal. Hence it is not easy to establish the 3-D space under concern. If one adds another source, non-commensurate to the two others, the problem is getting more complicated (e.g., electron microscope).

Alignment may request a conversion / transformation of observations. For example, it may be necessary to convert all data into optical paths in order to combine them. Alternatively, it may request an extraction of attributes, which may be the appropriate quantities to fuse, especially in the case of non-commensurate observations. Models may be necessary for aligning two sets of commensurate observations acquired on different supports, e.g., pin-point measurements integrated over a given period of time and

instantaneous measurements of the same entity / phenomenon integrated over a surface or within a volume.

Depending upon the case, corrections of changes in illumination of the object or in attenuation of the signal between the target and the sensor should be performed. This may occur in natural environment: a change in atmospheric constituents induces changes in light propagation, or in industrial environment: dust or paint aerosol may influence the illumination of the object to sense.

The concept of alignment is extended to a wider reference space (representation space). It includes the standardization of units in case of measurements, the calibration of sensors, the corrections of changes in illumination, the standardization in taxonomy if sources are attributes, or in syntax and lexicon if sources are rules, the selection of a common language for verbal or written reports etc.

Indeed alignment is part of the fusion process. It is sometimes considered as a pre-processing, but it should be stressed that this operation is solely performed in order that the information satisfies some constraints imposed by the objectives of the fusion process. That alignment is part of the fusion process may be hidden by the fact that information providers, including instruments makers, may supply data that are already aligned and ready for subsequent processing by customers.

ASSOCIATION

Let be two sources of information *S(1)* and *S(2)*. Each provides a representation, $(X_{S(1)})^t$ and $(X_{S(2)})^t$ at instant *t*. Let *S* be the union set of sources. Assume information is aligned for this set *S*. Associating the two representations $(X_{S(1)})^t$ and $(X_{S(2)})^t$ requires that they refer to the same object. There is no benefit of trying to fuse measurements, attributes, rules or representations that do not refer to the same phenomenon or entity.

The union of the representations is called *association* or concatenation. Association is made by an increase of the size of the state vector of the object.

Association is independent of the semantic level of the information. It is performed by an analysis of the degree of correlation / relation between the information to be fused and the entity under concern.

Some examples have been given previously, where sources are not exactly referring to the same object. In that case, though the sources are aligned, the representations cannot be associated. It can be a matter of period of observation for example. If one is observing a moving target within a limited period, any information relating to instants well before and well after is poorly correlated to the dynamics of the target.

On the contrary, quasi-simultaneous observations of different or same parts of the same human spine by X-rays scanner and nuclear-magnetic resonance imaging system refer to the same object. Once aligned for units (or gray level dynamics) and geometric superimposability, these observations can be fused for, e.g., a 3-D reconstruction of the spine, possibly given some additional knowledge.

Data concatenation is accomplished by juxtaposing all the data into the state vector, hence augmenting it. A straightforward example is given by a time-series of images from the meteorological geostationnary satellites, which are taking a picture of the Earth every half-hour or less. The raw data are processed by the meteorological offices, and are spatially superimposable once delivered to the customer. In that case, at each pixel, one can define a state vector by the concatenation of all the observations made at this pixel in the period under concern. Because the data provider has performed the alignment of data, the customer deals in this case with concatenation and subsequent analysis.

In some cases, the issue of association can be the selection of sub-sets of sensors, which are the most relevant for a given problem. A metric should then be defined for the comparison between sensors, and the choice of the most appropriate ones.

TOPOLOGICAL AND PROCESSING ISSUES

A fusion system can be a very complicated system. It is composed of sources of information, of means of acquisition of this information, of communications for the exchange of information, of intelligence to process the information and to issue information of higher content.

The issues involved may be separated in topological and processing issues. Despite the interconnection between both issues in an integrated fusion system design, they can be decoupled from each other in order to facilitate the development of a systematic methodology of analysis and synthesis of a fusion system according to Thomopoulos[20] [21]. Recent advances in technology and in the modeling of complex systems may render this separation useless or unrealistic (e.g., UML language).

---

[20] S. C. A. Thomopoulos. *Sensor integration and data fusion*. Journal of Robotic Systems, vol. 7, pp. 337-372, 1990.

[21] S. C. A. Thomopoulos. *Decision and evidence fusion in sensor integration*. In Advances in Control and Dynamic Systems, Ed. C. T. Leondes, vol. 49, part 5, pp. 339-412, Academic Press, 1991.

The *topological* issues address the problem of the spatial distribution of sensors, the communication network between sensors and places of processing and decision-making, the bandwidth and the global architecture. Also at stake are issues for the exchange of information, the availability and reliability of information at the time of the fusion. The cost of acquiring the information may also be relevant to the topological issues. In non-military applications, these issues are partly addressed by the vendors / distributors of information. They are also partly addressed by the customer, given its objectives and constraints, including the financial budget.

The *processing* issues address the question of how to fuse the data, *i.e.* select the proper measurements, determine the relevance of the data to the objectives, select the fusion methods and architectures, once the data are available, and according to the specifications issued by the project under concern.

There is no specific processing general techniques in data fusion. All mathematical tools may apply. Hall proposed taxonomy of algorithms for sensor fusion[22]. The first category of techniques deals with the positional fusion, *i.e.* the assessment of the state vector from the observations. The second category, called identity fusion, seeks to combine data to establish the identity of an entity. The third category includes ancillary techniques to support the processing in the level 1 of the JDL model. This taxonomy is not efficient. Hall himself wrote that positional and identity fusions may occur in a simultaneous or interleaved fashion, using similar algorithms.

In military applications, three stages of processing often appear, which may perform independent of the level of information being fused[23]. Correlation (first stage) applies a metric to each of the redundant parameters on which association is dependent to measure the degree to which that data is related, or associated, to an entity (e.g. a target track). If these parameters cannot be obtained from the source data then there is no way to fuse that data with the entity. Association (second stage) combines all of the correlations together and thresholds the result to decide if association exists between the source data and an entity. If they are associated, then the combination stage occurs. Combination (third stage) estimates the new state of an entity. It may use intermediate results from the preceding stages, particularly correlation, by aggregating and then merging the parameters. The multiple values for each redundant parameter are aggregated to form the single new updated value of that parameter. This results in a set of complementary parameters, which are then merged into the one unified representation of that entity.

---

[22] D. Hall. *Op. cit.*

[23] DSTO. *Op. cit.*

TYPOLOGY OF PROBLEMS IN DATA FUSION

In the literature, data fusion is often split into two categories of problems: "low-level" and "high-level", or three categories: "measurement level", "feature level" and "decision level"[24], which correspond to the semantic levels of the inputs and inferences.

This presentation may have at times practical advantages. However, it is not fully sustained by the concept of data fusion and should be avoided as much as possible to avoid confusion. Implicitly, this presentation implies that the concept contains a built-in hierarchy based upon the semantic level of the inputs and that of the inferences. This is not true at all. *Fusion may operate at any of the various semantic levels*, with *possible crossings* between levels.

This property is not fully expressed in the literature using the JDL model, as already discussed. (Beware not to confuse the semantic levels and the levels of the JDL model.) Together with the clear expression of the typology of problems, expressed below, this property impacts on the design of the architecture of a fusion system, on the selection of tools, suite of softwares and hardware (processing issues), communications (topological issues) and on the design of innovative procedures.

Three types of problems in data fusion are identified[25]. They clearly state that crossings between semantic levels are possible and frequent. Actually, the three semantic levels cited above are not the most appropriate to describe the fusion processes. Attributes, analyses and representations should be preferred.

FUSION OF ATTRIBUTES

Assume the sources of information are aligned and associated. Fusion of attributes consists in merging the attributes of a same object, derived from two representations $(X_{S(1)})^t$ and $(X_{S(2)})^t$ at instant t obtained by means of the sources of information $S(1)$ and $S(2)$, in order to obtain new attributes in the space of sources $S = S(1) \dot{E} S(2)$.

This is the case of classification processes that are performed on measurements (attributes) obtained by two different modalities or more, observing the same ensemble of objects (e.g., a piece of metal or a landscape). These fusion processes provide new attributes for these objects

---

[24] J. A. Benediktsson and D. A. Landgrebe. *Introduction to the special issue on data fusion*. IEEE Transactions on Geoscience and Remote Sensing, 37(3), pp. 1187, 1999.

[25] L. F. Pau. *Op. cit.*

(e.g., cracks or landuse category). Actually, representations may be obtained at various instants. Here, same instant *t* means that the time lag between the representations $(X_{S(1)})^t$ and $(X_{S(2)})^t$ is small enough with respect to the time scale of change of the attributes to fuse and of those resulting from the fusion process.

FUSION OF ANALYSIS

Assume the sources of information are aligned and associated. Fusion of analysis consists in aggregating representations $(X_{S(1)})^{t1}$ and $(X_{S(2)})^{t2}$, with $t2 > t1$, into a new representation $(X_S)^{t3}$, with $t3 \hat{I} [t1, t2]$, then in generating an analysis or interpretation of the object for further use at instant $t4$, with $t4 > t2$, or at a further step in an iterative process.

In the simpler case, the instants *t1*, *t2* and *t3* are identical. A typical case is that of a mobile target, co-operative or not, that is observed by two sensors *S(1)* and *S(2)* at the same instants. From each of the representations $(X_{S(1)})^t$ and $(X_{S(2)})^t$, an analysis can be performed on the trajectory and velocity. Prediction of these parameters at instant *(t+1)* can be made. Fusing the attributes provide a new representation, from which a new analysis can be generated for use at instant *(t+1)*. Kalman filtering is one of the well known tools for such cases.

In most real situations, the sources of information are asynchronous and representations are not available at the same time.

FUSION OF REPRESENTATIONS

Fusion of representations is defining and performing meta-operations applicable to representations $(X_{S(1)})^t$ and $(X_{S(2)})^t$ to obtain a new representation $(X_S)^t$. Fusion of representations includes fusion of decisions. This fusion of representations may be performed at any moment, *i.e.* combined with other types of fusion.

4. REPRESENTING A FUSION PROCESS - ARCHITECTURES

REPRESENTING A FUSION PROCESS

A fusion system is above all a system. As such, it obeys the general theory of systems. This Chapter does not discuss this theory. It focuses on the representation of a fusion process by simple schemes and architectures. These schemes should convey the major specific aspects of the fusion process. They are very useful, especially in education and training. Usually expressed in the form of graphics, they greatly help in expressing and understanding the fusion process.

According to Bass *et al.*[1], and in line with the standard ISO/IEC under construction[2], the architecture of an application is the structure, or the structures of the system, which comprises the components, the main externally visible properties of the components and the key relationships among those components. The architecture tells what happens and thus can be seen from various viewpoints. Hence it depends upon the interest of the reader, whether he is a computer man or a manager etc. The architecture encompasses our traditional understanding of blocks connected by data, communications, control or other type of links.

Adopting a common scheme / architecture for fusion process offers several advantages. Among others, it permits better understanding and foster co-operation between people because the language is the same; it supports analysis by capturing domain and knowledge and community consensus. Though architecture can be conceived without specific regard to a particular form of physical representation, such a representation is instrumental to make the architecture understandable to others. The representation should be as independent as possible from any application or technological solutions.

The JDL functional model may help in setting a scheme and architecture. It is well adapted to any situation of crisis, military or not. However not all applications in data fusion deal with the whole system, as the military people do.

---

[1] Bass, Clements and Kazman. *Software architecture in practice*. Addison-Wesley, 1998.

[2] ISO/IEC 10746, *Open distribution standard*. 10746-4: Architectural semantics.

In many applications, the work and responsibilities are shared, the data and services are bought and the whole system is not mastered by a single entity. Many applications only deal with fusion of decisions, and not with sensors. Others are interested by positional fusion, *i.e.* the assessment of the state vector (typically Level 1 in the JDL model). The JDL model is often too complex. A simpler scheme for sketching a fusion process will be more appropriate. This is also true for applications, in which fusion is performed at higher semantic levels.

Several other schemes have been proposed. In one scheme, pixel-level, attribute-level (or feature-level) and decision-level are used to describe the fusion process[3]. In pixel-level fusion, the data are combined at the pixel level of the sensors. In the attribute-level fusion, the features are extracted from each sensor data and combined. In the decision-level fusion, the fusion is performed on decisions. In a very similar scheme, the taxonomy is low-level fusion, middle-level fusion and high-level fusion, with reference to the semantic content of the information input in the fusion process.

The first scheme presents a first drawback. The pixel is only a support of information and has no semantic significance; measurements or observations would be more appropriate. The scheme can then be generalized to non-imaging sensors.

But the major drawback of these two schemes is that they are misleading: they do not consider fusion processes dealing simultaneously with these different semantic levels. The various natures of the information to be fused have already been emphasized. Many examples can be found in any domain, where information of various semantic levels is fused. The property of data fusion called "fusion of representations" also stresses that fusion can operate at the three different levels with possible mixing. Hence such schemes should be avoided. This book proposes a more general scheme.

THE FUSION CELL. SOME EXAMPLES

We have selected a scheme that integrates any of these levels. Houzelle, Giraudon[4] proposed a scheme that consists in a fusion cell for fusion of decisions. This scheme may be easily adapted to any input, and it allows all semantic levels (measurements, attributes, and decisions) to be simultaneous inputs to a fusion operation (Fig. 4.1).

---

[3] L. A. Klein. *Sensor and data fusion concepts and applications*. Tutorial texts, vol. TT 14, SPIE Optical Engineering Press, USA, 131 p., 1993.

[4] S. Houzelle and G. Giraudon. *Contribution to multisensor fusion formalization*. Robotics and Autonomous Systems, vol. 13, pp. 69-85, 1994.

This scheme considers three types of inputs: sources of information to be fused, auxiliary information, and external knowledge. Any fusion operation can be described by the means of this fusion cell. Actually, this cell may represent from simple to very complex operations. It can be combined with others to sketch combined fusion processes, as shown in the following.

Sources of information are the main inputs to the fusion cell. They are the inputs of the mathematical operations included in the fusion cell. They should be aligned. These inputs can be outputs of sensor: images or any other signal, and more generally measurements. They can also be attributes or decisions.

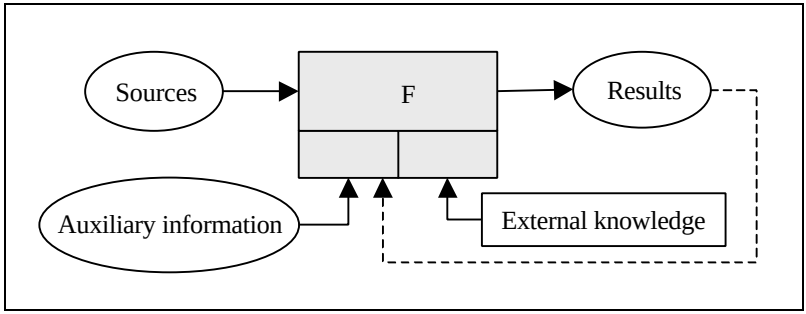


Figure 4.1. Representation of a fusion operation by a fusion cell

The auxiliary information brings additional information, resulting from the specific processing of a source, or from another fusion operation. In iterative processing, including time-dependent operations, the results may become inputs to the fusion operation in a subsequent step or instant. They will act as auxiliary information, since they are not original sources.

External knowledge is also additional information, whose objective is mainly to constrain or guide the fusion process by e.g., imposing *a priori* knowledge. *A priori* means that the knowledge is available prior to the fusion process. It can be made of process laws of mathematical foundation, or empirical laws that can be expressed or not in quantitative form. As an example, rules for decision processes (e.g., expert system, neural networks, and fuzzy logic) are such an external knowledge.

EXAMPLE. FUSION IN INDUSTRIAL PROCESSES

In this example (Fig. 4.2), several sensors *sensor 1*, *sensor 2*... *sensor n* are monitoring an industrial process. For example, these sensors can be distributed in space to monitor a plant or an airplane. They may measure similar quantities at various places. In other cases, they may be close in space and measure different quantities. The observations or measurements of these *n* sensors are fused in the cell *F*. The procedure obeys process laws,

which are inputs as external knowledge, in order to guide or constrain the fusion. The outputs may be attributes (e.g., effective attitude of a ship) or decisions (e.g., increase the speed of a conveyor belt).

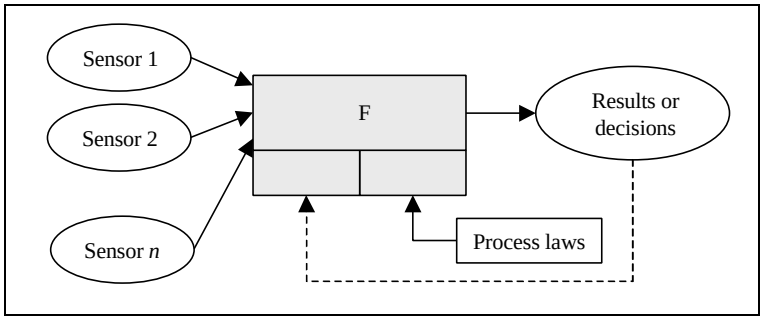


Figure 4.2. Scheme for an industrial process

The process laws may take into account the history of each measurement or result. In that case, the history becomes input as external knowledge by modifying the process laws. In other cases, this history is an input of the fusion cell as an auxiliary information.

The scheme in Figure 4.2 may represent the monitoring of an object, e.g., the trajectory of a rocket. Here, the outputs of the fusion cell at instant *t* become inputs for the following instant *t+1*, as auxiliary information (dotted lines, Fig. 4.2). In that case, the fusion cell may consist in a Kalman filter. The process laws contain the model of the dynamics of the observed system. At each time, the state vector at time *t+1* is related to the measurements at time *t* (input measurements) and the state vector at time *t* (auxiliary information).

The engine of modern vehicles works along this scheme (Fig. 4.3). The process laws are called engine cartography. This cartography has been established from fundamental parameters (usually rotation and load). Given these parameters, it returns the optimal ignition angle. The sensors measure e.g., temperature, pressure and flow in different places. These measurements are fused taking into account the rules given by the engine cartography. The result of the fusion process is e.g., the quantity of gas to be injected into the combustion chamber. In more sophisticated engines, the history of the results is often injected as an auxiliary information.

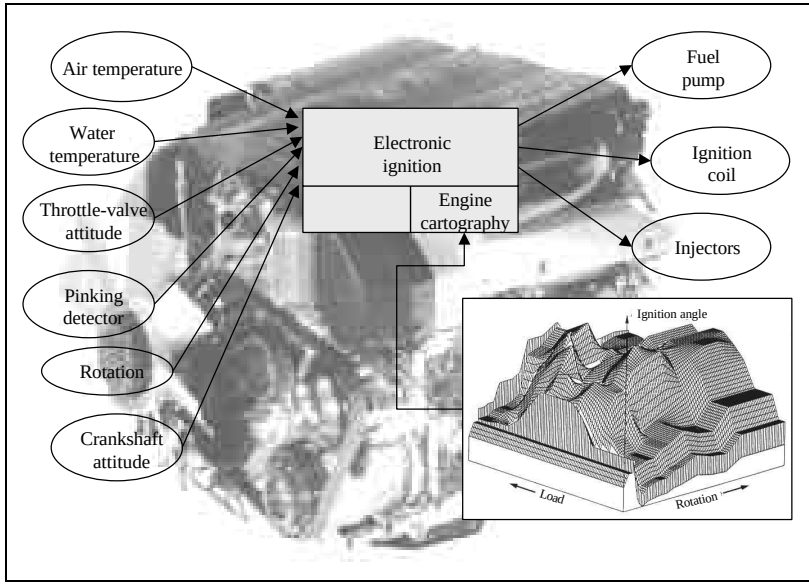


Figure 4.3. Example of a scheme describing a data fusion process: the engine of modern cars

EXAMPLE. MAPPING

Mapping the Earth from satellite images is another example of data fusion (Fig. 4.4). Several images of different nature (optical, radar), are inputs to the fusion cell. The fusion method is usually a classifier, and the outputs are maps of classes and of confidence level.

In this example, an image with a high spatial resolution is merged with multispectral optical image of lower spatial resolution and with radar images. From the image of best spatial resolution, an image of texture is extracted to help in classifying the original measurements. Though it will enter the classifier with the same weight than the other sources, it may be considered as an auxiliary information because it is derived from the original sources. The available geographical information is contained in a geographical information system and is a valuable input to the fusion process. The codebook for classification is give as annex knowledge as it is the case in a supervised classification process.

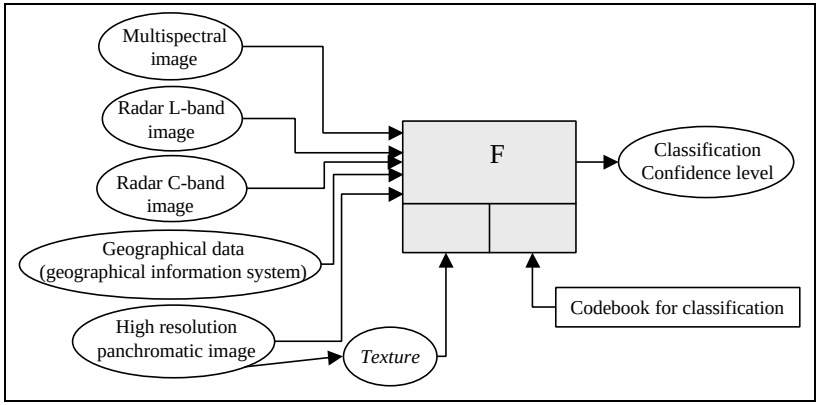


Figure 4.4. Typical scheme for the mapping of landscape using Earth observation satellite images

EXAMPLE. MAPPING BY FUSING SATELLITE IMAGES AND GROUND MEASUREMENTS

Beyer *et al.*[5] fused digital maps in raster format with measurements made at meteorological stations to construct the final raster maps of the solar radiation over Europe[6] with a pixel size of 5' of arc angle (approximately 10 km at latitude 45°). These maps were derived from the processing of images acquired by meteorological satellites. The site measurements are scarce in space but are more accurate than the satellite-derived maps. The latter offer a good description of the spatial distribution of the solar radiation, more accurate than what can be achieved by interpolation techniques.

In order to provide accurate maps the satellite-derived maps and the site measurements were fused as shown in Figure 4.5. After alignment for geographical absolute location, units, sampling supports in space and time, trends with seasons and latitude were removed from both types of data. The residuals are mostly isotropic and may be considered as random variables. Then a co-kriging is performed on the residuals, the distance being a function of the geographical location and of the elevation. This function is an external knowledge; the elevation of each pixel of the maps is an input to

---

[5] H.-G. Beyer, G. Czeplak, U. Terzenbach and L. Wald. *Assessment of the method used to construct clearness index maps for the new European solar radiation atlas (ESRA)*. Solar Energy, 61, 6, 389-397, 1997.

[6] *European solar radiation atlas*. Fourth edition, includ. CD-ROM. J. Greif, K. Scharmer. Published for the Commission of the European Communities by Presses de l'École des Mines de Paris, France, 2000.

the fusion process. Once the residuals interpolated, the trends are re-injected and the final maps are obtained.

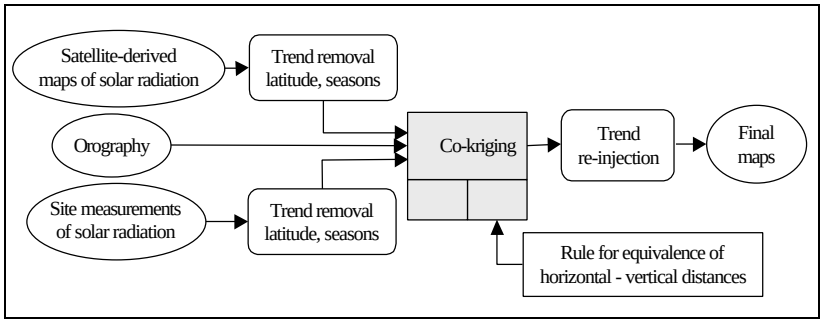


Figure 4.5. The fusion of raster maps and site measurements for the European solar radiation atlas

This example shows how the proposed scheme may deal with sources having different supports of information.

EXAMPLE. COMPRESSION OF INFORMATION

The following example is purely of academic interest; it does not describe any current technique in compression of information Let assume that color images should be transmitted. The three channels composing an image are called *R* (red), *G* (green), and *B* (blue). The images are originally coded in 24-bit (3 times 8 bits). Compression should be performed on these color images in order to decrease the necessary bandwidth. The compression should be applied before transmission and compressed color images are coded in 8 bits. The compression / re-coding algorithm calls upon rules, which are fixed but changes should be brought if necessary.

The algorithm should also respect the main contours and some of the colored transitions. Accordingly, the three channels *R*, *G*, *B* are converted into the intensity, hue, saturation space (*I*, *H*, *S*). The intensity *I* reveals the structures and contours of the objects, while the saturation *S* is assumed to reveal the colored transitions. A quantity *Q* is defined as follows

Q = R-G if the saturation S is greater than S_0

Q = R-B otherwise

The threshold S_0 is fixed but changes should be brought if necessary.

An index *ID* is defined that relates to the structures. Its is an input to the compression algorithm. This index is a mathematical combination of the wavelet coefficients (*C1*, *C2*) and of *Q*. The wavelet coefficients are obtained by two iterations of a wavelet transform WT applied to the

intensity *I* of the images; they identify the contours in the intensity *I* on a multi-scale basis.

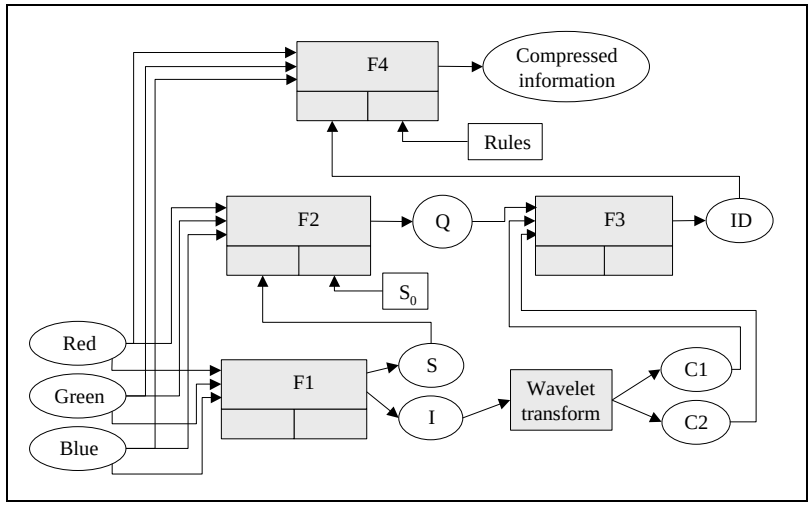


Figure 4.6. A scheme representing a data fusion process for data compression

According to these specifications, one may draw the architecture of the process (Fig. 4.6). The cell F1 performs the fusion of the three bands *R*, *G* and *B* and converts them in intensity *I* and saturation *S*. The cell F2 provides the quantity *Q*, which is combined with the wavelet coefficients in the cell F3, which results into the index *ID*. Finally this index enters the cell F4 together with the three bands *R*, *G* and *B*, and the compression is performed according to fixed rules.

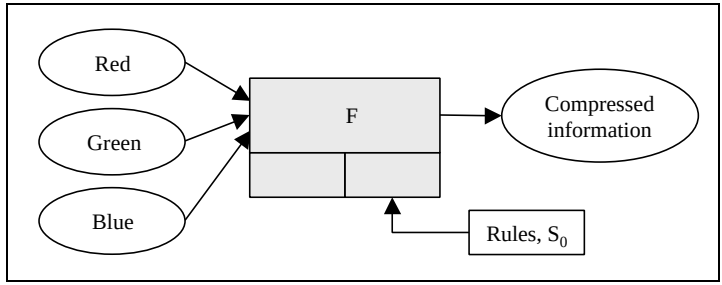


Figure 4.7. Another scheme for the same operation for compression

Figure 4.6 is a detailed presentation of the fusion process. It comprises several elementary fusion cells and helps in better understanding the

relationships between the data and the processes. For the same example, the fusion system may also be represented using a more condensed scheme, as in Figure 4.7.

This series of examples illustrates the large domain of applications of fusion. It shows the capabilities of the selected fusion scheme to represent fusion processes. Such a scheme does not replace the detailed descriptions that are usually requested to create a fusion system. It conveys the essential features of the fusion process and in this respect, is of great help to better understand the fusion process.

ARCHITECTURES

Fusion architecture describes the set of sources of information, how they are assembled, and how they are used, together with mathematical techniques and processing, in order to perform a fusion operation. This section does not intent to provide guidelines for implementation. The variety of data fusion applications is so wide and the implementation environments are so diverse that it is impossible to set up a blueprint for implementation. This section addresses the basis for understanding the key issues and problems in implementation.

The choice of architectures is not arbitrary. It depends on the nature of the information involved and the nature of the inferences sought. Usually three types of architectures are defined: centralized, decentralized and hybrid.

Centralized architecture may also be termed central-level fusion, or central fusion processing. Decentralized architecture is sometimes called autonomous architecture or, in case of sensor fusion, post-individual sensor processing fusion or sensor-level fusion. The term "distributed architecture" is ambiguous. It may refer to the case where pre-processing is performed at each sensor, before entering a centralized process, which still is a centralized architecture; it is also called distributed (federated) architecture or Level 1 processing if the JDL model is used. But it may also be equivalent to hybrid architecture.

CENTRALIZED ARCHITECTURE

The centralized architecture exploits in a single location, simultaneously or not, the set of data acquired by the set of sources of information (Fig. 4.8).

In this Figure, *Si* are the *n* sources. A source can be a set of measurements, attributes or decisions. All sources are inputs to the single fusion cell. The results *R* and quality parameters *Q* are obtained by the processing of all sources available at that moment. Of course, this architecture may include auxiliary information and external knowledge.

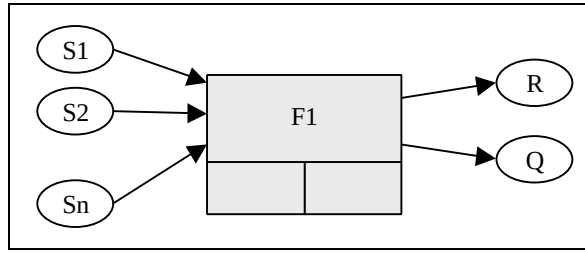


Figure 4.8. Centralized architecture. Si are the sources, R and Q the results and quality parameters

Stereo-photogrammetry has many applications, ranging from the study of manufactured objects to the monitoring of quarries and mapping of buildings. If such sensors are on board a satellite, or if the same spaceborne camera observes the same area under two different angles during e.g., two different orbits, the relief of the Earth can be reconstructed. Stereo-photogrammetry uses centralized architecture. Using two cameras observing the same object or area with different angles ("left" and "right" images), the relief of an object can be constructed (Fig. 4.9).

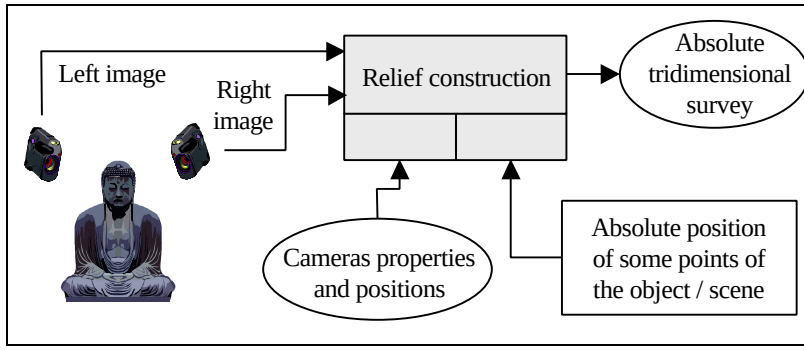


Figure 4.9. Centralized architecture for a stereo-photogrammetry process

The advantage of the centralized architecture is that theoretically it provides an optimal result, since the decision is made taking into account the whole knowledge available. The loss of information is minimal since the original information is fused directly without approximations *via* attributes, state vectors etc. The representation of objects, and further their discrimination, is more effective if the informations to fuse are not generated by independent phenomena.

However, if a particular source has a large error rate or a low signal-to-noise ratio in the case of a sensor, and depending upon the fusion technique, it may happen that this source contaminates the whole data set, and leads to a

decrease of the quality of the decision, compared to what would have been achieved without it.

Let use again the example of the human vision, which is a centralized fusion process. If one eye is deficient (e.g., very blurred vision), the other will ensure the vision in a degraded mode and the whole system will try to correct the vision of the other eye by calling much on the functioning one. It follows a great strain on the functioning eye, which will become rapidly tired. In this particular case, the performances in vision at the end of the day is finally worse than that attained by a single eye.

In Earth observation, such cases may be encountered as e.g., with imaging radar whose image quality is a function of various parameters, such as the rainfall before the instant of acquisition, or the surface state of the bodies of water. In most cases, using radar images, as inputs to a fusion operation will be highly profitable. In some cases, it may decrease the quality of the result. For example, if the wind is strong enough, it has been observed that rice fields cannot be perceived at certain growth states, because the clutter due to the wavelets make them confusing with other objects in the landscape. It is then more profitable to adopt another architecture.

The centralized architecture has some drawbacks with respect to processing. It requires all the data to be present on the processing site, which implies a large communications bandwidth. It also imposes a heavy processing load on the computer, which renews at any change of input.

DECENTRALIZED ARCHITECTURE

The decentralized architecture offers a large flexibility and modularity, and is often adopted for these reasons. It is also called autonomous because it involves independent processing of each source of information (or group of sources) until the fusion of some representation of higher semantic level takes place at a later stage (Fig. 4.10).

It should be selected when communication problems are at stake: small bandwidth, unsecured communications, which may be broken, etc. If the acquisition rate of information (sources) is very different between all sources, it may also be adopted to avoid re-processing all the sources while a few have changed, which is the case in the centralized scheme. The decentralized architecture will be adopted in risky domains, such as a battlefield or industrial hazards.

Each source *Si* enters a fusion cell, which may also include auxiliary information and external knowledge. As said before, a source *Si* is a set of inputs, which are composed of measurements, attributes, and / or decisions. The outputs of the local fusion cells (*F1, F2... Fn*) are results *Ri* and quality parameters *Qi*. These results and quality parameters are transmitted to the final fusion cell *F*. The results *Ri* are the inputs to this process. The quality

parameters *Qi* are auxiliary information and will help in deciding the weight of a source in the final process.

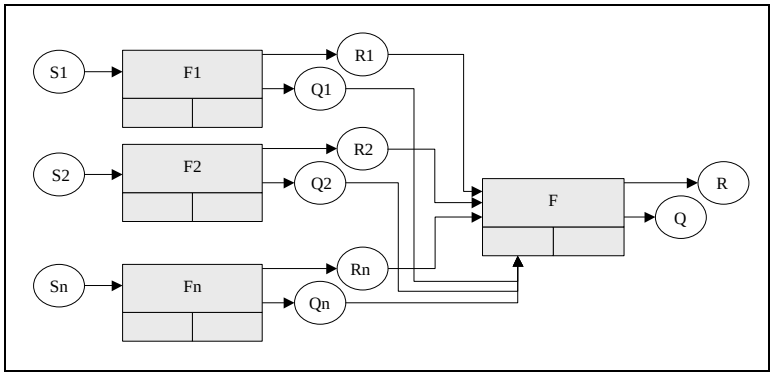


Figure 4.10. Decentralized architecture. After Mangolini[7]

One may note that each fusion process *Fi* is performed locally, using local intelligence. The fusion processing usually reduces the amount of information to be transmitted to the final fusion process. This accommodates for low communications bandwidth.

One may also note that this scheme is more robust to the loss of a source of information than the centralized scheme. From a practical point of view, it is easy with such architecture to remove, or not to take into account, a sensor whose confidence is questionable. It is much more difficult with a centralized architecture. In the case of strongly asynchronous information acquisition, i.e. very different time sampling of information from each source, the decentralized architecture gathers the locally fused information at the final central point, and thus does not need to renew the whole process at each acquisition time of the most rapid source.

The sources are processed independently from the others. Accordingly the results locally available *Ri* have a fairly low information content, depending upon the sources. It further results in the fact that the final result *R* has a lower quality and a lower information content than that would have been achieved with a centralized architecture.

---

[7] M. Mangolini. *Apport de la fusion d'images satellitaires multicapteurs au niveau pixel en télédétection et photo-interprétation*. Thèse de Doctorat, Université Nice - Sophia Antipolis, France, 174 p., 1994.

Figure 4.11 exhibits a typical decentralized architecture in the case of the management of a humanitarian crisis by an international organization, such as those that have been experienced in the past few years during civil wars.

In location X, all data and procedures are available for providing basic digital maps and geographical data that are needed. Inputs are mostly archived information in order to speed up the processing. This location X is usually far from the operation terrain. Besides providing relief to refugees, field teams (locations Z1, … Zj) collect information to better understand the present situation in several aspects, and send reports to the headquarters (location Y in the graph). There, a system processes all the data gathered, plus recent images acquired by spaceborne or airborne systems, to provide the best available situation plan for decision making.

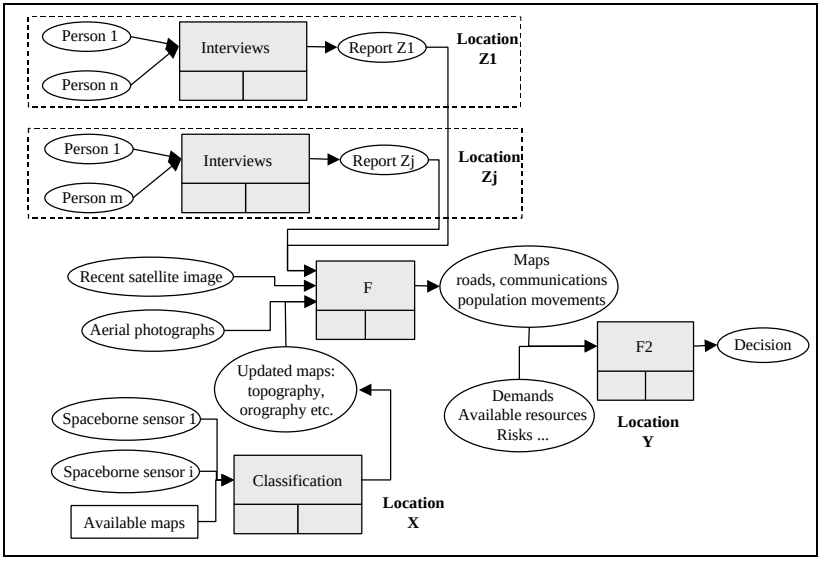


Figure 4.11. Typical decentralized architecture in a case of humanitarian crisis

In this case, the decentralized architecture is highly recommended for many reasons: communications lines may degrade suddenly, the number of sources of information varies, the quality of the collected information is highly variable and may necessitate interpretation and qualification by a skilled person, who should be located where necessary (e.g., location Z1 or X), the airborne surveys may be subjected to administrative / belligerents authorization, security factors and meteorology, the satellite imagery collection may be impeded by cloud coverage or availability of systems, etc.

HYBRID ARCHITECTURE

Other architectures may be designed that are a combination of centralized and decentralized architectures. They are called hybrid architectures and have various forms (Fig. 4.12).

In this Figure, the sources *S1*, *S2*... *Sn* are separated in two sub-sets: *S1*, ... *Si*, and *Sj*, ... *Sn* with possible overlaps. Each sub-set enters a fusion process having a centralized architecture. The results *R1* and *R2* are the sources of a final fusion process F, with the quality parameters *Q1* and *Q2* as auxiliary information.

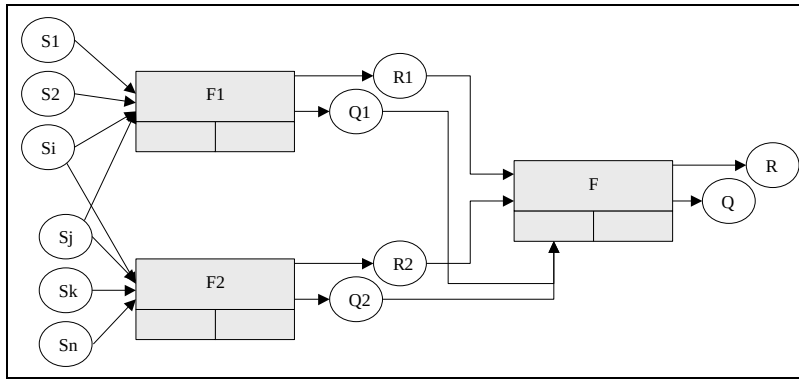


Figure 4.12. Hybrid architecture. After Mangolini (op. cit.)

Such architectures involve fusion of the sources at different semantic levels and at different processing stages. Depending upon the combination, such architecture is more or less close to a centralized or decentralized architecture, and so are its properties (advantages and drawbacks).

Weather forecasting, as discussed in the introduction, is an example of a system with hybrid architecture for fusion processes.

Another example is offered in the fusion of categorical data and observations (Fig. 4.13). Benthic communities denote the communities that are living on the sea floor close to the shoreline. They are mapped by means of various sampling techniques, which lead to different maps. Furthermore, there is a lack of standardized codes (categories) for depicting the communities. In addition, changes in communities, their size and location occurred in time. Consequently, maps of a same area are partly conflicting and partly in agreement. Fusion was used to reconcile this suite of maps and provide a synthesis map of categorical data for a given area together with a

series of maps reporting where conflict occurs, its nature and its magnitude[8].

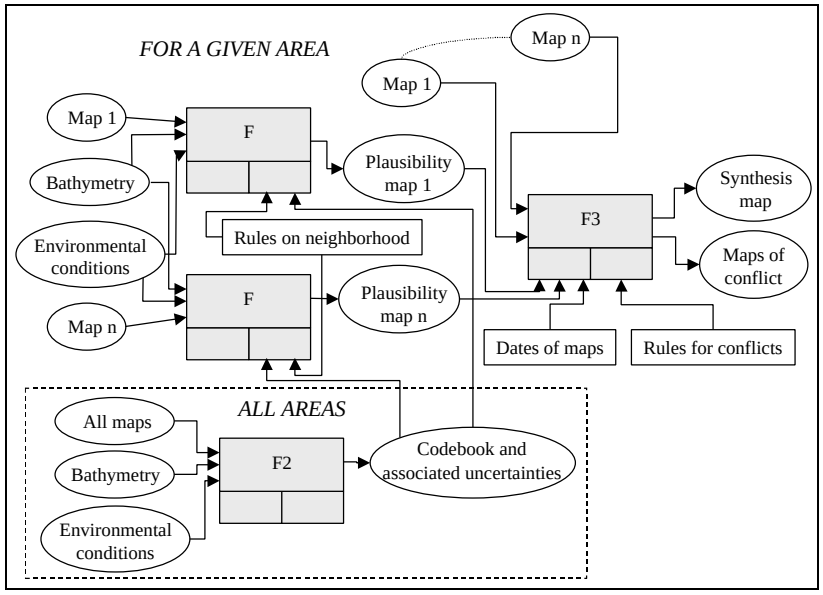


Figure 4.13. The fusion of a suite of maps differing in content for the construction of a synthesis map

Hybrid architecture was selected and the fusion is a three-stage process. The final fusion process is based upon fuzzy logic. The inputs are the original maps of communities; plausibility maps are input as auxiliary information as well as the dates of the maps. External knowledge of the possible conflicts helps in control the construction of the synthesis map. The plausibility maps are estimated from the previous stage of fusion, performed for each map independently. Each map is merged with other information of measurement type: the bathymetry, the state of turbulence of the sea, the currents and the quality of the water[9]. External knowledge is composed of a

---

[8] R. Méaille and L. Wald. *A Geographical Information System for some Mediterranean benthic communities*. International Journal of Geographical Information System, 4, 1, 79-86, 1990.

[9] A. Iehle, L. Wald and C.-F. Boudouresque. *Analyse et évaluation de la fiabilité de l'information dans le système d'information géographique des assemblages benthiques méditerranéens «MBA»*. Scientific Reports of the Port-Cros National Park, 16, 93-113, 1995.

codebook and its associated uncertainties and of some rules on the possible reciprocal neighborhood of the communities, the plausibility of each pixel of each map is constructed. The codebook is built in an initial stage. All maps of all areas and all geophysical information of influence: bathymetry, state of turbulence of the sea, currents and water quality, are fused through a classifier. A state vector of these geophysical measurements is associated to each category, together with the uncertainties of this classification.

SELECTION OF AN ARCHITECTURE

Centralized architecture should be preferred whenever possible because it provides the higher accuracy of the fused product. However, each architecture has advantages and drawbacks. Architecture should be selected on a case by case basis. Trade-off involve many factors[10], including the availability of smart sensors that perform data preprocessing, the availability of communications links and their bandwidth, and the computational abilities of the central processor / decentralized processors.

The application determines the phenomena observed, the type of sensors or of information utilized, and the inferences sought. These inferences in turn determine the types of techniques requested. Deployment and implementation constraints provide significant requirements that should be taken into account. Finally the capabilities of a system or suite for the acquisition or collection of data or information and the communication links (bandwidth, security, etc.) between the various elements also affect architecture selection.

Data fusion contributes to improved information accuracy, timeliness and content. Several major works have been performed to test and evaluate implemented, or modeled, data fusion systems, and to determine their contribution to the effectiveness in military or civilian applications. A hierarchy of measures is available in the military domain that relates performance characteristics of C3I systems to military effectiveness[11]. Dimensional parameters are the typical properties or characteristics that directly define the elements of the fusion system. Measures of performances (MOPs) are the measures that describe the behavioral aspects of the system and how well a fusion system performs. Measures of force effectiveness (MOFEs) quantify the ability of the total military force to complete its mission. It attempts to measure how well a fusion system, which is part of the total military force, satisfies an intended mission.

---

[10] D. Hall. *Mathematical techniques in multisensor data fusion*. Artech House, Boston, London, 1992.

[11] E. Waltz and J. Llinas. *Multisensor data fusion*. Artech House, 1990.

PART 2.

SOME TECHNIQUES IN FUSION OF IMAGES

5. SOME MATHEMATICAL TOOLS FOR THE FUSION OF IMAGES

Many popular or advanced techniques for the fusion of images share similar mathematical tools. Some of them are presented in this Chapter. Following Chapters illustrate how these tools may be used.

In this Part of the book, spectral means color, and more generally, electromagnetic radiation wavelength. Hence a spectral band is a portion of the electromagnetic spectrum. A multispectral image is composed of several images, each being acquired in a different spectral band. A spectral image may also be termed channel. More general terms are modality or mode instead of spectral band, and multi-modality image. A spectral signature (modality signature) is a state vector made of the values taken in each spectral band (modality).

CONVERSION RGB - IHS

THE COLOR SPACE

Color may be seen as a fusion process. Familiar objects, such as TV screens, PC monitors, color films, etc perform this process. Discussions of color usually involve three dimensions, known as *hue*, *saturation*, and *brightness*, as a descriptive tool[1]. Hue distinguishes between colors such as red, yellow, blue, etc. Saturation refers to purity, i.e. how the color is diluted by white light; it determines how pastel or strong a color appears, and distinguishes pink from red, sky blue from royal blue, etc. Brightness is equivalent to the intensity of the achromatic light; it is independent of hue and saturation. Artists may use another approach, specifying color as different *tints*, *shades*, and *tones* of strongly saturated, or pure, pigments.

Three primary colors (**X**, **Y**, **Z**) have been defined by the Commission Internationale pour l'Éclairage (CIE) that can be combined to define all light sensations we experience with our eyes. These CIE primaries form an international standard for specifying colors. Color models may be then developed to conveniently specify color range. Among other models, the RGB (red, green, blue) primaries have been defined for color TV monitors and raster displays.

---

[1] In the following, the author is indebted to the excellent book *Fundamentals of Interactive Computer Graphics* by J. D. Fooley and A. Van Dam, published by Addison-Wesley Publishing Company, 1982.

The RGB model uses a three-dimension cartesian co-ordinate system. The main diagonal, with equal amounts of each primary, represents the gray levels. This model does not relate to intuitive notions of hue, saturation and brightness.

The HSV (hue, saturation, and value) model of Smith[2] calls upon these notions. Figure 5.1 represents the six-sided cone defining the model. The top of the hexcone corresponds to V=1, which contains the maximum-value color. The point at the apex is black and has a co-ordinate of V=0. Complementary colors are 180° opposite one another as measured by H, which is the angle around the vertical axis, with red at 0°. The value of S is a ratio, ranging from 0 on the centerline (V-axis) to 1 on the triangular sides of the hexcone. For example, co-ordinates for yellow are $H=\pi/4$ and $S=1$ (V may be any value between 0 and 1). For white, $V=1$, $S=0$, and H can take any value between 0 and 2π.

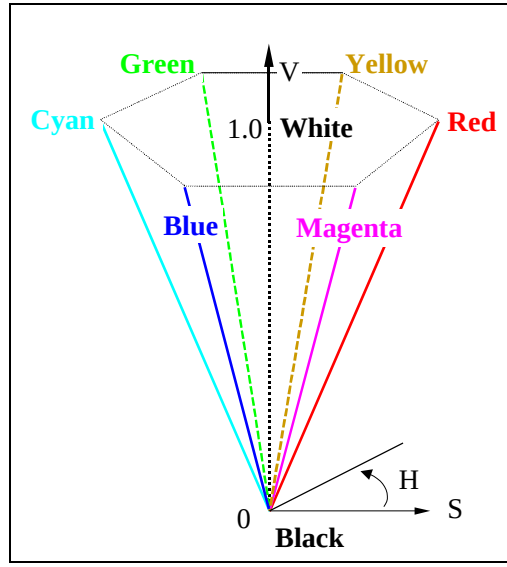


Figure 5.1. Color model HSV of Smith (op. cit.)

Other models have been derived from this HSV model. In the HSV model, the value V is equal to *max(R, G, B)* which means that two components are ignored to compute the value. It has been proposed to use instead a linear combination of the three primaries, whose result is called here *intensity I*.

[2] A. R. Smith. *Colour gamut transform pairs*. Computer Graphics, 12, 12-19, 1978.

This forms the IHS model that is currently used for the fusion of images. Several variations of this model have been proposed, which mainly differ in how the intensity is computed.

A SIMPLE MODEL FOR THE CONVERSION RGB-IHS

As an example the equations to convert RGB model into IHS model and reciprocally are given below, as found in Pohl and Van Genderen[3], and which are found in several commercial softwares.

The *RGB to IHS transform* can be performed using the following equations:

$$\begin{pmatrix} I \\ n1 \\ n2 \end{pmatrix} = \begin{pmatrix} 1/Ö3 & 1/Ö3 & 1/Ö3 \\ 1/Ö6 & 1/Ö6 & -2/Ö6 \\ 1/Ö2 & -1/Ö2 & 0 \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix} \qquad [5.1]$$

and $H = tan^{-1}(n2 / n1)$ and $S = Ö(n1^2 + n2^2)$

Note that H is not defined if $n1 = 0$, *i.e.* if $R+G = 2B$

Reciprocally, the *IHS to RGB transform* can be performed as follows:

$n1 = S\ cos(H)$

$n2 = S\ sin(H)$

$$\begin{pmatrix} R \\ G \\ B \end{pmatrix} = \begin{pmatrix} 1/Ö3 & 1/Ö6 & 1/Ö2 \\ 1/Ö3 & 1/Ö6 & -2/Ö2 \\ 1/Ö3 & -2/Ö6 & 0 \end{pmatrix} \begin{pmatrix} I \\ n1 \\ n2 \end{pmatrix} \qquad [5.2]$$

THE MODEL OF KING ET AL.

A more elaborated method has been developed by King *et al.*[4] to perform transformations between RGB and IHS spaces. It takes into account the fact that the relationship between IHS and the cartesian RGB co-ordinates is not linear and is functionally dependent upon the co-ordinates.

For the sake of presentation and not to go into many details, it is convenient to use an intermediate system ABC, which relates to RGB as follows:

---

[3] C. Pohl and J. Van Genderen. *Multisensor image fusion in remote sensing: concepts, methods and applications*. International Journal of Remote Sensing, vol. 19(5), 823-854, 1998.

[4] R. W. King, V. H. Kaupp, W. P. Waite and H. C. MacDonald. *Digital color space transformations*. In Proceedings of the IGARSS'84 Symposium. Published by the European space agency, ESA SP-215, pp. 6649-654, 1984.

$$\begin{pmatrix} A \\ B \\ C \end{pmatrix} = \begin{pmatrix} \ddot{O}(2/3) & -1/\ddot{O}6 & -1/\ddot{O}6 \\ 0 & 1/\ddot{O}2 & -1/\ddot{O}2 \\ 1/\ddot{O}3 & -1/\ddot{O}3 & 1/\ddot{O}3 \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix} \qquad [5.3]$$

and reciprocally

$$\begin{pmatrix} R \\ G \\ B \end{pmatrix} = \begin{pmatrix} \ddot{O}(2/3) & 0 & 1/\ddot{O}3 \\ -1/\ddot{O}6 & 1/\ddot{O}2 & 1/\ddot{O}3 \\ -1/\ddot{O}6 & -1/\ddot{O}2 & 1/\ddot{O}3 \end{pmatrix} \begin{pmatrix} A \\ B \\ C \end{pmatrix} \qquad [5.4]$$

The system rFq is the spherical counterpart of the ABC system, that is

$$A = r \sin q \cos F \qquad r = \ddot{O}(A^2 + B^2 + C^2)$$

$$B = r \sin q \sin F \qquad F = \arctan (B/A) \qquad [5.5]$$

$$C = r \cos q \qquad q = \arccos (C/r)$$

This system is similar to the IHS system. However, constraints exist, that request that all three components R, G, B are positive. Therefore the space described by the spherical system rFq should only include that region permitted to exist in the RGB system.

Six constraining surfaces *Ic* and *Sc* are defined as functions of F and q and corresponding to one of the co-ordinates R, G, B taking the value 0 or 1. *Ic* and *Sc* denote the maximum possible intensity and saturation for a vector in the direction specified by R, G and B.

$$R=0, \quad 2p/3 \pounds F \pounds 4p/3 \qquad Sc = \arctan[-1/\ddot{O}2 \cos F]$$

$$G=0, \quad 4p/3 \pounds F \pounds 2p \qquad Sc = \arctan[\ddot{O}2 / (\cos F - \ddot{O}3 \sin F)]$$

$$B=0, \quad 0 \pounds F \pounds 2p/3 \qquad Sc = \arctan[\ddot{O}2 / (\cos F + \ddot{O}3 \sin F)] \qquad [5.6]$$

$$R=1, \quad -p/3 \pounds F \pounds p/3 \qquad Ic = \ddot{O}3 / (\ddot{O}2 \sin q \cos F + \cos q)$$

$$G=1, \quad p/3 \pounds F \pounds p \qquad Ic=\ddot{O}6/(-\sin q \cos F + \ddot{O}3 \sin q \sin F + \ddot{O}2 \cos q)$$

$$B=1, \quad p \pounds F \pounds 5p/3 \qquad Ic=\ddot{O}6/(-\sin q \cos F - \ddot{O}3 \sin q \sin F + \ddot{O}2 \cos q)$$

The final set of equations completes the transformation:

$$I = r / Ic$$

$$H = F / 2p \qquad [5.7]$$

$$S = q / Sc$$

THE PRINCIPAL COMPONENTS ANALYSIS

The principal components analysis (PCA) is a mathematical transformation of a set of *N* images into a set of *N* new images. These *N* generated images are called principal components, or simply components. They are computed by linear combinations of the original images. These *N* components are orthogonal, that means that no component is linearly correlated with another. The total variance of the original *N* images is mapped onto the *N* components so that the first component corresponds to the largest amount of the total variance, with decreasing amount of variance going to each following component.

Let $\{B_n\}$, $n \hat{I} [1,N]$, be the set of original images and *C* the variance-covariance matrix of this set.

$$C(i, j) = covariance(B_i, B_j) \qquad [5.8]$$

As *C* is symmetric, it can be decomposed as follows:

$$V^t C V = \begin{pmatrix} d_1 & 0 & \ldots & 0 \\ 0 & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & 0 \\ 0 & \ldots & \ldots & d_N \end{pmatrix} \qquad [5.9]$$

where $\{d_k\}$ are the sorted eigenvalues so that:

$d_1 > \ldots > d_N$

and *V* is the unitary matrix whose columns are the eigenvectors:

$V = (n_1, \ldots, n_N)$, where $n_k = (n_{1,k}, \ldots, n_{N,k})^t$ is the eigenvector corresponding to d_k.

d_k is the amount of total variance that is explained by the k^{th} component. The sum of d_k, for $k=1...N$, is equal to the total variance.

The k^{th} component PCA_k is computed according to the k^{th} eigenvector:

$$PCA_k = \sum_{p=1}^{N} n_{p,k} B_p \qquad [5.10]$$

and the vector *PCA* is given by

$$PCA = V B \qquad [5.11]$$

or

$$\begin{pmatrix} \dots \\ \dots \\ PCA_k \\ \dots \\ \dots \end{pmatrix} = \begin{pmatrix} v_{1\,1} & \dots & v_{N\,1} \\ \dots & \dots & \dots \\ v_{1\,k} & \dots & \dots v_{N\,k} \\ \dots & \dots & \dots \\ v_{N\,1} & \dots & \dots v_{N\,N} \end{pmatrix} \begin{pmatrix} B_1 \\ \dots \\ \dots \\ \dots \\ B_N \end{pmatrix}$$

Reciprocally, the images B are retrieved by

$$B = V^{1}\, PCA \qquad [5.12]$$

The principal components analysis may be performed by using the correlation matrix, instead of the covariance. This implies a scaling of the axes. This helps in preventing some original images from dominating the transform because of their larger signal dynamics. The principal components analysis can also be found under the name Karhunen-Loeve technique.

THE WAVELET TRANSFORM AND MULTIRESOLUTION ANALYSIS[5]

The Fourier transform is likely the most known method for spatial analysis and does not need to be presented here. The wavelet transform is a more recent tool, which is a space-wave vector (or time-frequency) transform, while the Fourier transform only provides analysis in the wave vector (or frequency) domain. The wavelet transform may be combined with the multiresolution analysis, and both tools form a convenient means to describe, analyze and model the information contained in an image, or in a series of data.

THE WAVELET TRANSFORM

As the Fourier transform, the wavelet transform performs a decomposition of the signal on a base of elementary functions: the wavelets. The base is generated by dilations and translations of a single function y called the mother wavelet:

$$y_{a,b} = \hat{e}a\tilde{e}^{-1/2}\, y\,[(x\text{-}b) / a] \qquad [5.13]$$

where $a, b \in \hat{I}\ \hat{A}$ and $a > 0$. a is called the dilation step and b the translation step. Many mother wavelets exist. They are all oscillating functions that are well localized both in time and frequency. All the wavelets have common properties such as regularity, oscillation and localization, and satisfy an admissibility condition. For more details about the properties of the

[5] This section has been written in collaboration with Thierry Ranchin. Many thanks to him.

wavelets, one can refer to Meyer[6] or Daubechies[7]. Even if they have common properties, each of them leads to a unique decomposition of the signal related to the selected mother wavelet. In the one dimension case, the continuous wavelet transform of a function *f(x)* is:

$$WT_f(a,b) = \left\langle f, y_{a,b} \right\rangle = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{+\infty} f(x) \overline{y\left(\frac{x-b}{a}\right)} dx \qquad [5.14]$$

where $\overline{y\left(\frac{x-b}{a}\right)}$ is the complex conjugated of y. The computation of the wavelet transform for each scale a and each location b of a signal *f(x)* provides a local representation of *f(x)* and the information content is represented by the wavelet coefficient $WT_f(a,b)$. The process can be reversed and the original signal reconstructed exactly (without any loss) from the wavelet coefficients by:

$$f(x) = \frac{1}{C_y} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} WT_f(a,b) y_{a,b}(x) \frac{dadb}{a^2} \qquad [5.15]$$

where C_y is the admissibility condition of the mother wavelet. Discrete versions of the wavelet transform exist and are applied to signals using filters.

THE MULTIRESOLUTION ANALYSIS

The multiresolution analysis a means to describe and model the signal in the time-frequency domain or in the space-wavevector domain or in any domain with similar duality. It makes use of space (or time) transforms or filters. This section does not describe the multiresolution analysis in its mathematical aspects. It introduces the multiresolution analysis to the reader *via* specific cases, which are of high value in fusion of images.

Figure 5.2 is a very convenient illustration of the multiresolution analysis and more generally of pyramidal algorithms[8]. The basis of the pyramid is

---

[6] Y. Meyer. *Ondelettes et opérateurs 1: Ondelettes*. Hermann, Paris, France, 215 p., 1990.

[7] I. Daubechies. *Ten lectures on wavelets*. CBMS-NSF regional conference series in applied mathematics 61, SIAM, Philadelphia, USA, 357 p., 1992.

[8] S. G.Mallat. *A theory for multiresolution signal decomposition: the wavelet representation*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 11(7):674-693, 1989.

the original image. Each level of the pyramid is an approximation of the original image computed from the original one. When climbing the pyramid, the successive approximations have coarser and coarser spatial resolutions. The computation of the approximations is done using a base of functions, called the scale functions. The base is generated following the same scheme than the one used for the generation of the wavelet base. Hence scale and wavelet bases have the same properties.

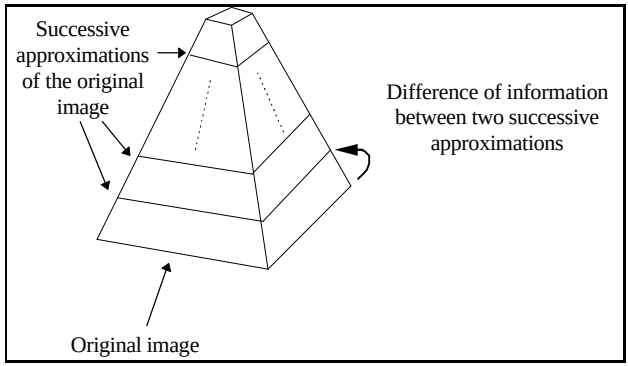


Figure 5.2. Pyramid representing the multiresolution analysis combined with the wavelet transform

The basis of the pyramid is the landscape measured by the sensor. The wavelet coefficients are produced by the application of the wavelet functions: they describe the differences existing between two successive approximations of the same image (*i.e.* two successive levels of the pyramid). The approximations are produced by the application of the scale functions. Approximations are also called contexts. This phase is called *analysis*.

If the process of the multiresolution analysis is inverted, the original image can be exactly reconstructed, from one approximation and from the different wavelet coefficients describing the differences in signal between this approximation and the original image: this phase is called *synthesis*.

As we are processing images, the wavelet and the scale functions are applied first in columns and then in lines (rows). This leads to a representation of the information using the scheme proposed in Figure 5.3. Here a dyadic wavelet transform is assumed, that is that the resolution of any approximation is half that of the previous approximation.

The dilation of both the wavelet and the scale function is obtained by the sub-sampling of the original image. Hence if the original image comprises e.g., 768 lines by 1024 columns, the first approximation is 384 lines by 512 columns, as well as the three wavelet coefficients images.

The first context image contains all the scales greater than half the original spatial resolution (1/2 in Fig. 5.3). The three wavelet coefficients images represent the structures with sizes comprised between the original spatial resolution and half this resolution for the diagonal (C^D), vertical (C^V) and horizontal (C^H) directions. In the second context image are represented all the scales greater than a quarter of the original resolution, and the wavelet coefficient images contain the scales between half and a quarter of the original resolution.

If the multiresolution analysis is performed once more, the context image 1/4 will be decomposed into a context image 1/8 comprising all the scales greater than one eighth of the original resolution and three wavelet coefficients images in diagonal (C^D), vertical (C^V) and horizontal (C^H) directions representing the structures with sizes comprised between a quarter and one eighth of the original resolution

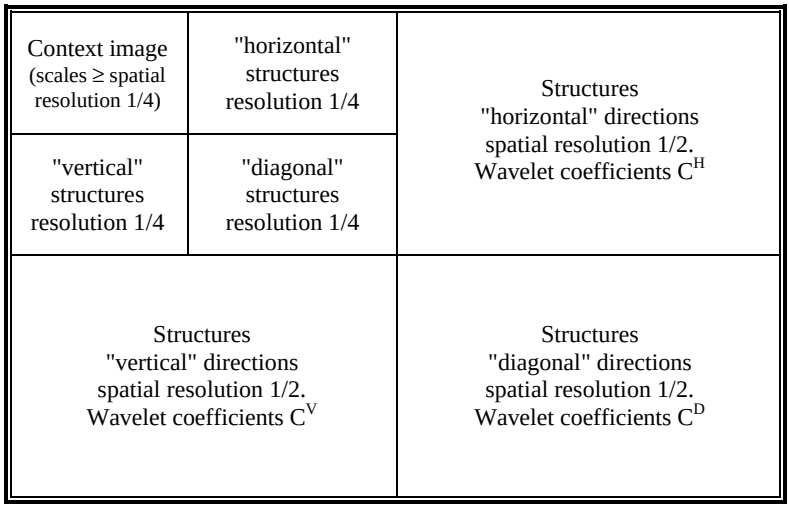


Figure 5.3. Presentation of a multiresolution analysis using the Mallat algorithm. Original resolution of the image is 1. (Taken from Ranchin and Wald[9])

It should be noted that the pyramidal approach is one of the many possible implementations of the multiresolution analysis. The multiresolution analysis may call or not upon wavelets that can be constructed in a great deal of ways. Hence, the multiresolution analysis comprises a very large

9 T. Ranchin T. and L. Wald. *The wavelet transform for the analysis of remotely sensed images*. International Journal of Remote Sensing, 14(3):615-619, 1993.

number of possible implementations. Many of them are suitable for the application of the ARSIS concept for image fusion, which is discussed in following Chapter.

The following sections propose two different practical implementations of the multiresolution analysis and the related wavelet transforms.

PRACTICAL IMPLEMENTATION OF THE ALGORITHM OF MALLAT

The algorithm of Mallat can be implemented using a filter bank structure (Fig. 5.4). This algorithm is one of the possible implementations of the pyramidal approach.

In Figure 5.4, $f_j(x, y)$ represents the original image, where x is the column (beginning from 1), and y is the line or row (beginning from 1). The index j denotes the ranking of the current approximation. $f(x, y)$ is the original image, $f_1(x, y)$ is the first approximation etc.

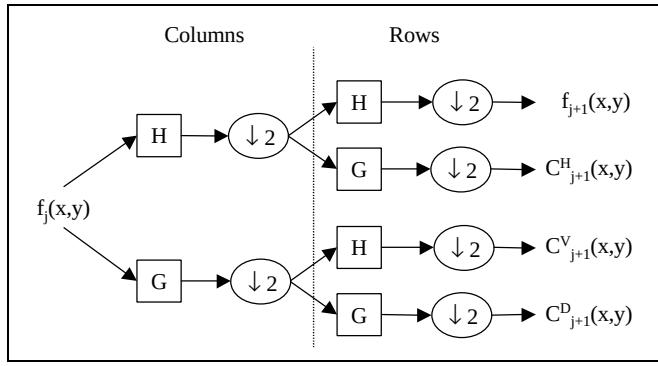


Figure 5.4. Implementation of the analysis phase of the Mallat's algorithm into a filter bank structure

In this Figure, H and G are two filters. The columns and rows are processed separately. Filter H is applied on the columns of $f_j(x, y)$. Same for filter G. Both resulting images are re-sampled (operation $\downarrow 2$, Fig. 5.4): one column over two is removed. Then filters H and G are applied again on each re-sampled image. The resulting four images are re-sampled again: one row over two is removed. This results into four images:

- $f_{j+1}(x, y)$ is the approximation (context) with half the spatial resolution of the original $f_j(x, y)$ one;
- the three wavelet coefficients images $C^H_{j+1}(x, y)$, $C^V_{j+1}(x, y)$ and $C^D_{j+1}(x, y)$.

The filters *H* and *G* may be selected among those designed by Daubechies[10]. Here the four-tap filters are choosen. Table 5.1 gives the coefficients of the filter *H*. All these coefficients have to be divided by $\sqrt{2}$ for normalization purpose.

H(0)	*H(1)*	*H(2)*	*H(3)*
0.482962913145	0.836516303738	0.224143868042	-0.129409522551

Table 5.1. Values of the coefficients of the filter H for the wavelet (Daubechies, op. cit.)

The filter *H* is applied as shown in Figure 5.5, which presents its application along a row. Operations along columns are similar. The new value $f_{new}(x, y)$ for the current pixel (x, y) is computed as a multiplication between the coefficients of the filters and the pixels:

$$f_{new}(x, y) = H(3)\, f(x-2, y) + H(2)\, f(x-1, y) + H(1)\, f(x, y) + H(0)\, f(x+1, y) \quad [5.16]$$

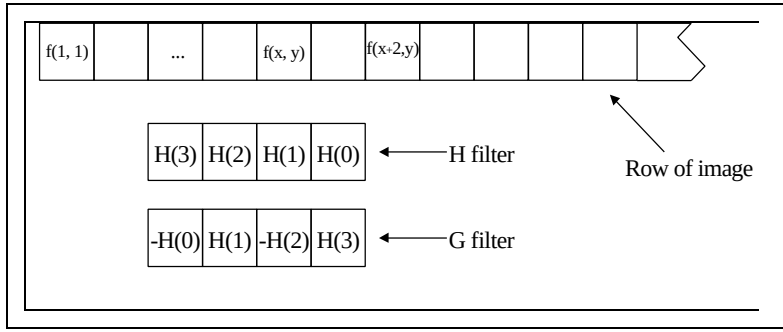


Figure 5.5. Position of the filters for the analysis. f(x, y) denotes the function on which the filter is applied, e.g. $f_j(x, y)$. The column x is odd. Standard solutions can be adopted for the borders

Then, due to the sub-sampling, the next pixel to be computed is $(x+2, y)$. The new value of pixel $(x+1, y)$ is temporarily set to 0. Actually, this pixel is not processed at all because it will be removed by sub-sampling (Fig. 5.4). Once the entire image processed for the columns, the same process is applied to all rows along the column direction.

[10] I. Daubechies. *Orthonormal bases of compactly supported wavelets*. Communications on Pure and Applied Mathematics, vol. XLI, 909-906, 1988.

Filter *G* is applied in a similar way. This filter derives from the filter *H* (Fig. 5.5). The coefficients are the same in absolute value, but their sign and order differ.

From the approximation $f_{j+1}(x, y)$ and from the three wavelet coefficients images, $C^{H}_{j+1}(x, y)$, $C^{V}_{j+1}(x, y)$ and $C^{D}_{j+1}(x, y)$, one can exactly reconstruct the original image $f_j(x, y)$ (Fig. 5.6).

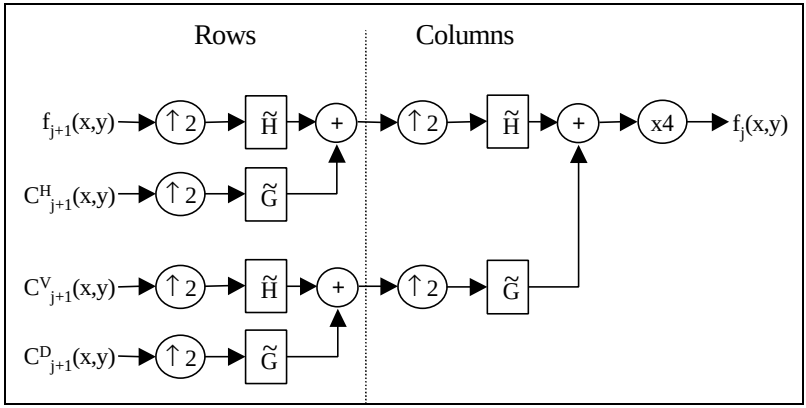


Figure 5.6. Implementation of the synthesis of the Mallat's algorithm into a filter bank structure

In the synthesis, an over-sampling $\uparrow 2$ is necessary. It is obtained by adding a zero between the pixels (Fig. 5.7). In the case of orthogonal filters as presently, $\tilde{H}$ and $\tilde{G}$ are the same filters than those used in the analysis (*i.e.* *H* and *G*). Firstly an over-sampling in columns is applied on the approximation and the three coefficients images. Then either the filter *H* or *G* is applied for each pixel including those set to zero in the analysis phase. Results are summed two by two as shown in Figure 5.6. An over-sampling is applied in lines, prior to the application of filters $\tilde{H}$ and $\tilde{G}$. The final summation provides the original $f_j(x, y)$ (or synthesized) image after a multiplication by 4.

The application of the filters $\tilde{H}$ and $\tilde{G}$ is performed as shown in Figure 5.7, in a similar way than for the analysis phase.

It is recommended to check the good implementation of the Mallat's algorithm. The following scheme can be employed, given any image. Apply analysis (Fig. 5.4) with one iteration: the first approximation $f_{j+1}(x, y)$ is obtained. Then perform synthesis (Fig. 5.6) on $f_{j+1}(x, y)$. The resulting image should be identical to the original, except for the borders of the image. Computing the difference between both images, pixel by pixel, can

check this. The whole checking procedure should be performed for more than one iteration.

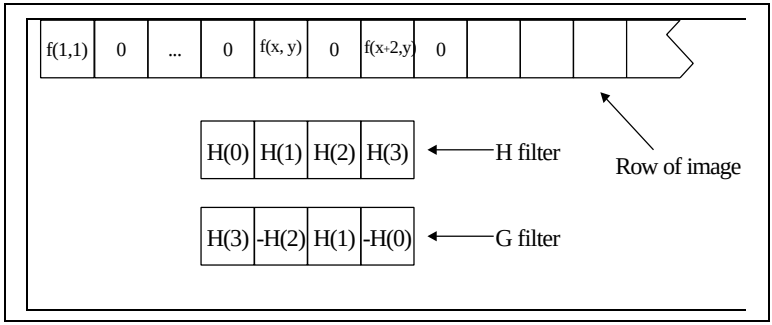


Figure 5.7. Position of the filters for the synthesis. f(x, y) denotes the function on which the filter is applied, i.e., $f_{j+1}(x, y)$, $C^{H}_{j+1}(x, y)$, $C^{V}_{j+1}(x, y)$ or $C^{D}_{j+1}(x, y)$. The column x is odd. Standard solutions can be adopted for borders of image

THE "À TROUS" ALGORITHM FOR THE MULTIRESOLUTION ANALYSIS AND WAVELET TRANSFORM

Actually, the discrete approach of the wavelet transform and of the multiresolution analysis can be done with several different algorithms. The Mallat's algorithm uses an orthonormal basis, which may be well suited to several problems in data fusion, but is not shift-invariant, which may cause some problems.

Another popular wavelet transform is the "à trous" (with holes) transform[11]. A sequence of approximations is constructed, by performing successive convolutions with a filter obtained from a scaling function (the filters *H* and *G* in the Mallat's algorithm). This scaling function may be a B_3 cubic spline[12], leading to the following 5x5 filter:

---

[11] P. Dutilleux. *An implementation of the "algorithme à trous" to compute the wavelet transform*. In Compte-rendus du congrès ondelettes et méthodes temps-fréquence et espace des phases, Marseille 14-18 septembre 1987, Springer-Verlag ed., pp. 298-304, 1987.

[12] J. L. Starck and F. Murtagh. *Image restoration with noise suppression using the wavelet transform*. Astronomy Astrophysics, 288, 342-350, 1994.

$$1/256 \begin{pmatrix} 1 & 4 & 6 & 4 & 1 \\ 4 & 16 & 24 & 16 & 4 \\ 6 & 24 & 36 & 24 & 6 \\ 4 & 16 & 24 & 16 & 4 \\ 1 & 4 & 6 & 4 & 1 \end{pmatrix} \quad [5.17]$$

The wavelet coefficients are the differences between two consecutive approximations, $C_{j+1}(x, y) = f_j(x, y) - f_{j+1}(x, y)$. The synthesis equation follows:

$$f_j(x, y) = f_{j+n}(x, y) + \sum_{k=1}^{n} C_{j+k}(x, y) \quad [5.18]$$

where n is the number of iterations.

Note that in this scheme, and contrary to the Mallat's algorithm, all approximations and wavelet coefficients have the same number of pixels than the original image. It does not form a pyramid as in Mallat's algorithm but a parallelepiped. The application of the filter should take this into account. Practically speaking, the size of the filter should grow as the successive approximations are constructed. The present filter should apply to the original image $f(x, y)$ to obtain the first approximation $f_1(x, y)$. Then null values (0) are introduced in the filter in-between the present coefficients in both directions. This larger filter is applied to the first approximation $f_1(x, y)$; this results into the second approximation $f_2(x, y)$. Then zeros are again introduced into the filter, which again doubles in size in both directions. This iterative process leads to the sequence of approximations and further of wavelet coefficients.

Contrary to the wavelet of Daubechies presented above, the "à trous" algorithm is not orthogonal, that is that the wavelet coefficient $C_{j+1}(x, y)$ for a given scale j retains information from the neighboring scales. Otherwise said, $C_j(x, y)$ and $C_{j+1}(x, y)$ are correlated. The "à trous" algorithm is certainly easier to implement than the Mallat's algorithm. Nevertheless, their properties are different, and the influence of these properties on the result of the fusion should be carefully analyzed.

6. FUSION OF IMAGES

INTRODUCTION

The general approach in the fusion of images is to create a new set of images *I*, usually of reduced dimension, from the original sets of images:

$$I = f(A, B, C, D, ...) \quad [6.1]$$

where *A, B, C, ...* are the original sets of images and characteristics that may be derived from them. These sets may originate from various modalities, e.g., panchromatic, X-rays, electron microscope, taken at different instants and with different times of integration and may have different space resolutions. Within a set, all images are geometrically aligned (see later) and have the same pixel size. As said in Chapter 3, here the term image comprises any information that is presented in a raster format, or gridded format in 2 dimensions. The grid cell is called pixel.

A classical example of fusion of images is the classification process. Several images of commensurate or non-commensurate measurements and possibly of other information are inputs to a classifier. If the classification if of supervised type, a codebook exists that is input to the fusion process as an external knowledge. The result is an image of taxons and possibly another image of the related accuracy (or plausibility, or reliability etc.). In an unsupervised procedure, the state vectors of the pixels are grouped on similarity properties. The final classification is performed by querying additional constraints to the operator. The unsupervised classification is usually an iterative fusion process with successive refinements until operator satisfaction is met. In classification processes, the original dimension of the information is reduced. In this example, the semantic level of the fused product is higher than that of the original set of images.

Another example is given by the construction of digital elevation model (DEM), which represents the relief of an object or a terrain relative to a reference. This elevation model is constructed by stereo-photogrammetry. This is one of the major applications of the images acquired by space-borne systems observing the Earth. Cloud cover impedes the observation by optical systems, resulting into "loss" of usable data and gaps in the constructed elevation model. The missing parts can be recovered by performing interferometry using radar systems in lieu of optical systems. In the example illustrated in Figure 6.1, fusion is performed on two optical images to obtain a first DEM (actually an image of parallax), then on two radar images to obtain a second DEM (actually an image of coherence). Quality parameters are also available for each of the DEMs. A Bayesian

approach is used to optimize parameters for the images of parallax and coherence for the fusion of these derived images. The resulting image is the final digital elevation model, which is more complete and more accurate than the two others are[1].

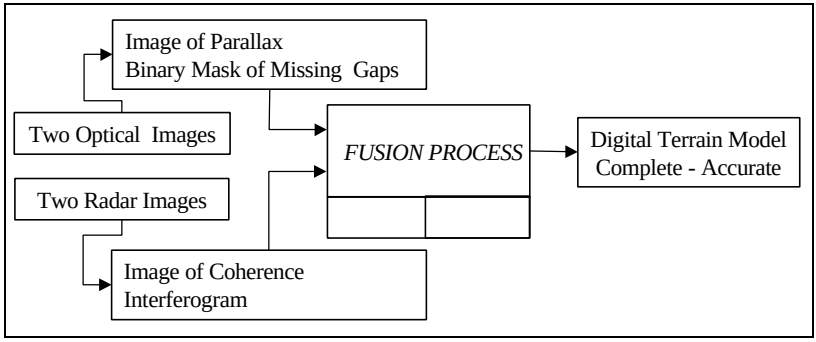


Figure 6.1. Example of a fusion process for the construction of a digital elevation model

Many other approaches in image fusion exist. Some include extraction of features from each image and then fusion of features without referring any more to image. An example is given by the mapping of roads by fusing several sets of images, where the final product is a symbol "road" in the form of vectors in a geographic information system.

The objects of the fusion of images are various. Some have already been seen. They depend upon the domain of applications. In robotics, the major thrust for fusing images is the acquisition of the relief for accurate displacements or moves of the robot. In environment, classification is the most usual fusion process when knowledge of land use and its characteristics is at stake. If visual analysis and interpretation are under concern, the fusion aims at creating a reduced set of images, which contains all the information of interest present in the original sets of images. Fusion may be performed to create new sets of images in various modalities with a better spatial resolution, which are close to similar observations if existent.

This Chapter presents very briefly the classification and identification as a fusion process. Then it focus on techniques calling upon fusion to display information of interest scattered in several images acquired by various modalities. Color space is used in that purpose. Some techniques are

---

[1] In L. Wald. *Data fusion for a better exploitation of data in environment and Earth observation sciences*. Presented to UNISPACE III, July 2000, Vienna, Austria. By courtesy of Issam Tannous.

extensively discussed. They are useful to deal with a large number of commensurate modalities, e.g., hyperspectral optical sensors acquiring images in several hundreds of channels, or with non-commensurate images.

Following Chapter focuses on methods for the synthesis of images having the best spatial resolution available in the original sets of images.

GEOMETRICAL ALIGNMENT OF IMAGES

The alignment has been discussed in Chapter 3. The images should be aligned to performing correctly the fusion process. The exact constraints depend upon the method; they will be presented on a case by case basis in the following pages. However the geometrical alignment is almost always requested for the fusion of images; it is discussed now for once.

Many techniques for image fusion perform calculations on a pixel basis. Hence the images should be perfectly superimposed in order to get maximum accuracy. The geometrical alignment is often the most critical step in the alignment process. It is also called co-registration, superimposition, geometrical correction, conflation or navigation. The terms warping or rubber sheeting are sometimes used, but they refer to specific techniques. The geometrical alignment can be performed on an absolute basis. An absolute reference is selected that exists outside the specific problem under concern. All images are aligned in this reference. An example is the latitude - longitude system or any geographic projection system. The alignment may be relative: an image is selected, which serves as a reference, and all images are aligned on that one. In this case, the size of the pixel of the reference image plays a major role.

Some systems deliver images that are already co-registered. Otherwise there are several approaches to perform geometrical alignment.

Assume that the systems for image acquisition are perfectly described by analytical models whose parameters are all known. Then the alignment is made by the appropriate combination of the models. For example, if the set B of images is to be aligned on the set A, and assuming that f_A and f_B represent the models, the alignment of B is made by performing $f_A[f_B^{-1}(B)]$. The number of parameters is usually very large, and, very often many of them are not precisely known. To overcome this shortcoming, one calls upon external knowledge if existent (e.g., range of variation of each parameter, mean value, value at previous instant or iteration), and on some features present in both sets of images. These features are assumed to represent the same object and to have the same locations. Using this additional information, one may find the parameters, which minimize a cost function.

Very often, the models are unknown to the persons performing the fusion process. This is the case when images are commercial products or originate from a commercial acquisition system.

In some cases, one may use the technique of landmarks. A landmark designates a pixel or a set of pixels, which have remarkable properties: very bright pixel in a dark context, geometrical arrangements like lines, crosses, etc. It is also called homologue points in stereo-photogrammetry or ground control point. Of interest are the landmarks that appear in one of the images of each set of images *A* and *B*. Then one assumes that any landmark present in both sets represents the same object and should have the same geometrical location in the reference space (here selected as being that of *A* for the sake of the simplicity). For example, if a table is seen under similar viewing geometry in both sets of images of an indoor scene with permanent, or well-controlled and modeled, illumination, and if it has not been moved between the times of acquisition of the images, its angles may constitute landmarks. This hypothesis is not always valid and should be checked carefully for each landmark. Many difficulties arise in the case of objects, which are seen under very different viewing geometry, or objects, whose shapes are changing during the time lag between the acquisition of images in two different modalities, such as the inside of a living stomach or cloud fields, or objects, whose limits / borders may vary according to the modalities, such as X-rays versus electron microscope.

If several landmarks are found, one may estimate a model for the conversion of the geometry of *B* into that of *A*. The complexity of the models depend upon the number of landmarks. In the simplest case, with at least three landmarks, one may perform a least-square fitting of straight line (polynomial of first order) on the co-ordinates (line, row). The detection of landmarks is usually done manually; hence the amount of landmark amounts to a few tens at most. Some methods perform automatic detection of pairs of landmarks; the resulting number is very large and the landmarks are well distributed within the reference space. The space can then be divided into triangles, each landmark being the summit of a triangle. On each triangle, a bi-linear model can be estimated. Some constraints may be taken into account such as the continuity of the models and their first derivatives on the sides of the triangles. The geometrical alignment model is the composition of these local models. Additional local models need to be estimated for the part of the reference space that lies outside the convex envelope defined by the landmarks. Then the co-ordinates $(x, y)_B$ in the space *B* are given by

$$(x, y)B = fB[(x', y')A] \qquad [6.2]$$

where $(x', y')_A$ are the co-ordinates in the reference space and f_B the model.

The case of oblique viewing without enough knowledge of the acquisition model and its parameters should be treated differently. Here oblique viewing implies that some facets of the object or some slopes in the landscape are hidden. The lines of sight differ between the two sets of images *A* and *B*. This means that an object may not appear in one of the set or not have the same appearance in both sets. Oblique viewing is the standard mode of acquisition in radar imagery. If the creation of the image is mostly due to the changes in relief, then the problem my be solved by the knowledge of the lines of sight relative to the observed scene if a digital elevation model is available and accurate enough. In the case of radar imagery in very steep relief, the area covered by the digital elevation model should be much larger than the reference space in order to take into account the steep high relief outside the reference space, whose echoes are present in the radar image. One simulates the oblique viewing of the elevation model using the parameters of the set *A* and then of the set *B*. Two synthetic images are obtained: A^S and B^S. The technique of landmarks is applied to *A* and A^S, and to *B* and B^S. Here landmarks are composed of crests and troughs. Then all geometrical models are known and one may convert the space of *B* into that of *A*.

Once the geometrical model known, one may find useful or necessary to resample the images of the set *B* to project them into the reference space:

$$B'(x', y')_A = g[B(x, y)_B], \text{ where } (x, y)_B = f_B(x', y')_A \qquad [6.3]$$

Several resampling operators are available. Very good results are attained by truncated versions of the sine cardinal. A bi-cubic function offers a very good trade-off between the accuracy and the computing time. In some cases of oblique viewing, the value of the pixels that are shadowed by the relief may be inferred from other images, given some assumptions and models. An example is the resampling of radar imagery in steep relief using optical imagery[2]. This also demonstrates that a fusion process may be nested into another one.

Assume that *A* has the highest spatial resolution *h* (*i.e.* the smallest pixel size) and *B* the lowest spatial resolution *l*. The techniques described in this Chapter and the following require that the images B_l be exactly superimposed onto the images A_l at the resolution *l*. For any pixel $(x, y)_l$ of the reference space at resolution *l*, one should be able to construct the state vector of the various modalities of the sets A_l and B_l by concatenation. For

[2] L. Castagnas. *Application of the multiresolution analysis to the fusion of satellite images: example of SPOT and ERS-1 data*. In Proceedings of the 1993 IEEE International Conference on Systems, Man and Cybernetics, vol. 3, pp. 684-686. IEEE n° 93CH3242-5, 1993.

the sake of the simplicity, in the following, the images are assumed to be geometrically aligned and the term "images of lowest resolution" B_l denote the projected resampled images.

CLASSIFICATION - IDENTIFICATION

Estimation of the identification of objects by means of techniques of correlation - association is a common task in fusion of images. It is also called classification, or pattern recognition. The fusion process benefits very often from the properties of complementarity of the sources. In some cases, redundancy of sources is helpful to increase the quality of the identification, especially if some sources are noisy or unreliable or inaccurate.

Comparison of observations with models describing the physics of a phenomenon by the means of estimation methods such as data assimilation or Kalman filtering is a task of identity estimation in terms of the given models. In such approaches, data are usually commensurate.

Other approaches are more appropriate to the fusion of non-commensurate information. Many classifiers call upon laws pertaining to statistics and probabilities and are powerful tools to merge commensurate or non-commensurate images. For example, it is common to use optical and radar images as inputs to a classifier. Additional inputs may be some "texture" parameters (e.g., local variance) computed on some of the original images. Other features may also be taken into account by the classifier (e.g., boundaries or hydrography by the means of a geographical information system.)

Usual techniques in classification of commensurate or non-commensurate data are cluster analysis, classical inference, Bayesian inference or the Dempster-Shafer theory. Cognitive-based methods aim at reproducing the human inference process. Knowledge-based systems, or expert systems, or fuzzy logic belong to this family.

There is a wealth of literature about classification or pattern recognition[3]. Many classifiers exist in commercial softwares. They may request more or less computer resources; it usually remains reasonable. The input images must be geometrically or geographically aligned. They should have the same pixel size (usually that of the image having the best spatial resolution).

[3] See the special issue on data fusion, *IEEE Transactions on Geoscience and Remote Sensing, 37(3), 1999.*

COLOR COMPOSITING - THE IHS AND PCA METHODS

It is very common to allocate the three basic colors (Red, Green, and Blue) to three modalities or spectral bands. The color compositing permits to visualize the combination of these three modalities and to understand better their relationship. A component may be actually any combination of the available information. If the number of modalities is greater than three, it is common to perform a principal component analysis, and to allocate the first three components to the R, G, and B axes. Other combinations are possible, such as sums, differences, ratios etc. Since no physical law is implied, color compositing is commonly used to visually merge homogeneous images or heterogeneous ones, such as optical and radar images. Here again, each component may be a combination of the original data.

To perform a color compositing, images should be superimposable (geometric alignment). Usually the spatial resolution does not have a major importance. Hence the images are resampled to fit the image having the best spatial resolution. They should also be aligned for signal dynamics; their histograms, or at least their range of values, should be similar. Most of the commercial softwares offer such functionalities. There is no particular difficulty.

The problem to solve is the creation of a triplet *(R, G, B)* at spatial resolution *h* from the sets of images and derived characteristics, also at spatial resolution *h*.

$$(R, G, B)_h = f(A_h, C_h, ...) \qquad [6.4]$$

Two techniques are mainly used. They pertain to the projection and substitution type and are detailed hereafter. Other combinations are possible, such as sums, differences, ratios, ratios of differences and sums etc., which are performed on a case by case basis.

THE IHS METHOD

The IHS method applies to four images, three of them belonging to a set *C*, one to the set *A*. It is based on an analogy between the three images (also called channels or bands) of the set *C* and the color primaries, as discussed previously. The fourth image *A* plays a role apart. It is usually an image of the same type than the images of the set *C* but with a higher spatial resolution, or an image non-commensurate with the images of the set *C*. If the set *C* comprises more than three bands, then one should select three bands among them. Alternatively the original bands, or a selection of them, may be combined in order to give three bands. The ways of combining them are diverse and depend upon the purpose of the fusion process. If the set *C* comprises only two bands, one trick is to create a third one by combining

the two available by, e.g. summing up these two. The image *A* itself may result from a combination of several images.

Each of the three bands C_i is labeled as blue, green and red respectively. Then, these color components are converted into intensity (*I*), hue (*H*) and saturation (*S*) components using for example, one of the models discussed previously. The next step is the substitution of the intensity by the image *A*. The scheme is shown in Figure 6.2:

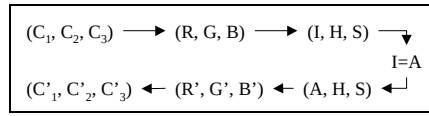


Figure 6.2. Scheme of the IHS method

The substitution of the intensity by *A* implies to match the dynamics range of *A* to that of *I*, which can be done by histogram matching, or variance and mean matching, or other techniques. Actually, it is recommended to perform the alignment of the dynamics on the *R*, *G*, *B* and *A* images before their conversion into the IHS model.

Figure 6.3 displays a typical scheme for an IHS process. Here, three images, called R_l, G_l and B_l, of low spatial resolution *l* are fused with another one HR_h of higher resolution *h*.

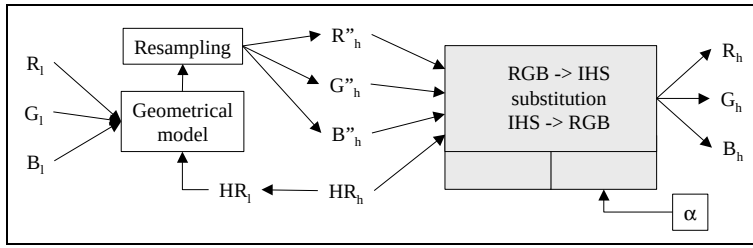


Figure 6.3. Typical scheme for a fusion process based upon the IHS method

The first step is the alignment in gray levels of the four images R_l, G_l, B_l and HR_h. Then the image HR_h is resampled to the resolution *l*. The geometrical alignment is performed by the means of the landmark technique, the estimation of the geometrical model, its application to images R_l, G_l and B_l, and finally the resampling of the results at resolution *h*. The geometrically aligned outputs R''_h, G''_h and B''_h and the image HR_h are the inputs of the fusion cell. R''_h, G''_h and B''_h are converted into I'_h, H_h and S_h. The substitution of the intensity I'_h by HR_h ($I_h = HR_h$) and the inverse conversion from the IHS system into the RGB system provide a new set of images (R_h, G_h, B_h), which includes more spatial details (more high frequencies) than the original set (R_l, G_l, B_l).

Refinements can be made which include the substitution of the intensity I by a linear combination of the image A and the original intensity I':

$$new\ intensity\ I = a\ A + (1-a)\ I' \qquad [6.5]$$

At that stage, the fusion is accomplished and the process may stop there if visual analysis of the fused image is performed in the IHS model. Otherwise, the last step performs the inverse model converting the new IHS components into blue, green and red components.

Some commercial softwares for the processing of images from e.g., Adobe or JASC companies, propose a function called transparency, which acts similarly to the IHS method. The new intensity is a linear combination of the original intensity and of A, which can be user-adjusted.

The IHS method can be applied to images not resulting from measurements, but displaying other types of information e.g., attributes or taxons (see the following example), not obeying to a relationship of order.

THE PCA METHOD

The PCA method calls upon the principal components analysis and is similar in essence to the IHS method. It applies to one plus two or more images, and is more general than the IHS. It is recommended either to equalize the dynamics of the signal in each image, including A, in order to make them similar, or to perform the PCA with the correlation matrix and matching the dynamics of A with that of the first component. This alignment of dynamics ensures better performances but may cause a distortion of the spectral content.

The PCA provides N components. The first component corresponds to the largest amount of the total variance. Very often, this component may be seen as a rough approximation of the average value of the images at each pixel. In the PCA method, the first component acts as the intensity in the IHS method. Accordingly, it is replaced by the image A. Refinements can be brought in this substitution step by performing a combination of this image A and of the first component. At that stage, the fusion is accomplished. An inverse PCA transformation may be performed to provide the fused N images in the original co-ordinates system.

If visual analysis is at stake and since very often, most of the variance is contained in the first three components, one does not perform the inverse transformation. Only the three first components are retained. They are assimilated to the three color primaries RGB discussed earlier, and displayed for analysis.

This method is appropriate in the case of a very high number of images C_i provided e.g., by a hyperspectral imager, which are to be fused with other

modalities. Like the IHS method, it deals indifferently with commensurate or non-commensurate images. However, contrary to the IHS method, it should only be applied on images whose contents are ruled by a relationship of order.

The principal components analysis is a form of orthogonal transformation, based on the analysis of the covariance matrix. Other transformations exist, orthogonal or not; they can be used in lieu of the PCA using the same principle. Examples are the Fourier transform (in the modality domain) or empirical orthogonal functions.

AN EXAMPLE OF THE IHS METHOD

In the course of the realization of the European solar radiation atlas[4], a clickable map for Europe has been produced, which serves as an interface for the user to select some geographical points. The map has been made in raster format from vector geographical databases. Its scale is small: the pixel size is approximately 10 km (actually 5' of arc angle).

To help the user in orientating himself, it was decided to draw landmasses and large water bodies (lakes and seas). Nations were also drawn on the map, one color per nation and geographical co-ordinates (latitude, longitude) were available at any instant (Fig. 6.4). However, this was not sufficient and incorporating major orographic features (*i.e.* terrain relief) into the map brought an additional visual help. Preservation of the color of a nation is important to well distinguish them. Accordingly, the IHS technique was selected as the fusion process. Modulating the intensity of the color has done the incorporation of the orographic features.

The scheme of the fusion process is shown in Figure 6.5. It was performed by means of a commercial software for image processing.

The raster map was split into the three *R*, *G*, and *B* components. Setting the illumination source at its northeast, with an elevation above horizon of approximately 30 degrees illuminated the digital elevation model (DEM). This produced a shadowed DEM, which enhanced the relief and steep slopes. The *R, G, B* components were converted into *I, H, S* components.

Using a transparency function of the software, a new intensity was produced, which is a linear combination of the shadowed DEM and the original intensity (see Eq. 6.5). The water bodies were preserved.

---

[4] *European solar radiation atlas*. Fourth edition, includ. CD-ROM. J. Greif, K. Scharmer. Published for the Commission of the European Communities by Presses de l'École des Mines de Paris, France, 2000.

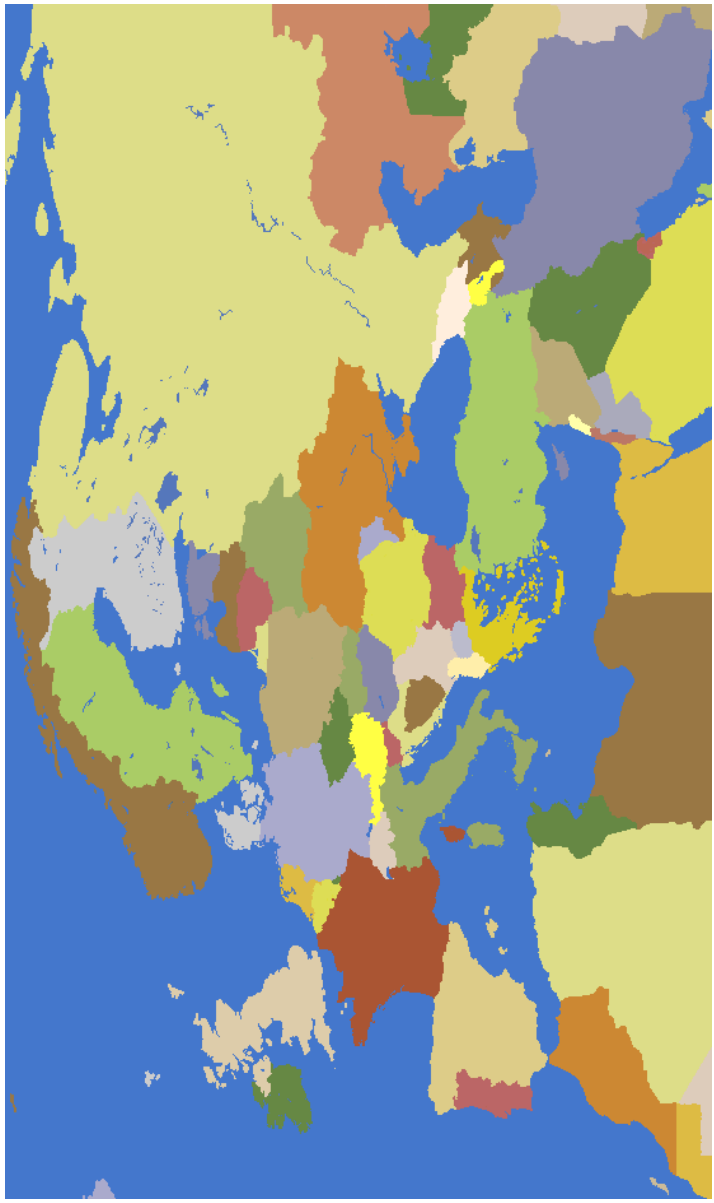


Figure 6.4. Part of the map of the countries in the European solar radiation atlas (Presses de l'École des Mines, 2000)

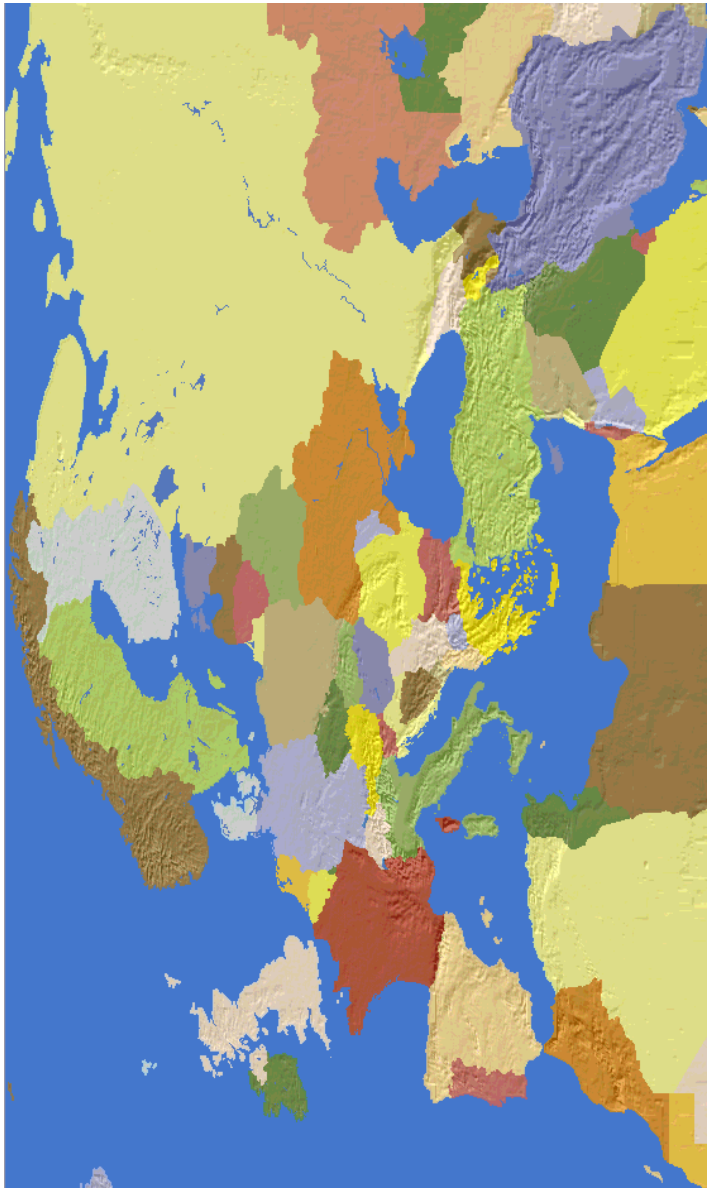


Figure 6.6. Part of the map serving as the user- interface in the European solar radiation atlas. Orographic features have been incorporated (Presses de l'École des Mines, 2000)

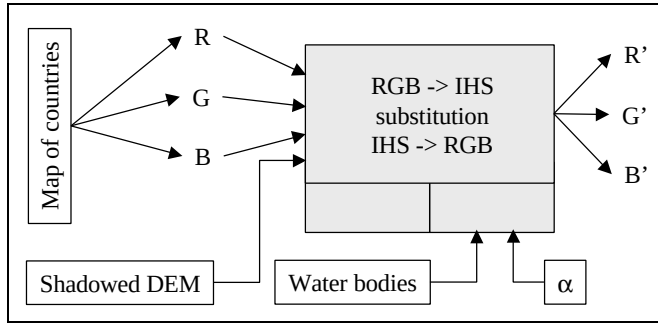


Figure 6.5. Scheme of the IHS fusion used to produce the interface map in the European solar radiation atlas.

The final product was a trade-off between the appearance of the orographic features and the preservation of colors of nations. The new *I', H, S* components were then converted into new *R', G', B'* components, which in turn were combined to produce an 8-bits image.

The resulting map is shown in Figure 6.6. Major orographic features are clearly visible, and are a valuable help to locate sites in large countries.

VISUAL ENCRUSTATION

Assume two sets of images *A* and *B*. Encrusting some elements originating from *A* into images originating from *B*, color composite or not, is a trivial form of fusion. The elements of *A* may be measurements or attributes or both. The images supporting the encrustation may be a combination of the original images *B*, or a combination of attributes derived from these images or both.

Encrustation is very useful in the visual analysis of several sets of non-commensurate images. The content of the images may be measurements or attributes; a relationship of order may exist or not for each image. Each image has its specific interest regarding some elements of the scene under concern. Encrustation is one approach to enhance the perception and visual analysis of these elements and to create a composite scene showing most, if not all, of the information of interest.

There are several techniques, which meet this general approach. It is impossible to describe all of them; this book focuses on one example. The methodology shown can be used for other applications, which are dealing with the analysis of some geometrical features perceived by one or more modalities in relationship with an environment better seen by other modalities. Examples are analysis of cracks in metals or detection of farms

in the South of Africa for the monitoring and management of the water resources.

Optical imagery, including color composites, is of very high interest for image analysts involved in surveillance and detection of objects of 100 m in size or less. Such images are comfortable because they offer a fair amount of similarity with what can be observed by a human eye from an airplane. Figure 6.7 exhibits an optical image of the international airport of Marseille (France) taken by the satellite SPOT in the panchromatic band. The pixel size is 10 m. The airport is partly built on a lake, which appears in dark tones in the upper left corner of the image. The runways are in clear tones; the airport is clearly visible.

Figure 6.7. SPOT image of the airport of Marseille, France. Copyright CNES - SPOT Image (1992).

Radar imagery is also used by image analysts. It offers unique advantages. Beyond its all-weather and night capabilities, it may image sharp objects of small size, even smaller than the pixel size.

An appropriate combination of optical and radar images permits the analyst to spot these objects into a familiar view of the landscape and ease and speed up the analysis of the satellite radar scenes. Applications are in the surveillance domain and may deal with humanitarian catastrophes, whether they are natural (e.g., floods) or man-made (e.g., wars, guerillas). The fused product may be used to detect the camps of refugees and to assess the

number of refugees, which helps in shaping and sizing the international aid resources and efforts.

Figure 6.8 exhibits a radar image obtained by the satellite ERS over the airport of Marseille. The two images are not contemporary, though both taken in 1992. The original size of the pixel is 12.5 m in the radar imagery. Both images have been geometrically aligned. The ERS image has been resampled for a better illustration of the method, with a pixel size of 10 m.

This radar image is more difficult to interpret than the optical one. There is a "salt and pepper" effect, called speckle. It is inherent to this type of ERS radar imagery. It makes objects difficult to distinguish. Indeed only the runways and taxiways appear clearly in this image in dark tones. White spots are due to strong echoes. They are very often related to large buildings but do not delineate them perfectly. The waves at the surface of the lake return the radar signal rendering impossible the detection of the shoreline.

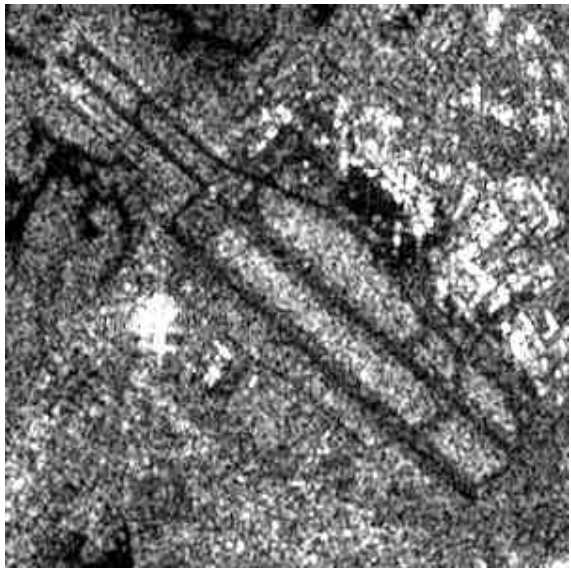


Figure 6.8. ERS image of the airport of Marseille, France. Copyright ESA (1992)

All these comments may appear as drawbacks. Nevertheless the radar image has the capability of imaging objects of size smaller than the pixel size, provided they comprise square angles and are located in flat areas of low radar cross-section. Such elements are of interest in airports. The method described hereafter tends to make them more visible to the image analyst by fusing radar and optical images.

Figure 6.9 shows the scheme of the fusion method. A multiresolution analysis is applied to the SPOT image PAN using a wavelet transform "à trous". In this algorithm, only a scale function is used (see Chapter 5). This algorithm provides at each step one context and one non-directional wavelet coefficient image.

All images have the same size. Three iterations are performed, providing three images of wavelet coefficients C_{PAN1}, C_{PAN2}, C_{PAN3} and one image of context PAN_3. Segmentation is made by the means of a multiscale algorithm, using these four images as inputs. The parameters of the cost function are also inputs (rules). They are such that the segmentation process delivers selected areas of the images that are large and homogeneous. Here, homogeneous means that transitions and gradients are small relative to the surface of the area. Runways of the airport and the airfield are such selected areas; zones of buildings are excluded. The first fusion process provides a representation of the areas where elements of interest may be present.

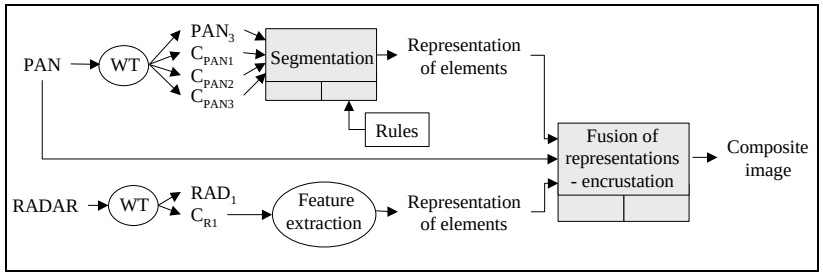


Figure 6.9. Scheme of the encrustation method for the example of Marseille airport

A multiresolution analysis is also applied to the radar image using one iteration of the algorithm "à trous". The images of wavelet coefficients always have histograms that are centered at zero and present a very sharp distribution. Figure 6.10 is a typical example of the histograms of the image of the wavelet coefficients for this example.

This histogram is assumed to be a representation of the probability density function of the wavelet coefficient image. It can be adjusted to a theoretical probability density function obeying a generalized Gaussian law.

The contents of the histogram can be interpreted as follows:

- the values close to zero (central peak of the histogram) represent noise or very small variation of the image;
- the high values of the histogram (left and right parts of the histogram) represent the strong variations of the image *i.e.* well marked structures as borders between different elements.

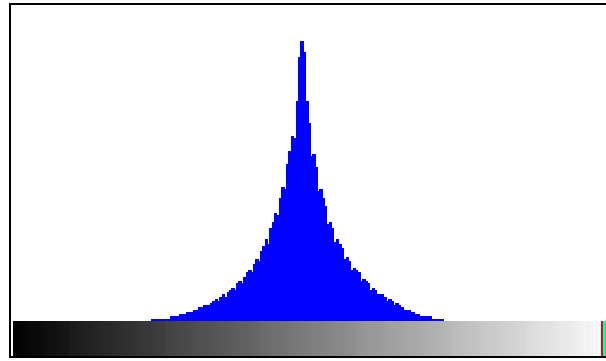


Figure 6.10. Typical histogram of the wavelet coefficients image representing the information at a given scale

The method for the detection of elements of interest is based on this observation. Recall that these elements present strong transitions and have sizes lower than the pixel size. A thresholding is applied on the wavelet coefficients. Only values that are greater than the absolute value of the threshold are kept. They are the well-marked structures of small sizes. This threshold depends on the application and on the size of the structures to detect (as well as the number of iterations of the wavelet transform) to be performed). Some tests are always needed to fit this value to the aim or it can be user-adjusted. The output of this process is a representation of the possible elements of interest.

The final stage is the fusion of the representations together with the visual marking of the retained elements into the panchromatic image PAN (encrustation). The fused product is presented to the image analyst for visual analysis. An enlargement of the resulting image is shown in Figure 6.11.

The retained possible elements of interest are encrusted as bright points. Three of them are shown into a circle. The objects that are represented by these features are actually much smaller than the pixel size (10 m). One is a vehicle: a standard 4 wheel drive vehicle, aiming at keeping the sea birds away from the runway. The second one is a power converter, a set of three cubes of 2 m in size made of metal. The third one is the ILS antenna, an aid to navigation. It is made of metal, very thin and of a few meters high.

Such a fusion of representations of elements derived from optical and radar images has an application in image visual analysis. Through an appropriate interface, the analyst may select in an interactive manner the range of scales of interest and the thresholds for the characterization of well-marked structures. The joint exploitation of the two sets of images is thus possible

and offers better performances than the combined results of individual exploitation.

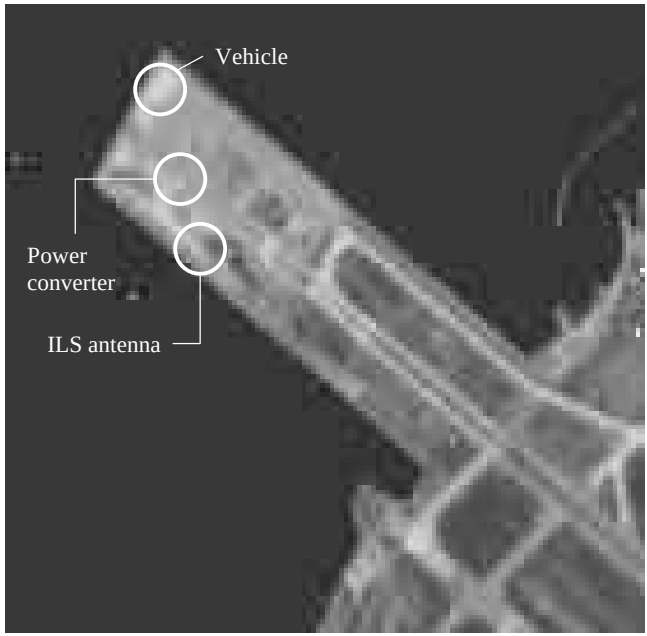


Figure 6.11. Enlargement of the panchromatic image with the encrustation of features extracted from the radar image. Some features are identified.

7. FUSION FOR THE SYNTHESIS OF IMAGES WITH A HIGHER SPATIAL RESOLUTION

INTRODUCTION

In various applications, the benefit of having images with the highest spectral resolution (or the largest number of relevant modalities) and the highest spatial resolution has been demonstrated. On the one hand, the high spatial resolution is necessary for an accurate description of the shapes, features and structures. On the other hand, depending on the application and the level of complexity of the observed scene, the different objects are better identified if high spectral resolution images are used. Hence, there is a desire to combine the high spatial and the high spectral resolutions with the aim of obtaining the most complete and accurate description of the observed scene.

However the sensors offer either high spectral resolution and low spatial resolution, or low spectral resolution (broadband) and high spatial resolution. Hence research has developed, which aims at proposing algorithms for fusing both types of images, in order to synthesize images with the highest spectral and spatial resolutions available in the sets of images.

These images should be as close as possible to reality and should simulate what would be observed by a sensor having the same modalities but the highest spatial resolution. The accurate synthesis of the multispectral character is very important to many applications, including those calling upon classification or the reproduction of the natural colors. Classification processes often use bases of spectra (multi-modality signatures), which result from measurements or simulations by models or from the experience of image analysts. In the course of the classification, the observed spectra are compared to the known ones and a decision is taken according to their similarities. Accordingly, any error in the synthesis of the multi-modality signatures at the highest spatial resolution induces an error in the decision.

This research is very vivid in Earth observation. Many methods have been developed. Some of them are used in the military domain. Synthetic products are also available by the providers of satellite images and the civil market is becoming more and more important. The concept is also gaining attention from instrument makers, especially for space-borne missions. By integrating fusion techniques in the processing software, the instrument makers can design instruments made of several sensors, each of them being well adapted to one aspect of the mission, e.g., one sensor with a high

spatial resolution and another with a high spectral resolution. Hence some trade-off may be avoided in the design. It results into lighter, cheaper and more reliable systems.

Some meteorological satellites are observing the Earth by means of passive sensors working in several bands of frequencies a few GHz (millimetric wavelengths). For instrumental reasons, the lower the frequency, the larger the size of the pixel. Presently all spectral data are fused after they have been resampled to the largest pixel size found in the set of images, *i.e.* approximately 100 km. The fusion provides estimates of rainfall, sea surface temperature and surface wind. Meteorologists would prefer to have such gridded parameters at a better spatial resolution, that is a smaller size of the grid cell. The realization of an instrument meeting their needs would be very costly. Fusion methods for synthesizing images in GHz frequencies with a better spatial resolution are an efficient low-cost alternative to the making of an instrument.

Several space-borne systems have dual sensors offering multispectral capabilities and low spatial resolution on one side, and a panchromatic band with a high spatial resolution on the other side. The SPOT system is one of them. It presents a panchromatic band with a spatial resolution of 10 m (SPOT-P) and three spectral bands XS1 (green-yellow), XS2 (red) and XS3 (near infrared) with a resolution of 20 m. Such multispectral images are very useful to map the different types of land use (e.g., fields, forests, roads...). The improved spatial resolution allows these features to be better delineated, meaning that the synthesized images are more useful for applications such as mapping, precision farming, surveillance, national security etc.

The Landsat space-borne system offers a panchromatic band with a spatial resolution of 15 m, six spectral bands located between blue and near-infrared with a spatial resolution of 30 m, and a thermal infrared band with a spatial resolution of 60 m. Several authors have used one or more of the six spectral bands, or the panchromatic image, to synthesize a thermal infrared image at a better spatial resolution. Recent Earth observation commercial missions (Ikonos, Orbview) provide broadband images (panchromatic) with a high spatial resolution of 1 m, and three multispectral images with a lower spatial resolution of 4 m, taken in the blue, green and red bands. Ikonos has an additional near infrared band at 4 m.

Customers have expressed a great interest in obtaining high spatial resolution landscapes with objects having their natural colors. Here, natural colors mean the colors that are perceived by the human eye. All these examples demonstrate that such fusion methods are of a large concern, outside the research community in mathematics.

These fusion methods are sometimes called "band sharpening". Care should be taken. "Band sharpening" may be performed to increase the utility of a set of images for visual analysis, while the synthesis of image aims at producing actual images of higher spatial and spectral resolutions.

Care should be taken with the term "spatial resolution". A spatial resolution is expressed in meters. Objects or features as wide as or bigger than the resolution can be distinguished in the image. The size of the pixel of the original image is most often equal to the spatial resolution of the instrument. However it is often said that the lower the size of the pixel, the better the resolution. A high resolution h (e.g., 1 m) means a small pixel size, while a low resolution l (e.g., 5 m) means a larger pixel size. Nevertheless, it remains that $h < l$, *i.e.*, that the resolution of 1 m is less than the resolution of 5 m. This is certainly confusing but there is a general consensus on using the term "resolution" in this way. One may add that the mathematicians do not care of the size of the pixel in meters. For example, in the multiresolution analysis (see Chapter 5), the original resolution of the image is set to 1. Then in the course of the analysis, and assuming a dyadic case, the resolution will be successively equal to 1/2, 1/4, 1/8 etc.

THE GENERAL PROBLEM

Let denote the acquired images of lowest spatial resolution by B_l, and the images of highest spatial resolution by A_h. The subscripts l and h denote the spatial resolution of images B or A, *i.e.*, low and high resolution, respectively. $B^{interp}{}_h$ denotes the result of the interpolation (resampling) of B_l from resolution l to h. Each set of images A and B is composed of several images acquired by various modalities, e.g., panchromatic, X-rays, electron microscope, nuclear-magnetic resonance, taken at different times and with different times of integration and have different space resolutions. Within each set, the images are geometrically aligned and have the same pixel sizes. Within the set B, B_{kl} denotes the image acquired by the modality k (hereafter called the spectral band k). In the following, for the sake of the simplicity, the set A is assumed to have only one image A_h, unless mentioned otherwise. The problem may be easily extended to a set A comprising several images or to more than two sets of images.

The fusion methods aim at constructing synthetic images $B^*{}_h$, which are close to reality. The methods should perform a high-quality transformation of the multispectral content of B_l, when increasing the spatial resolution from l to h.

The general problem is relevant to the fusion of representations and is the creation of a new set of images B^* from the original representations A and B:

$$B^* = f(A, B)$$

In addition, these synthetic images B^* must respect the three following properties.

First property. Any synthetic image B^*_h once degraded to its original resolution l, should be as identical as possible to the original image B_l, that is

$$D_1(B_{kl}, B^*_{kl}) < e1_k \quad [7.1]$$

where D_1 is the distance between B_{kl} and B^*_{kl}. Approximation induced by the resampling of B^*_{kh} into B^*_{kl} should be taken into account: the limit $e1_k$ is determined by the requested degree of accuracy. $e1_k$ should be small for all modalities; this ensures the similarity between the sets B_l and B^*_l. An example of D_1 is the square root of the mean of the squared differences $(B_{kl} - B^*_{kl})$ on a pixel basis. A typical value for $e1_k$ is 0.05 times the mean value of B_{kl}. Depending upon the objectives, other distances may be used in order to enhance specific properties in the image, e.g., structures or shapes.

Second property. Any synthetic image B^*_h should be as identical as possible to the image B_h that the corresponding sensor would observe with the highest spatial resolution h, if existent:

$$D_2(B_{kh}, B^*_{kh}) < e2_k \quad [7.2]$$

where D_2 is the distance between B_{kh} and B^*_{kh} for the modality k.

The second property does not imply an accurate synthesis of the multi-modality properties of the set B when increasing the spatial resolution. This should be an additional property.

Third property. The multispectral (or multi-modality) set of synthetic images B^*_h should be as identical as possible to the multispectral (or multi-modality) set of images B_h that the corresponding sensor would observe with the highest spatial resolution h, if existent:

$$D_3(B_h, B^*_h) < e3 \quad [7.3]$$

where D_3 is the distance between the sets B_h and B^*_h.

Several methods have been published. They differ in the way they respect the three properties. Considering the methods that are used and known, one may distinguish three groups of methods: the projection and substitution methods, the relative spectral contribution methods and the methods relevant to the ARSIS concept. Evidently, there are some hybrid methods belonging to more than one group. These three groups are discussed in the following sections.

The problem may be seen as the inference of the information that is missing to the images B_{kl} for the construction of the synthesized images B^*_{kh}. The missing information is linked to the high frequencies in the representations

A and *B*. Since this information is to be inferred from the modalities in *A*, a relation should exist between the modalities in *B* and at least one of the modalities of *A*, relative to the high frequencies. The images of the sets *A* and *B* do not need to be commensurate. Some studies have been published where images acquired in thermal infrared bands have been synthesized with a better spatial resolution with a satisfactory quality by the means of images acquired in the visible range.

The geometrical superimposability of images is usually of importance for such fusion methods, especially since they are dealing with the addition / combination of high frequencies. The images B_l and A_l should be geometrically aligned, as discussed in previous Chapter, once all images are set to the lowest available spatial resolution. Some acquisition systems provide images of different spatial resolutions that are already co-registered. Otherwise this can be done by means of standard methods available in public or commercial software packages for image processing. Some providers of images arrange for their products to be co-registered. The images of lowest resolution B_l are projected into the geometry of the images of highest resolution degraded to the lowest resolution A_l. During the process, a resampling of the multispectral images *B* is made. The resampling operator has an influence upon the final result. In most cases, a bi-cubic interpolator offers a good compromise between the accuracy of the result and the required computer time.

A few authors have assessed the influences of respectively the quality of the co-registration and the resampling operator on the final results. The relative discrepancies between the results are a few per cent; these influences can be kept very small provided the co-registration is accurate enough and the operator is appropriate enough. In the following, for the sake of the simplicity, the term "image of lowest resolution" B_l will denote the projected resampled image of lowest resolution.

The images *A* and *B* may not have been acquired simultaneously. Changes between the two sets may arise from

- the fact that the images are acquired by different modalities;
- the differences in spatial resolution;
- the differences between illumination or acquisition conditions;
- the changes in the observed scene itself.

As for the two latter causes, as long as the time-lag is small with respect to the time scale of the variations in small-size features of each cause, its influence upon the quality of the transformation of the spectral content when enhancing the spatial resolution (*i.e.* the synthesis of B^*_h) is low or negligible. Such a time scale is greatly variable; it depends upon the objects themselves as well upon the spatial and spectral resolutions with which they

are observed. If the time lag is large, the user must weight its consequences. He should know precisely the merging method to be used, because all methods do not take into account in the same way the small structures to be injected from the images of highest resolution into the images of lowest resolutions. This point is briefly discussed in the following section and more extensively in Chapter 9. An analysis of the influence of a large change with time in the observed scene upon the fused product is made for several methods. It is illustrated by the case of the building yard of the gigantic river dam of the Three Gorges in China. Several sets of images are available, with time lag ranging from a few months to several years, showing the constantly evolving yard. This study demonstrates how outputs of the methods may be differently affected by the time lag.

THE SPOT IMAGES

For a better understanding, the fusion methods are illustrated by the case of the SPOT system. As said before, this system presents two sets of images:

- one panchromatic image, called *P* or *PAN*, with a spatial resolution of 10 m and a wide spectral band (from 450 to 810 nm);
- three other images, called *XS1*, *XS2* and *XS3*, with a spatial resolution of 20 m only, i.e. twice less than that of *P*. Each spectral band *XS* (modality) has a narrower spectral window than the band *P*. Hence this second set *B* offers a better spectral resolution.

The peak-normalized spectral response functions of these bands are denoted $S_k(l)$, where l is the wavelength and k the spectral band (P, XS1, XS2 or XS3). They are displayed in Figure 7.1.

The equivalent radiance L_k in the band k is

$$L_k = \frac{\int L(l)\, S_k(l)\, dl}{\int S_k(l)\, dl} \qquad [7.4]$$

where $L(l)$ is the spectral density of radiance (W m^{-2} sr^{-1} μm^{-1}).

The bands XS1 and XS2 are in the visible range, as well as the panchromatic band P of course. The band XS3 is in the near infrared domain outside the visible range. The peak-normalized spectral response function for the half-sum of the bands XS1 and XS2 is drawn in Figure 7.4.

Of interest to the discussion are also the following points:

- the equivalent radiances L_{XS1} and L_{XS2} are similar for a spectrally uniform reflector: $L_{XS1} \sim L_{XS2}$;
- the average value of the equivalent radiances of the bands XS1 and XS2

is very close to the equivalent radiance of the band P for a spectrally uniform reflector: $L_{(XS1+XS2)/2} \sim L_P$;

- the band P is covering more near infrared wavelengths than does the band XS2 (approximately 750 to 810 nm);
- the bands XS3 and P are separate.

Finally note that this is valid for the systems SPOT-1, -2 and -3. In the system SPOT-4, the bands XS are noted B (B1, B2, and B3). There is an additional B4 band in the near infrared range. The spatial resolution of the bands B1, B3 and B4 is 20 m. There is no panchromatic band and the band B2 has a spatial resolution of 10 m. In this case, the set *A* is comprised of the image *B2* and the set *B* of three images: *B1*, *B3* and *B4*. Contrary to the previous systems, there is no spectral overlap between the sets *A* and *B*, except for a very small one between the bands B1 (XS1) and B2 (XS2) as shown in Figure 7.1.

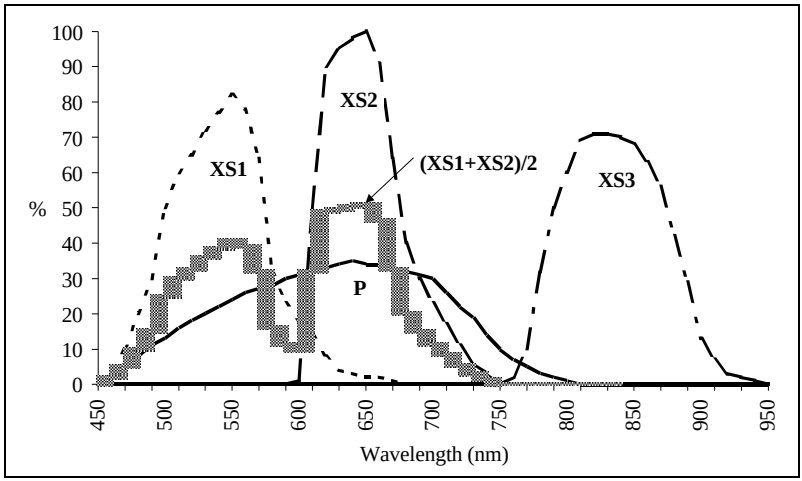


Figure 7.1. Relative spectral responsivity of the bands of the instrument HRV of the SPOT system (SPOT 1, 2 and 3)

PROJECTION AND SUBSTITUTION METHODS

These methods have already been discussed in previous Chapter. The set of images B^{interp}_{kh} is projected into another space, where one of the components exhibits most of the structures present in the set B_l. This component is replaced by the image A_h. The inverse projection is performed and the synthetic images B^*_{kh} are obtained. In the case of the method IHS, the component to be replaced is the intensity I_l. Prior to substitution, the three

components IHS are interpolated to the resolution h and the component I^{interp}_{h} is replaced by the image A_h.

The substitution component is usually A_h but may be a linear combination of A_h and of the component to be replaced (or any other function). From a practical point of view, the construction of the new components in the projection space requests that the images B^{interp}_{kh} exhibit similar encoding dynamics. This constraint may induce spectral distortion in the synthesis of B^*_{kh}. The similarity in encoding dynamics is also necessary between the substitution component and the component to be replaced. This alignment procedure is often performed by matching histograms.

Among these projection and substitution methods, the IHS and PCA methods are the most used.

Some refinements were proposed making use of the wavelet transform[1]. Instead of replacing the intensity I^{interp}_{h} by the image A_h, it is possible to synthesize, in the sense of the multiresolution analysis, the intensity I_h at resolution h, by adding the appropriate wavelet coefficients extracted from A_h to I_l. Then, the inverse projection is performed to obtain the synthetic images B^*_{kh}. In this approach, the wavelet transform "à trous" is the easiest tool to use. This approach also holds for the PCA method. It has the advantage to respect the first property, while the original method does not.

To achieve good results using such techniques from the point of view of performing a high-quality transformation of the multispectral content when increasing the spatial resolution, it is necessary that

- the component to be replaced comprises a very high percentage of the information related to the high frequencies structures;
- the correlation between the component to be replaced and the substitution component is very high, for all frequencies and not only for the highest frequency.

Whatever the method, looking to the equations makes it obvious that, except in rare cases, such methods cannot achieve high-quality transformation of the multispectral content when increasing the spatial resolution from l to h. Maybe the clearest demonstration of this is given by looking to the equations, when reducing (e.g., resampling) the spatial resolution of the synthesized images B^*_{kh} back to the original resolution of the set B, that is looking to the first property. It appears that the influence of the high resolution image A_h is not limited to the high frequencies that have

---

[1] J. Núñez, X. Otazu, O. Fors, A. Prades, V. Palà, and R. Arbiol. *Multiresolution-based image fusion with additive wavelet decomposition*. IEEE Transactions on Geosciences and Remote Sensing, 37(3), 1204-1211, 1999.

been injected to increase the resolution; it covers the whole set of frequencies that are present in A_h. Hence the substitution component may include frequencies that should not appear in a given band. This comment explains the two constraints for success listed above.

Figure 7.2 illustrates the IHS method. The case is a part of the city of Riyadh (Saudi Arabia). A color composite of a SPOT XS sub-scene acquired on May 16, 1993 (upper left image) is displayed. In this picture, one can clearly see the large interchange of two urban highways (middle left). The highways are in black. The interchange is enhanced by the presence of vegetation (here in red). Lots in the sandy areas (upper part) are ready for further constructions of buildings; the network of asphalt streets is perfectly visible. On the contrary, details in the housing in the lower half part cannot be seen. The large white-blue rectangular shape in the lower left part below the interchange is a mall.

The color composite is performed by a dynamic allocation of color codes to color classes respective to the frequencies of the triplets. The coding for the three color composites is the same. Their colors can be compared; they represent the multispectral information contained in the set of images $B^{interp}{}_h$ or $B^*{}_{h(IHS)}$ or $B^*{}_{h(Model2)}$.

The upper right corner of Figure 7.2 exhibits a panchromatic image taken by the Russian camera KVR-1000 on September 7, 1992. This image A_h has a spatial resolution of 2 m and has been acquired more than eight months before the SPOT scene. It displays much more details than the SPOT XS image. See for example the details in the mall. But there are also two striking features. Firstly, the details of the highway interchange do not appear in the KVR image likely because of some saturation and defects in the KVR film, which has been digitized. Secondly, the lots in the sandy areas were not finished at that date: a few asphalt streets can be seen and the others are sandy paths that are hardly visible in the KVR image.

Here the ratio between h and l is 10. The images in the lower half of Figure 7.2 are color composites of the three synthetic images $B^*{}_{kh}$ obtained by two different fusion methods: the IHS method and the ARSIS-Model2 method, which is discussed later.

In the IHS composite (lower left) more details appear compared to the original XS color composite (see e.g., the mall). However the IHS procedure has introduced two large visible defects: the highway interchange is less visible because of the defects in KVR and the network of asphalt streets in the lots is no longer visible because of the time lag between both images. This means that if one has to interpret this fused product, one would falsely conclude that this part of the district is still under work and that the streets are not ready for car traffic or building construction.

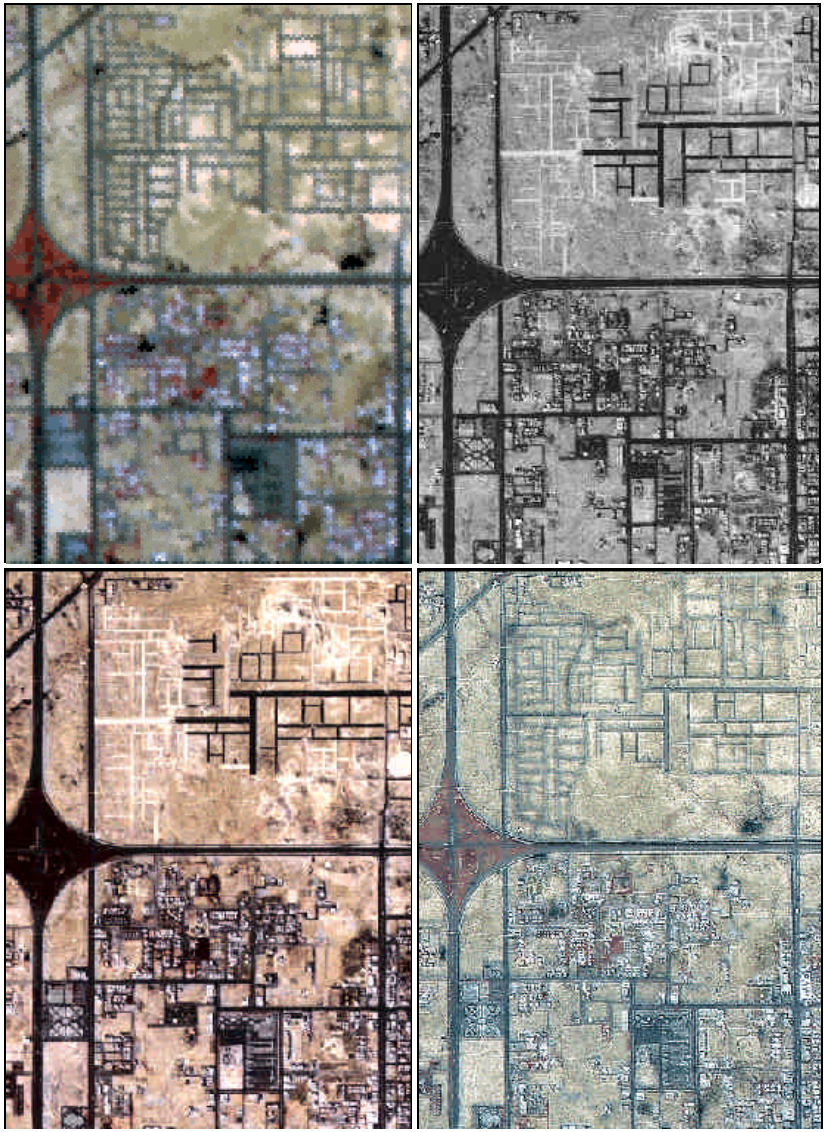


Figure 7.2. Scenes of the city of Riyadh (Saudi Arabia). From left to right, top to bottom: SPOT-XS color composite (spatial resolution is 20 m, © CNES SPOT-Image 1993), KVR-1000 panchromatic image (spatial resolution is 2 m, © Sovinformsputnik 1992), and color composites of synthetic product: IHS and ARSIS-Model2 methods.

This image illustrates the fact that the structures seen in the fused products B^*_{kh} output from the IHS method are those present in the substitution component, and only them. This clearly demonstrates that the fused products will not respect the first property or the second one.

As for the third property, one has to look to the colors of the IHS fused product and compare them to the colors of the composite of the original XS images. One should not expect these colors to be identical. Since the synthetic method should only add high frequencies to the original images of the set *B*, one may expect the colors of the color composites B_{kl} and B^*_{kh} to be fairly close. This is not the case at all for the IHS fused product. The color composite should exhibit more bluish tones.

In the color composite (lower right) made from the products synthesized by the ARSIS-Model 2 method, it is possible to distinguish the structures of the mall, and all the buildings in this area. The colors of this composite are close to that of the composite of the original XS images. It means that the statistical distribution of the spectra is similar between the sets B_l and $B^*_{hModel2}$. The change in the spectral content induced by the increase of spatial resolution is of higher quality than for the IHS method. The roads on the interchange, including the details of the lower left loop, are now visible. The asphalt streets that are not visible in the KVR image can be distinguished as they are in the original XS color composition. Contrary to the IHS method, the ARSIS-Model 2 method is capable of taking into account the changes in structures due to the time lag.

RELATIVE SPECTRAL CONTRIBUTION

In this group of methods, the relationships between the various modalities are exploited. The P+XS method is one of these methods. It is presented first because it fully illustrates this concept in a pragmatic manner.

THE P+XS METHOD

This method has been devised by the CNES, the French space agency, for the system SPOT to produce multispectral images with a resolution of 10 m^2. It is founded on the assumption that the half-sum of the radiances in XS1 (modality *1*) and XS2 (modality *2*) is equal to the radiance in the panchromatic band P (Fig. 7.1). It can be applied to other systems provided their modalities obey the two assumptions made.

---

2 Anonymous. *Guide des utilisateurs de données SPOT*, 3 tomes, Editeurs CNES et SPOT-Image, Toulouse, France, 1986.

Assume that L_{Ah}, L^*_{1h} and L^*_{2h} are the radiances of a 10 m pixel in respectively the image A_h and in the fused products B^*_{1h} and B^*_{2h}. The previous assumption is:

$$L_{Ah} = \frac{L^*_{1h} + L^*_{2h}}{2} \qquad [7.5]$$

A second hypothesis is necessary that describes the distribution of the radiance L_{Ah} between L^*_{1h} and L^*_{2h}. The P+XS method assumes that the ratio of the radiances in both bands is constant with the change in resolution from *l* to *h*:

$$\frac{L^*_{1h}}{L^*_{2h}} = \frac{L^*_{1l}}{L^*_{2l}} \qquad [7.6]$$

This assumption is entirely valid only if the pixel of size 20 m is composed of four pixels of size 10 m having all the same spectral behavior. The final equations are then:

$$L^*_{1h}(x,y)_h = 2L_{Ah}(x,y)_h \frac{L_{1l}}{L_{1l} + L_{2l}}(x0,y0)_l$$

$$L^*_{2h}(x,y)_h = 2L_{Ah}(x,y)_h \frac{L_{2l}}{L_{1l} + L_{2l}}(x0,y0)_l \qquad [7.7]$$

where $(x0, y0)_l$ is the pixel in the set B_l corresponding to the pixel $(x, y)_h$ in the set A_h. These formulae produce the synthetic images B^*_{1h} and B^*_{2h}. They do not apply to the band XS3 of the SPOT system because there is no overlap between the bands P and XS3 (Fig. 7.1). The synthetic image B^*_{3h} is created by a simple duplication of pixels of the image B_{3l}. There is no fusion process in this particular case.

RELATIVE SPECTRAL CONTRIBUTION

The method "relative spectral contribution", which applies to radiances, restrains itself to the sub-set of the *N* spectral bands B_k lying within the spectral range of the A_h image. Furthermore, it is assumed that

$$A_l \approx \sum_{k=1}^{N} B_{kl} \text{ for } k \text{ belonging to the sub-set.} \qquad [7.8]$$

Actually, this may happen in radiances. Anyway, this equation may be generalized, and the components B_{kl} may be adjusted in digital counts (gray levels) for this equation to apply. An intermediate multispectral image, B''_{kh}, is computed:

$$B''_{kh} = \frac{B^{interp}_{kh} A_h}{\sum_{j=1}^{N} B^{interp}_{jh}} \quad [7.9]$$

where B^{interp}_{kh} is a resampled version (e.g., cubic interpolation) of B_{kl} at the higher resolution h and assuming that $\sum_{j=1}^{N} B^{interp}_{jh}$ is not equal to zero.

To obtain the final result (i.e. the multispectral image B^*_{kh} at the resolution h), the intermediate multispectral image B''_{kh} has to be adjusted in such a way that the mean value of each component of the synthesized multi-modality image B^*_{kh} is the same than the mean value of each original component B_{kl}. Therefore:

$$B^*_{kh} = B''_{kh}\, m(B_{kl}) / m(B''_{kh}) \quad [7.10]$$

where $m(B_{kl})$ and $m(B''_{kh})$ are the mean values of the images B_{kl} and B''_{kh} computed over all pixels, assuming that $m(B''_{kh})$ is not equal to zero.

The Brovey transform is a simplification of this method, where the last step (adjustment of mean value in Equation 7.10) is omitted. It is found in several commercial softwares, but beyond its own limitations, is not always well implemented. Ideally, it should deal with radiances and the software should request the calibration coefficients, while it is often performed on gray levels. This usually causes additional distortions in the spectral content.

The "color normalized" method is a version of the Brovey transform. It is in use in the Department of Defense in the USA. It is defined for three images B_k, preferably in true colors, and one panchromatic image A_h:

$$B^*_{kh} = \frac{3(B^{interp}_{kh} + 1)(A_h + 1)}{3 + \sum_{j=1}^{3} B^{interp}_{jh}} - 1 \quad [7.11]$$

The values are image gray levels. The dynamics of each image: A_h and B^{interp}_{kh}, for $k=1, 3$, should be similar. The small additive constants prevent division by zero. The spectral content is usually incorrectly synthesized, when increasing the spatial resolution.

Another type of adjustment of the intermediate image B''_{kh} is sometimes proposed. The adjustment is a function of the mean value found for the spectral class under concern. The interpolated multi-modality images B^{interp}_{kh} are classified into M classes by e.g., an unsupervised classification. For each class C_c ($c \in [1, M]$), and each modality k, the mean value $m(B_{kl})_c$ is computed. Once the image B''_{kh} is computed according to the Equation

7.9, its dynamics is adjusted so that the mean for class C_c and modality k of the synthesized component B^*_{khc} is the same than for the original component:

$$B^*_{khc} = B''_{kh}\, m(B_{kl})_c / m(B''_{kh})_c$$

assuming that $m(B''_{kh})_c$ is not equal to zero. Assuming that every pixel (x, y) of the image belongs to one class and only to one, the synthesized image B^*_{kh} is given by:

$$B^*_{kh}(x, y, c_0) = \sum_{c=1}^{M} B^*_{khc}(x, y)\, d(c - c_0) \qquad [7.12]$$

where d is the Dirac distribution. Of course, B^*_{kh} is less continuous than B_{kl}; the larger the number of classes, the more continuous B^*_{kh}.

The method "relative spectral contribution" imposes the mean of the synthesized image to be equal to that of the original image. This is a major drawback since observations show that, for the same landscape or scene, the mean of an image is not invariant with the spatial resolution. This remark holds for the class-specific adjustment version.

THE GENERALIZED RELATIVE SPECTRAL CONTRIBUTION

The relative spectral contribution method can be generalized by relaxing the constraint imposed by the Equation 7.8. This equation becomes

$$A_l \approx \sum_{k=1}^{N} a_k\, B_{kl} \text{ for } k \text{ belonging to the sub-set.} \qquad [7.13]$$

where a_k is the ratio of the integral of the spectral band k and of the integral of the spectral band A. This ratio depends upon the sensor (see e.g. the case of the SPOT system given above, Fig. 7.1). Then

$$B^*_{kh} = \frac{B^{\text{interp}}_{kh}\, A_h}{\sum_{j=1}^{N} a_j\, B^{\text{interp}}_{jh}} \qquad [7.14]$$

The P+XS method is an example of the generalized method.

Instead of using the integrals of the spectral windows, some authors have proposed to define a_k by linear regression over targets, whose spectra are known. Assume that A_h is a multispectral image, whose spectral bands are much larger than the N spectral bands B_k and are covering these bands B_k. Checking the first property is the first constraint imposed on the synthesis method:

$$B^*_{kl} = B_{kl}\,,\ k = 1...N \qquad [7.15]$$

The second constraint is

$$A_{jh} = (1 / D\lambda_j) \int_{\Delta\lambda_j} B^*_{\lambda h}\, d\lambda \qquad [7.16]$$

where λ is the wavelength and $D\lambda_j$ is the spectral band j of the set of images A_h. This creates a number of equations with a larger number of unknowns.

To solve the problem, it is assumed that

$$B^*_{kh} = \sum_{j=1}^{M} a^i{}_j \; f_i(\lambda) \qquad [7.17]$$

where $f_i(\lambda)$ are orthogonal functions and M the dimension of the space. Using models, one simulates several spectra under a variety of conditions. Then a base of orthogonal functions f_i is defined by the means of, e.g. a principal component analysis, to represent this space of spectra. One defines M according to the requested accuracy.

Equation 7.13 holds on a physical basis as far as the spectral range covered by the N bands is the same than that covered by A. If there is a gap between two bands B_l, or if the N bands do not cover the whole band of A, then some objects having a strong spectral signature in this gap will be unnoticed in the bands B_l but will be present in A_h. Because of the construction of the images B^*_{kh}, these objects will falsely appear in these bands.

The relative spectral contribution methods, generalized or not, do not tell what to do when the spectral range of the modality k in the set B lies outside the spectral range of the modality A. The P+XS method recommends a simple duplication of pixels values for the images *XS3*, whose spectral band lies outside the range of the panchromatic band P. The duplication has the property of not changing the spectral content of the original images B_{kl}. However, other interpolation methods, including bi-cubic resampling techniques, of these images from the resolution l to h provide better results regarding the spatial and spectral aspects. The resulting images respect the first property. These methods are recommended instead of the duplication technique.

The relative spectral contribution methods, generalized or not, usually induce a spectral distortion during the synthesis. This distortion occurs at all scales, thus making the synthesized image B^*_l different from the original image B_l. This can be easily seen from Equation 7.7 in the P+XS method. The B^*_{1h} (*i.e.*, L^*_{1h}) image is function of the B_{1l} (L_{1l}) image, but also of B_{2l} (L_{2l}) and A_h (L_p). The influence of the latter images is not limited to the high frequencies but covers the whole spectrum of frequencies. It creates a stronger relationship between synthesized images than that existing between original images. This can be formally demonstrated by the means of a Fourier transform. The influence of the other spectral bands on the

synthesized image may range from low to high, depending mostly upon the landscape and also of the modulation transfer functions of the sensors. If the representation of an object is fairly close in the different spectral bands, the influence will be low.

A further drawback is that these methods cannot resolve local anti-correlations between spectral bands with a high accuracy in the synthesizing of the spectral content. Using additional equations and calling upon additional knowledge may prevent these problems.

THE ARSIS CONCEPT

The ARSIS concept is based upon the use of multiscale techniques in order to inject the high frequencies that are missing into the images of lowest resolution. Here multiscale techniques refer to mathematical tools, which are calling upon convolution and image filtering. These tools perform a hierarchical description, modelling and synthesis of the information content relative to spatial structures in an image.

The ARSIS concept assumes that the missing information is linked to the high frequencies in the representations *A* and *B*. It searches a relation relating these high frequencies and models this relation. A method belonging to the ARSIS concept performs typically the following operations: *i)* the extraction of a set of information from the set *A*, *ii)* the inference of the information that is missing to the images B_{kl} using this extracted information and *iii)* the construction of the synthesized images B^*_{kh}.

The methods developed within this concept respect the first property by construction though it depends upon the techniques used. The two other properties may be included in their design.

The concept is called ARSIS, after the acronym of its French name "*amélioration de la résolution spatiale par injection de structures*" (improvement of spatial resolution by structure injection)[3]. It has been designed in a generic way, transcending the mathematical tools used for its implementation.

---

[3] M. Mangolini, T. Ranchin, and L. Wald. *Procédé et dispositif pour l'amélioration de la résolution spatiale d'images à partir d'autres images de meilleure résolution spatiale*. French patent n° 92-13961, 20 novembre 1992, and T. Ranchin. *Applications de la transformée en ondelettes et de l'analyse multirésolution au traitement des images de télédétection*. Thèse de Doctorat en Sciences de l'Ingénieur, Université de Nice-Sophia Antipolis, 146 p., 1993.

The ARSIS concept makes use of a multiscale analysis for the description and the modeling of the missing information between the images A_h and B_l. The multiscale method mostly used for its various implementations is the multiresolution analysis, together with the wavelet transform (see Chapter 5 for a description of these mathematical tools).

Most examples of practical implementation of the concept ARSIS found in the literature are based on a multiresolution pyramidal approach[4]. Other tools for the multiscale analysis exist that have been used in the same purpose by various authors. Examples are filter banks instead of wavelet transform, gaussian filters, generalised Laplacian pyramid[5] or the second derivative of an apodisation function.

Figure 7.3 illustrates the ARSIS concept in the case of a multiresolution pyramidal approach. The multiresolution analysis is applied to the two images *A* and *B*. A scale by scale description of the information content of both images is thus obtained. The high frequencies between A_h at the resolution *h* (the bottom of the left pyramid) and A_l at the resolution *l* (the first level of the pyramid) are represented by the wavelet coefficients (the details). As seen in Chapter 5, the multiresolution analysis is such that given these wavelet coefficients and the image A_l, one may synthesize in an exact manner the image A_h.

In a similar manner, if the wavelet coefficients were available between the resolutions *h* and *l* for the right pyramid, and starting from B_l, one would be able to synthesize in an exact manner the image B_h (the dotted line at the bottom of the right pyramid). Since these wavelet coefficients for the image *B* are unknown (right pyramid), one solution to this problem consists in inferring them from the wavelet coefficients of the image *A* (left pyramid). Then, the synthetic image B^*_h may be constructed.

The missing information to be injected in the pyramid *B* from pyramid *A* is located in the missing bottom of the pyramid *B* (dotted line). Only this part is needed to improve the spatial resolution of the image B_l. But, if the missing information is set equal to that provided by the image *A*, the synthesized image B^*_h will not be equivalent to "what would be seen by the sensor B if it has the spatial resolution of the sensor A". Hence, in order to improve the quality of the synthesized image, a transformation should be

---

[4] See a review in T. Ranchin, and L. Wald. *Fusion of high spatial and spectral resolution images: the ARSIS concept and its implementation*. Photogrammetric Engineering and Remote Sensing, 66(1), 49-61, 2000.

[5] B. Aiazzi, L. Alparone, S. Baronti, R. Carlà, and L. Mortelli. *Pyramid-based multi-sensor image data fusion*. Wavelet Applications in Signal and Image Processing, Proceedings SPIE Conference, vol. 3169, pp. 224-235, 1997.

applied to convert the information provided by the multiscale representation of image *A* into the information needed for the synthesis of image *B*.

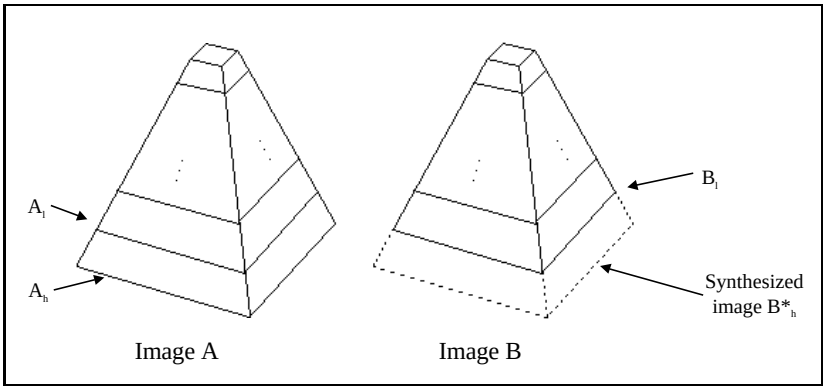


Figure 7.3. The use of the multiresolution pyramidal approach for the fusion of high spatial and spectral resolution images in the ARSIS concept

METHODS NOT CALLING EXPLICITLY ON THE MULTISCALE ANALYSIS

The HPF method, the methods of Pradines and Price and their derivatives and the LMVM method are examples of implementation of the concept ARSIS, which are not based explicitly on the multiscale analysis. They are relevant to the ARSIS concept because high frequencies are extracted from the image A_h by the means of moving windows and are injected into the images B_{kl} to synthesize images at higher resolution B^*_{kh}.

In addition, these methods have in common that the spectral behavior of the high frequency information is not taken into account. In other words, there is no model for the transformation of the high frequencies of image *A* into those of image *B*. This induces spectral distortion during the synthesis. The implicit hypotheses are *i)* that the correlation between A_l and B_l is large, ***ii)*** that this correlation is positive and *iii)* that these two hypotheses hold for A_l and B_l.

The HPF method

In the HPF method[6], a high pass filtering (HPF) is applied to the high spatial resolution image A_h in order to extract the high frequencies

[6] P. S. Chavez Jr., S. C. Sides, and J. A. Anderson. *Comparison of three different methods to merge multiresolution and multispectral data: Landsat TM and SPOT Panchromatic*. Photogrammetric Engineering & Remote Sensing, 57(3), 265-303, 1991.

representing the small structures between scales h and l. Then, these high frequencies are introduced in the multispectral image B_l by addition, which leads to the synthetic image B^*_h. The HPF filter is constructed by computing the second derivative of an apodisation function.

In the case of the images taken by the system SPOT with a ratio l/h of 2, this filter is a Laplacian filter. It is applied to the high resolution image, and its results are added to the low resolution images. The filter is a 3x3 matrix, and the coefficients are:

$$\begin{pmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{pmatrix} \qquad [7.18]$$

Refinements include an adjustment of the dynamics of the image A_h prior to the extraction of the high frequencies to adapt for the dynamics of B_l. This adjustment can be performed on the whole image or within the moving window.

The method of Pradines

In the method of Pradines[7], the relative spatial distribution of the high resolution signal is injected into the low resolution spectral image for each low resolution pixel. In the SPOT case, the 20 m pixel XS is shared in four 10 m pixels using the relative distribution observed in the image P for these four pixels (Fig. 7.4).

The relative spatial distribution may be described in several ways. The basic equation of this method is

$$B^*_{kh} = A_h \, B_{kl} / A_l \qquad [7.19]$$

where the ratio B_{kl} / A_l applies to the pixel at scale l containing the current pixel at scale h.

Several other authors have refined the method[8]. Good results are attained if the correlation between the images A_l and B_{kl} is high. The results are often noisy because the pixels at scale l are processed independently. One may filter the resulting images B^*_{kh} but this limits the benefit of increasing the

---

[7] D. Pradines. *Improving SPOT image size and multispectral resolution*. Earth Remote Sensing using the Landsat Thematic Mapper and SPOT Systems, Proceedings SPIE Conference, 660, pp. 78-102, 1986.

[8] See e.g., Liu J. G. and J. M. Moore. *Pixel block intensity modulation: adding spatial detail to TM band 6 thermal imagery*. International Journal of Remote Sensing, 19(13):2,477-2,491, 1998.

resolution. Another solution is to apply Equation 7.19 to interpolated images, that is

$$B^*_{kh} = A_h \, B^{interp}_{kh} / \langle A_h \rangle_{l/h} \quad [7.20]$$

where $\langle A_h \rangle_l$ is the average value of A_h performed on a moving window centered on the current pixel at resolution h and which size is the ratio l/h.

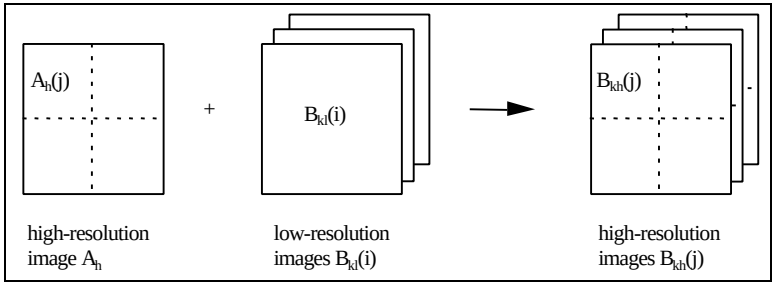


Figure 7.4. Principle of the method of Pradines

The local correlation modeling method and the Price method

These methods are similar and can be seen as an extension of the method of Pradines[9] [10]. Compared to that of Pradines, they offer the possibility of reproducing local anti-correlation between the images A_l and B_{kl}. A linear relationship is searched for between the moving windows centered on the current pixel at resolution l for both images A_l and B_{kl}. Without entering details, the equations are

$$B^*_{kh} = B^{interp}_{kh} + a \, (A_h - (A_l)^{interp}_h) \quad [7.21]$$

where the coefficients a and b are computed by linear regression over a moving window of variable size (typically, 3x3 or 5x5 low resolution pixels):

$$B_{kl} = a \, A_l + b \quad [7.22]$$

---

[9] J. C. Price. *Combining multispectral data of differing spatial resolution*. IEEE Transactions on Geoscience and Remote Sensing, 37(3), 1199-1203, 1999.

[10] C. Diemer, and J. Hill. *Local correlation approach for the fusion of remote sensing data with different spatial resolutions. In* Proceedings of the third conference "Fusion of Earth data: merging point measurements, raster maps and remotely sensed images", Sophia Antipolis, France, January 26-28, 2000, Thierry Ranchin and Lucien Wald Editors, published by SEE/URISCA, Nice, France, pp 91-98, 2000.

and where $(A_l)^{interp}{}_h$ is the image A_l interpolated at resolution h. The image $(A_l)^{interp}{}_h$ has no high frequencies.

The LMVM method

The LMVM method is built with respect to the first property, like the previous ones. Once brought back to the resolution l, the $(B^*_{kh})_l$ image reproduces the original mean of the image $(B_{kh})_l$.

The principle of the method LMVM (local mean and variance matching) is to adjust locally the mean and the variance of the image A to the same quantities of the image B and then to replace locally the image B by the image A[11].

The equations apply to a moving window. The working resolution is h and the images B_k are interpolated. The size s of the window is user-defined and expressed in pixels at resolution h. Noting $<A_h>_s$ and $stdev(A_h)$ respectively, the mean value and the standard deviation of A_h over that window, the equations are:

$$B^*_{kh} = (A_h - <A_h>_s)*stdev(B^{interp}{}_{kh}) / stdev(A_h) + <B^{interp}{}_{kh}>_s \qquad [7.23]$$

If $stdev(A_h) = 0$, then $B^*_{kh} = <B^{interp}{}_{kh}>_s$.

THE GENERAL SCHEME FOR MULTISCALE ANALYSIS

It is difficult to sketch the general scheme for the application of the ARSIS concept. In the methods above-mentioned, the modeling of the missing information from the image A to the image B is performed on moving windows of these images themselves. It is possible to focus more on the modeling of the missing high frequencies, expressed by Fourier coefficients or wavelet coefficients or other appropriate spatial transform.

Figure 7.5 presents the general scheme that applies on the case of use of a multiscale analysis. This case is used in the following for a better description of the ARSIS concept. Similar schemes can be drawn in other cases, where other tools or other strategies are used.

Inputs to the fusion process are the modality A at high spatial resolution (A_h, resolution n°1) and the modality B at low spatial resolution (B_{kl}, resolution n°2).

---

11 S. De Béthune, F. Muller, and J.-P. Donnay. *Fusion of multispectral and panchromatic images by local mean and variance matching filtering techniques*. *In* Proceedings of the second conference "Fusion of Earth data: merging point measurements, raster maps and remotely sensed images", Sophia Antipolis, France, January 28-30, 1988, Thierry Ranchin and Lucien Wald Editors, published by SEE/URISCA, Nice, France, pp 31-36, 1998.

Three models appear in this scheme. The Multi-Scale Model (MSM) performs a hierarchical description of the information content relative to spatial structures in an image. An example of such a model for remotely sensed images is the combination of the wavelet transform and multiresolution analysis (see Chapter 5). When applied to an image, the MSM provides one or more images of details, that is the high frequencies, and one image of approximation, that is the lower frequencies. The first iteration of the MSM on the modality *A* gives one image of the structures comprised between the resolution n°1 and n°2 (details image) and one image of the structures larger than or equal to the resolution n°2 (approximation image). The spatial variability within an image can thus be modeled. The Multi-Scale Model is built in such a way that it can be inverted (MSM^{-1}) to perform a synthesis of the high-frequency information.

The Inter-Modality Model (IMM) deals with the transformation of spatial structures with changes in modalities. It models the relationships between the details or approximation observed in the image *A* and those observed in the image *B*. This model may relate approximations and/or details for one or more resolutions and one or more modalities.

The High Resolution Inter-Modality Model (HRIMM) performs the transformation of the parameters of the Inter-Modality Model with the change in resolution. This operation is not obvious since many works have demonstrated the influence of the spatial resolution on the quantification of parameters extracted from imagery. To our knowledge, no particular attention was paid to this point, except for the model ARSIS-RWM[12] where a multiscale synthesis of the parameters of their model IMM from resolution n°3 to resolution n°2 is performed. Otherwise, the High Resolution Inter-Modality Model is often set identical to the IMM.

In this scheme, the operations are performed as follows. First, the MSM is used to compute the details and the approximations of image *A* (Step 1, Fig. 7.5). The same operation is applied to image *B* (Step 2). The analysis is performed for several resolutions, up to *n* in Figure 7.5 - that is (*n-1*) iterations for the analysis of the image *A* and (*n-2*) iterations for that of B_{kl}. These analyses provide several approximation images and several images of details for *A* and *B*.

---

[12] T. Ranchin, L. Wald L., and M. Mangolini. *Efficient data fusion using wavelet transforms: the case of SPOT satellite images*. In Proceedings of SPIE 1993 International Symposium on Optics, Imaging and Instrumentation. Mathematical Imaging: Wavelet Applications in Signal and Image Processing. San Diego, California, USA, July 11-16 1993, vol. 2034, pp. 171-178, 1994.

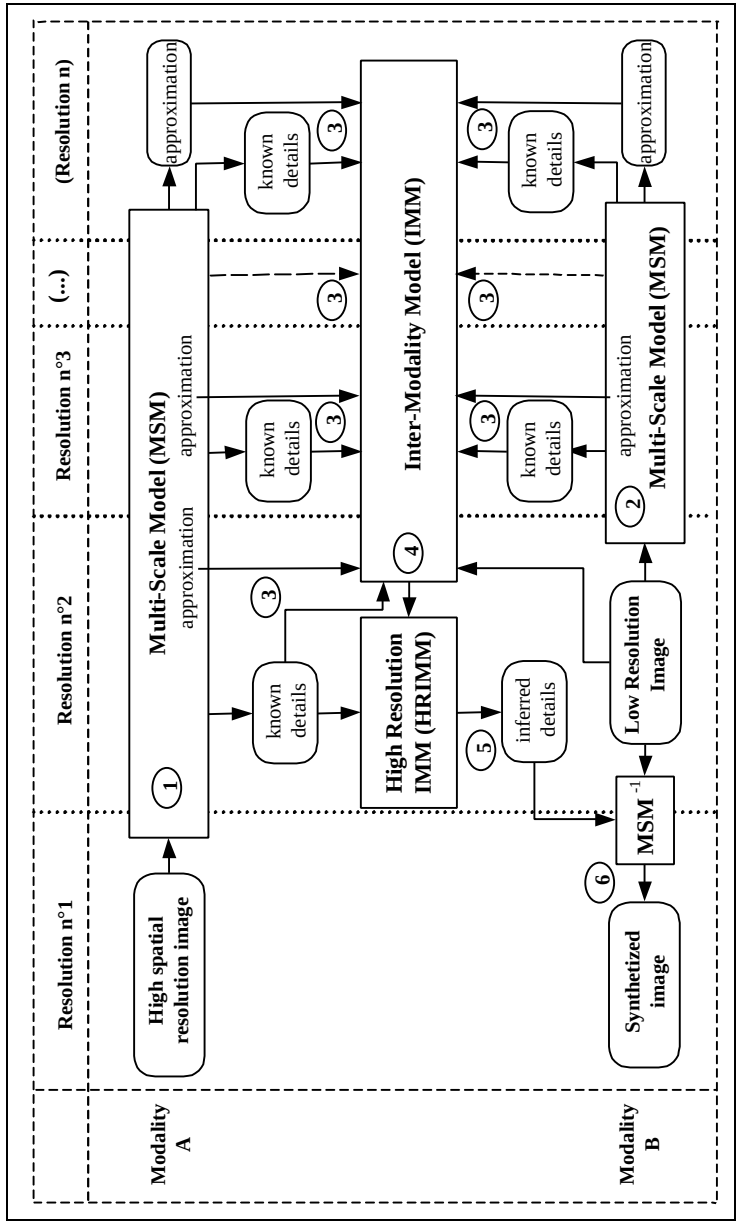


Figure 7.5. General scheme for the implementation of the ARSIS concept using a multiscale model (MSM) and its inverse (MSM^{-1}). See text for further comments

The approximations and the known details at each resolution are used to adjust the parameters of the Inter-Modality Model (Step 3). From the IMM is derived the High Resolution Inter-Modality Model (Step 4, at resolution n°2), which converts the known details of image A into the inferred details of image B_k (inferred details, Step 5). Finally, the inversion of the MSM (MSM^{-1}) from resolution n°2 to resolution n°1 performs the synthesis of the image B^*_{kh} (Step 6).

Figure 7.5 shows a case where the low resolution l (resolution n°2) is attained by the first iteration of the multiscale analysis applied to the high resolution h (resolution n°1). In the dyadic case, $l = 2\ h$. This is not always the case by far. The scheme can be easily drawn for the general case, where the image A_h is at resolution n°1, the image B_{kl} at resolution n°p and the synthetic product B^*_{kh} at an intermediate resolution n°q.

The scheme shown in Figure 7.5 applies to the case of sets A and B comprising several modalities. The scheme may apply to each modality of B separately. It may also apply to all modalities of B at the same time. In that case, the MSM is performed on each modality separately. Then, the multi-modality aspects are taken into account by the models IMM and HRIMM (Steps 3, 4 and 5). Finally, the inverse MSM applies to each modality separately.

Figure 7.6 details one possible scheme of the application of the ARSIS concept to the case of the SPOT imagery. The set of images is composed of a panchromatic image P at the spatial resolution of 10 m and three multispectral images *XS1*, *XS2*, *XS3* at the spatial resolution of 20 m.

The process is applied to each image *XSi* separately. Two iterations of the multiresolution analysis using the wavelet transform are applied to the original panchromatic image P and one iteration is performed on the original image *XSi*.

Inter-modalities models are computed for the transformation of each panchromatic wavelet coefficient image $C_P{}^D{}_{20\text{-}40}$, $C_P{}^V{}_{20\text{-}40}$, and $C_P{}^H{}_{20\text{-}40}$ into each wavelet coefficient image $C_{XS}{}^D{}_{20\text{-}40}$, $C_{XS}{}^V{}_{20\text{-}40}$, and $C_{XS}{}^H{}_{20\text{-}40}$ (see proposed models below). Then, these models are applied to the wavelet coefficient images $C_P{}^D{}_{10\text{-}20}$, $C_P{}^V{}_{10\text{-}20}$, and $C_P{}^H{}_{10\text{-}20}$ for the computation of the missing wavelet coefficient images $C_{XS}{}^D{}_{10\text{-}20}$, $C_{XS}{}^V{}_{10\text{-}20}$, and $C_{XS}{}^H{}_{10\text{-}20}$. Finally, the synthesis step reconstructs the high spatial resolution image *XSi* (XSi-HR).

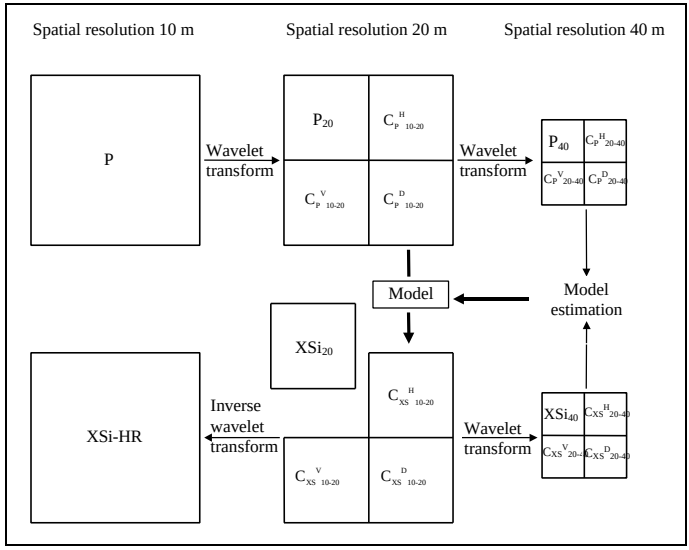


Figure 7.6. Application of the ARSIS concept to the SPOT imagery

THE INTER-MODALITIES MODELS

The inter-modalities model (IMM) is obviously a key point for an efficient synthesis. This model uses the available information to infer the missing details. The general form of the model is

$$C_{Bk}{}^{Z}{}_{h\text{-}l} = f(C_B{}^{Z}{}_n, C_A{}^{Z}{}_n, A_h, A_l, A_n, B_{jl}, B_{jn}...) \quad [7.24]$$

where

- $C_{Bk}{}^{Z}{}_{h\text{-}l}$, etc. are the details of the modality k of the set B for the scales comprised between h and l;
- Z = *D*, *V* or *H* if necessary, *i.e.* if the multiscale model is directional;
- n denotes the successive resolutions of the iterative multiscale model for scales greater than l;
- A_n the approximation of A at the resolution n and B_{jn} the modality j of the set B at resolution n.

More than the tools selected to perform the multiscale analysis, the IMM is what makes the difference between the various methods. Good results depend upon the accuracy of the IMM. Many models can be proposed. For example, the model RWM takes into account the physics of both images

and the relationship existing between the wavelet coefficients images[13]. Some authors used neural networks applied to the known wavelet coefficients at several scales to infer the parameters of the model. The IMM can have various forms and may take into account more than one scale, as shown in the general equation.

The simplest model (Model 1) is the identity model:

$$C_{Bk}{}^{Z}{}_{h\text{-}l} = C_{A}{}^{Z}{}_{h\text{-}l} \qquad \text{for Z = D, V or H} \qquad [7.25]$$

and in the case of the SPOT system

$$C_{XS}{}^{Z}{}_{10\text{-}20} = C_{P}{}^{Z}{}_{10\text{-}20} \qquad \text{for Z = D, V or H} \qquad [7.26]$$

Hybrid versions can be devised by adjusting the dynamics of the images A_h and B_{kl} prior to the multiscale analysis. The Model 1 is that implicitly used in the HPF and Pradines methods. It does not take into account the multi-modalities differences in high frequencies information between the image A and the images B_k and gives poor results. This can be seen directly from the equations but also from experiments. The following models are more accurate.

Assume that p is the ratio of two successive scales in the multiscale analysis. In the dyadic case, $p=2$.

The following model (Model 2) is based on the adjustment of the means and variances of the details images computed between the scales *(ph)* and *(pl)*. The parameters $a_k{}^{Z}$ and $b_k{}^{Z}$ are defined as:

$$a_k{}^{Z} = \sqrt{\frac{v(C_{Bk}{}^{Z}{}_{ph-pl})}{v(C_{A}{}^{Z}{}_{ph-pl})}}$$

$$b_k{}^{Z} = m(C_{Bk}{}^{Z}{}_{ph\text{-}pl}) - a_k{}^{Z}\, m(C_{A}{}^{Z}{}_{ph\text{-}pl}) \qquad \text{for Z = D, V or H} \qquad [7.27]$$

where v and m denote the operators *variance* and *mean* of an image.

The missing details are given by:

$$C_{Bk}{}^{Z}{}_{h\text{-}l} = a_k{}^{Z}\, C_{A}{}^{Z}{}_{h\text{-}l} + b_k{}^{Z} \qquad \text{for Z = D, V or H} \qquad [7.28]$$

In the case of the SPOT system, where $l=2h$ and if a dyadic multiresolution analysis is used ($p=2$):

$$a_k{}^{Z} = \sqrt{\frac{v(C_{XSk}{}^{Z}{}_{20-40})}{v(C_{P}{}^{Z}{}_{20-40})}}$$

---

[13] T. Ranchin, L. Wald L., and M. Mangolini. *Op. cit.*

$$b_k^Z = m(C_{XSk}{}^Z{}_{20\text{-}40}) - a_k^Z\, m(C_P{}^Z{}_{20\text{-}40}) \qquad \text{for Z = D, V or H} \qquad [7.29]$$

$$C_{XSk}{}^Z{}_{10\text{-}20} = a_k^Z\, C_P{}^Z{}_{10\text{-}20} + b_k^Z$$

In the case of the fusion of images with a ratio *l/h* of 4 and if a dyadic scheme is used for the multiresolution analysis, two wavelet coefficients images need to be synthesized. Then the equations for the Model 2 are:

$$a_k^Z = \sqrt{\frac{v(C_{Bk}{}^Z{}_{4h-2l})}{v(C_A{}^Z{}_{4h-2l})}}$$

$$b_k^Z = m(C_{Bk}{}^Z{}_{4h\text{-}2l}) - a_k^Z\, m(C_A{}^Z{}_{4h\text{-}2l})$$

$$C_{Bk}{}^Z{}_{2h\text{-}l} = a_k^Z\, C_A{}^Z{}_{2h\text{-}l} + b_k^Z \qquad \text{for Z = D, V or H} \qquad [7.30]$$

$$C_{Bk}{}^Z{}_{h\text{-}2h} = a_k^Z\, C_A{}^Z{}_{h\text{-}2h} + b_k^Z$$

Let take the example of the space-borne system Landsat and the fusion of Landsat TM6 (60 m, thermal infrared band) with Landsat panchromatic band P (15 m). The images of the wavelet coefficients between 60 and 30 m, and between 30 and 15 m should be computed. The equations are:

$$a^Z = \sqrt{\frac{v(C_{TM6}{}^Z{}_{60-120})}{v(C_P{}^Z{}_{60-120})}}$$

$$b^Z = m(C_{TM6}{}^Z{}_{60\text{-}120}) - a^Z\, m(C_P{}^Z{}_{60\text{-}120})$$

$$C_{TM6}{}^Z{}_{30\text{-}60} = a^Z\, C_P{}^Z{}_{30\text{-}60} + b^Z \qquad \text{for Z = D, V or H} \qquad [7.31]$$

$$C_{TM6}{}^Z{}_{15\text{-}30} = a^Z\, C_P{}^Z{}_{15\text{-}30} + b^Z$$

Some authors use a hybrid Model 2 where the dynamics of the image A_h is stretched to have the same mean and variance as the image B_{kl}. Because the mean of any wavelet coefficient image is null (less than 10^{-3}), and wavelet transform is linear, the results of the Model 2 and the hybrid Model 2 are very similar. Another variation results from the adjustment of the probability density function of A_h (expressed as the cumulative histogram) to that of B_{kl}.

In the Model 3 a_k^Z and b_k^Z are now computed using an adjustment between $C_{Bk}{}^Z{}_{ph\text{-}pl}$ and $C_A{}^Z{}_{ph\text{-}pl}$ using either least square fitting or axis of inertia. The form of the model is still linear:

$$C_{Bk}{}^Z{}_{h\text{-}l} = a_k^Z\, C_A{}^Z{}_{h\text{-}l} + b_k^Z \qquad \text{for Z = D, V or H} \qquad [7.32]$$

cov denotes the *covariance* operator. For the least square fitting the parameters a_k^Z and b_k^Z are given by:

$$a_k^{\,Z} = \sqrt{\frac{\mathrm{cov}(C_{Bk}^{\,Z}{}_{4h-2l})}{v(C_A^{\,Z}{}_{4h-2l})}}$$

$$b_k^{\,Z} = m(C_{Bk}^{\,Z}{}_{ph\text{-}pl}) - a_k^{\,Z}\, m(C_A^{\,Z}{}_{ph\text{-}pl}) \qquad \text{for Z = D, V or H} \qquad [7.33]$$

For the first axis of inertia (first principal component), these parameters are:

$$a_k^{\,z} = \frac{v(C_{Bk}^{\,Z}\,ph\text{-}pl) - v(C_{Ak}^{\,Z}\,ph\text{-}pl)}{2\ \mathrm{cov}(C_{Bk}^{\,Z}\,ph\text{-}pl,\, C_{Ak}^{\,Z}\,ph\text{-}pl)} + \frac{\sqrt{[v(C_{Bk}^{\,Z}\,ph\text{-}pl) - v(C_{Ak}^{\,Z}\,ph\text{-}pl)]^2 + 4\mathrm{cov}(C_{Bk}^{\,Z}\,ph\text{-}pl, C_{Ak}^{\,Z}\,ph\text{-}pl)}}{2\,\mathrm{cov}(C_{Bk}^{\,Z}\,ph\text{-}pl,\, C_{Ak}^{\,Z}\,ph\text{-}pl)}$$

$$b_k^{\,Z} = m(C_{Bk}^{\,Z}{}_{ph\text{-}pl}) - a_k^{\,Z}\, m(C_A^{\,Z}{}_{ph\text{-}pl}) \qquad \text{for Z = D, V or H} \qquad [7.34]$$

Several experiments showed that both adjustment methods provide similar results. They also demonstrated that the Model 3 gives better results than the Model 2.

The Models 2 and 3 rely on the following assumptions:

- for a given modality *k*, there is a strong linear relationship between the details of the image *A* and those of the image B_k for a range of scales *[ph, pl]*, with *p* greater than 1;
- this relationship also holds for the range of scales *[h, l]* and the parameters of the relation are exactly the same. That is the High Resolution Inter-Modality Model is identical to the IMM.

These assumptions are less stringent than those imposed by the projection and substitution methods and by the relative spectral contribution. This partly explained the better results obtained by the methods developed within the concept ARSIS.

Since the physics are taken into account, the models should apply to radiances, if applicable. However if one uses a linear model such as the Models 1, 2 or 3, and if the calibration law of the sensors is linear, identical results are obtained using directly digital counts (gray levels).

The ARSIS concept is now better understood and is now employed in many applications. This concept is obviously a good framework for the development of accurate methods for the construction of high spatial resolution multispectral images, which are close to the images that the corresponding sensor would observe with the highest resolution. It is a good and open framework with still many places for the development of different cases of applications and approaches for implementation.

Different methods can be developed based on this concept, depending upon the multiscale description and synthesis model MSM, the Inter-Modality Model relating the content of both representations *A* and *B* and the High Resolution Inter-Modality Model transforming the parameters of the IMM when increasing the spatial resolution. Though all examples of models given above are dyadic cases, that is the ratio *l/h* is a power of 2, the ARSIS concept is applicable to any value of the ratio, provided one can find appropriate filters for the analysis and synthesis steps of the multiresolution analysis (see e.g. Fig. 7.2).

A number of studies demonstrate the general superiority of methods belonging to the ARSIS concept over other families of methods. However, the achieved quality is not always satisfactory. Further investigations are needed to improve these methods or to design new ones that perform better

There are several ways of improvement. One deals with the modeling of the content of the information. Several tools exist for the multiscale analysis and for the modeling of the high frequencies in the time-frequency domain. They have different properties and some may be more adapted than others, resulting in a better quality of the synthesized images

The modeling of the inter-modality behavior of the small-size structures (high frequencies) is central in the ARSIS concept. The models IMM and HRIMM presently available are rather straightforward. Though they already produce satisfactory results, better than other methods, efforts should be made to improve them and finally provide better synthesized images. They are mostly based upon statistical adjustment of some properties representing the signal dynamics. Physical laws should be taken into account in these models. Efforts should be made on the HRIMM. Knowledge is mostly inexistent on this model. It is believed that the improvement in the IMM will benefit to improvements in the HRIMM.

One possible improvement is to design methods and models that take into account all modalities simultaneously, as it is done in the two other groups of methods: the projection and substitution methods and the relative spectral contribution methods. Presently, methods are constructed for each modality separately, without considering the multi-modality aspect.

ILLUSTRATION IN URBAN MAPPING

In this section, airborne images are used to illustrate the fusion for the construction of multispectral images of highest spatial resolution. The method used is the ARSIS-RWM method. The urban area under concern is the oldest part of Marseille, France.

Airborne images were acquired in 1993 by the CNES, the French space agency, to simulate the future satellite SPOT 5. The original images were

processed to simulate multispectral bands XS similar to those of the systems SPOT 1-3, with a spatial resolution of 5 m.

The panchromatic image *P* (A_h) has a resolution of 2.5 m (Fig. 7.7). Many details can be seen. The old harbor is seen in the middle left in black tones. Docks are clearly visible, but boats are not discernible. The network of streets appear clearly though sometimes a street is masked by high buildings. One can distinguish vehicles. Blocks of houses are well defined; their inner courts are visible.

Figure 7.7. Panchromatic image of the oldest part of the city of Marseille, France. The spatial resolution is 2.5 m. Ó CNES 1993

In the middle of the upper part is a large building in clear tones, having an "L" shape. It is a commercial center; one can see some structural elements on the roof. A garden is enclosed in the interior of the "L". It is hardly visible in dark gray levels but some paths can be distinguished in white. This garden contains the oldest remains of the city founded by the

Phoenicians. A magnification of this sub-area is shown in Figure 7.9 (upper left).

Analysis of such panchromatic images is very useful for urban mapping. However, several published studies reveal that image analysts find profitable to have also color composite along the panchromatic band. The color composite better displays vegetation and trees along the streets. They may also offer details in the black areas of the panchromatic image. A color composite has been built from the three original images *XS* with a resolution of 5 m and is displayed in Figure 7.8. In the old harbor, docks and boats are in blue. Streets are in blue as well as large buildings and bare soils. Blocks of houses are in green, vegetation is in red. The large red spot in the upper middle is the garden. The blue mass on its side is the commercial center.

The color compositing is performed by a dynamic allocation of color codes to color classes respective to the frequencies of the triplets[14]. This permits a better exploitation of the information contained in the sets of images, and explains why the color composites may vary from one image to the other.

Compared to the panchromatic image, the benefit of this color composite is obvious, even if at 5 m. The most striking feature is likely the enhancement of vegetation and it greatly helps the analysis. However, the spatial resolution is too low for urban mapping. Docks and boats are confused; trees cannot be individualized.

Three multispectral images *XS*1*, *XS*2* and *XS*3* were synthesized at the resolution of 2.5 m by the means of the method ARSIS-RWM using the panchromatic image *P*. The color composite of these synthesized images is shown in Figure 7.8 (lower image). This image offers much more details than the color composite of the original images *XS*. Docks are more separate; streets are better defined and trees are better delineated. The color tables of both images are not the same, but the colors of both images are very close. This means that the statistical distributions of the spectra are very close between both images, which is expected. It demonstrates the quality of the transformation of the spectral content when increasing the spatial resolution.

Compared to the panchromatic image, the use of color at this high resolution is highly profitable. Streets are more visible, because on the one hand of the various hues of the asphalt, and on the other hand of the trees

---

[14] M. Albuisson. *Codage trichrome et classification*. In Outils micro-informatiques et télédétection de l'évolution des milieux : troisièmes journées scientifiques du réseau de télédétection de l'UREF, pp. 167-173. Presses de l'Université du Québec, Sainte-Foy, Québec, Canada, 444 p., 1993.

bordering them. Even boats are more distinguishable because they offer some colors that are not noticeable in the panchromatic image.

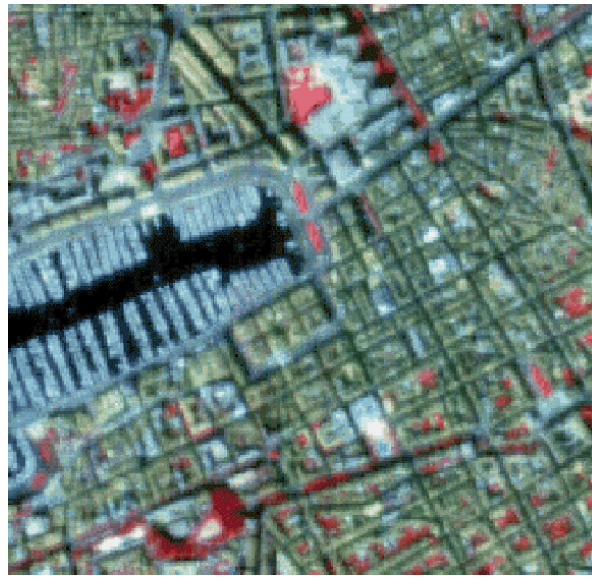
A magnification of the commercial center and the garden area is shown in Figure 7.9. It helps in judging the quality of the fused product and the benefit of the fusion. The panchromatic image is in the upper left. The color composite of the original images *XS* is in the upper right. Comparing those two, one may see that the panchromatic image offers more structural details, owing to its better spatial resolution: see e.g., the structural elements on the roof of the commercial center, or the network of the streets.

On the other hand, owing to its better spectral resolution and its multi-modality properties, the color composite displays information that cannot be seen in the panchromatic image. Vegetation is an example of such information: it appears in red. The color composite shows the garden, the flowerbeds close to the old harbor (bottom left) at the entrance of the famous avenue "la Canebière" (ranging from lower left to middle right) and the trees along the large avenue on the right part, perpendicular to "la Canebière". Looking to both images, one feels the need of fusing both images to obtain a better description of the city.

The color composite of the fused products is presented in the middle left in Figure 7.9. The colors are similar to those of the color composite of the images *XS*. It combines the high spatial resolution of the panchromatic image *P* with the spectral resolution of the images *XS*.

The flowerbeds are better delineated. The trees along the avenue are better seen; the width of the streets can be assessed with more accuracy. This high-quality transformation of the spectral content of set of images *XS* when increasing the resolution allows the application of a classifier, automatic or not, in order to extract the roads and the buildings. Hence, these synthesized images can be used for classification, or for other methods that need to use the multispectral content provided by the whole set of images with the best spatial resolution available.

Further processing may be performed on these synthesized images *XS**. A Laplacian filter was applied to sharpen the contours. The resulting color composite is displayed in the middle right in Figure 7.9 and exhibits striking features. The elements on the roof of the commercial center are well delineated. Each tree appears individually and vehicles are visible. In the garden, the paths are clearly distinguished. The small spots in blue tones are the Phoenicians remains. This color composite can be compared to the photograph of the garden exhibiting the paths and the remains (lower left). The author took this photograph with his back towards the commercial center and looking to the West, *i.e.* the left side of the airborne image.



(a)

(b)

Figure 7.8. Color composites of the same area of Marseille. (a) Original images with a spatial resolution of 5 m. Ó CNES 1993 (b) Synthesized images with a spatial resolution of 2.5 m.

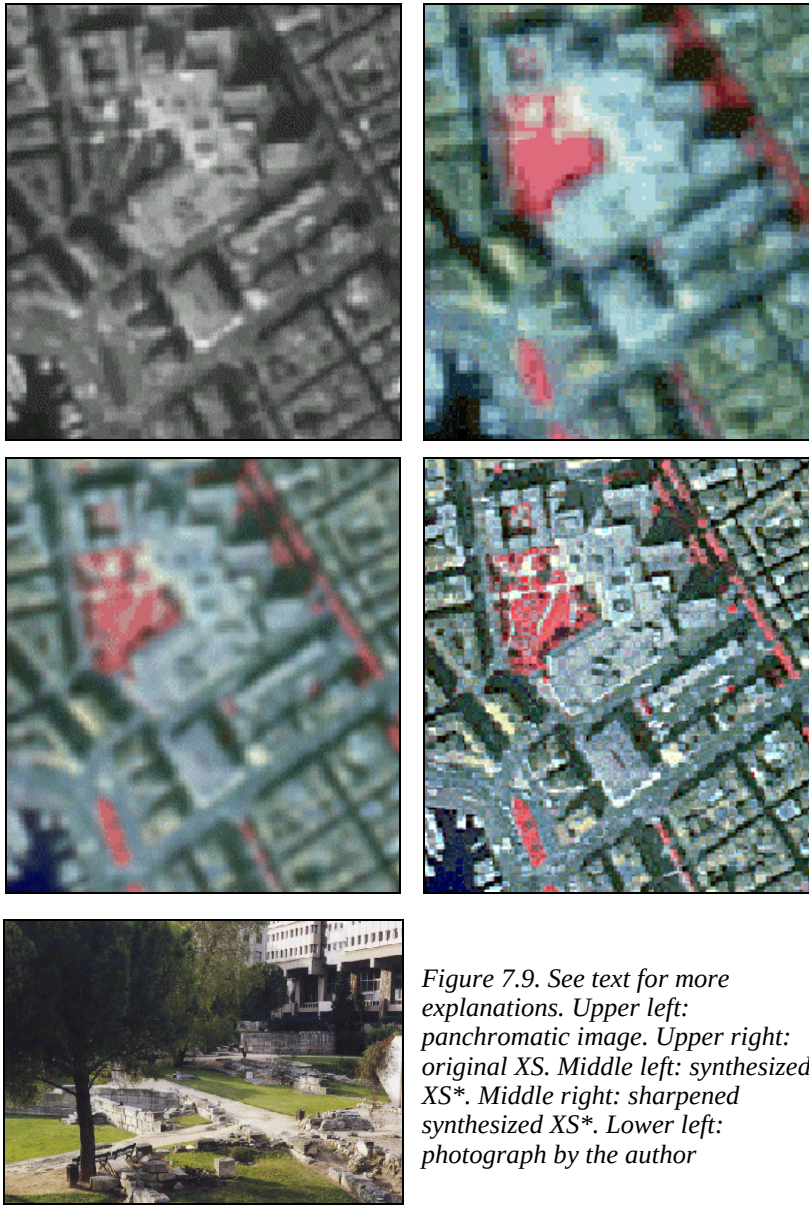


Figure 7.9. See text for more explanations. Upper left: panchromatic image. Upper right: original XS. Middle left: synthesized XS. Middle right: sharpened synthesized XS*. Lower left: photograph by the author*

The benefits of the fused products for the analysis and mapping of the center of the city is demonstrated through this example. The use of color images having a high spatial resolution permits clearly a more accurate interpretation of the features in the city. Furthermore, such fused products may be the object of further image processing techniques, without creation of visible artifacts. This is exemplified by the analysis of that part of the city, wherein remains of the old Phenician city are visible. It illustrates the capability of the synthetic images to support further image processing dealing with the high frequencies.

8. ASSESSING THE QUALITY OF SYNTHESIZED IMAGES

Previous Chapter presents several methods for the fusion of sets of multispectral images B_{kl} at a low spatial resolution l and sets of images A_h at a higher spatial resolution h but with a lower spectral content. These methods aim at synthesizing sets of multispectral images B^*_{kh}, which are as close as possible to the reality B_{kh}. The three properties that should be respected by the synthesized sets are listed in previous Chapter.

Producers, *i.e.* providers of fused products, and customers, *i.e.* users of such fused products, may hesitate to select one of these methods or fused products. Commercial softwares often propose several different methods and it is not obvious for non -specialists to select one method or another for a given case. It follows that usually producers often use methods, which are not the most suitable for their customers.

Several comparisons between methods have been published and are regularly published. However results poorly disseminate in the community and there is lack of knowledge among producers regarding these methods, their advantages and limits. The lack of standardization of protocols for comparison does not add to the clarity of the results. Some efforts have been made recently but a lot still remain.

Behind the choice of a method lie needs for quality. Not neglecting aspects related to software complexity, implementation and maintenance and computation time and other constraints, the quality of a fused product is the driving cause for the selection of a method in industrial systems and production lines.

Thus, the problem is twofold. Firstly, how to assess the *a priori* quality of a synthesized set of images B_h produced by a given fusion method? This may translate into: what is the typical quality one may expect by running a given method over given cases? The answer helps in selecting a method. Secondly, how to assess *a posteriori* the quality of a synthesized set of images effectively produced within a given industrial process?

Quality assessment needs a protocol. We will see later that the same protocol may answer both questions: the *a priori* and *a posteriori* assessment of quality. Such a protocol and the associated quantification of the quality may help in

- system requirements by providing a framework for users to better specify their needs for information;
- information communication by allowing producers, customers and other

persons from all backgrounds to communicate the usefulness of an image to perform a task;

- and analysis by providing an instrument for developing other system performance tools or for assessing the effects of changes in the fusion methods or sensor design or image chain or production line on image quality.

A protocol for quality assessment should have very clear objectives. The objectives of the protocol discussed hereafter are the assessment of the performances of the fused products with respect to the three properties listed in previous Chapter. The typical approach for the assessment of the quality by the means of visual analysis performed by a panel of investigators is also reported. Such an approach is tailored to the needs and objectives of a specific community of users. The actual spatial resolution of the fused product was assessed in this way in the military community.

QUALITY ASSESSMENT NEEDS A REFERENCE

Quality is assessed with respect to a reference image. In the case of the assessment of a method (*a priori* assessment of fused products), sets of actual multispectral images B_h at high resolution h are usually available. The fused products B^*_{kh} are made from the images A_h and B_{kl} and are compared to these references B_{kh} through a visual analysis or computation of similarities and discrepancies, in an automated way or not.

WHAT TO DO IF NO REFERENCE IS AVAILABLE?

Such a reference is not always available and should be created. This is the usual case of the assessment of a fused product (*a posteriori* assessment). One of the most common approaches to this shortcoming consists in interpolating low resolution images B_{kl} up to the high resolution h, and assuming that these images constitute the reference. In any case, are the interpolated images representatives of what would be observed by a similar sensor with a higher resolution, and these interpolated images cannot constitute a valid reference. It follows that this approach is not valid and should not be used. It is in itself a paradox: if interpolated images are assumed to be the reference, why should one bother with fusion methods?

Other protocols try to avoid establishing images of reference, mostly by using some statistical quantities or features derived from the original data set and from the synthesized images. One example is the use of the histograms of the synthetic products, which are compared to the original ones. The histograms for images taken by the SPOT system over the city of Barcelona (Spain) are presented in Figure 8.1. These images are displayed and discussed in following Chapter. On the upper half are the histograms of the original images *P* and *XS1*. For the latter, the resolution is 20 m only: it

contains four times fewer pixels than *P* or the synthetic images *XS**. For the comparison, the histogram of *XS1* has been normalized to the others by multiplying the number of pixels by four. Though the resolution is increased by a factor of two relative to that of *XS1*, the histograms of the images *XS*1* synthesized by two different fusion methods are expected to be close to that of *XS1* in shape. This is true for the histogram of the image $XS*1_{RWM}$ synthesized by the ARSIS-RWM method (lower right). Its highest frequency is close to four times that of the histogram of *XS1*. The modal values are the same and the shapes of both histograms are very similar. On the contrary, the histogram of the image $XS*1_{P+XS}$ synthesized by the P+XS method (lower left) is much closer to that of the image *P*, both in shape and in peak. It indicates the discrepancies between the actual image *XS1(10 m)* and $XS*1_{P+XS}$ and the spectral distortion induced by the P+XS method.

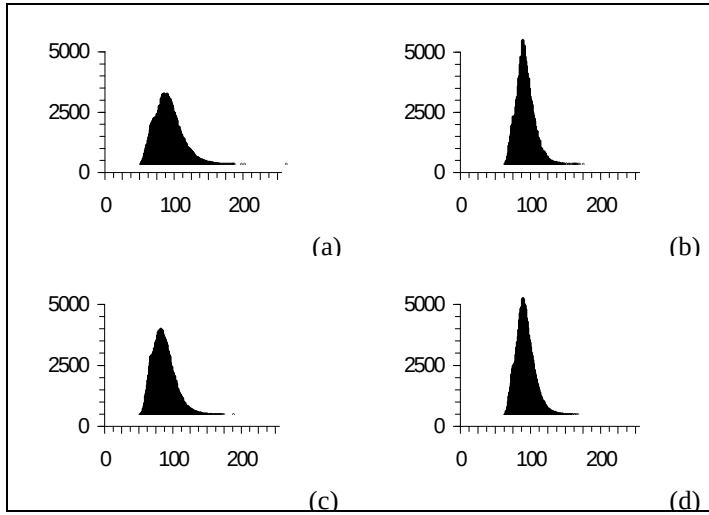


Figure 8.1. Comparison of histograms of original and synthetic SPOT images. Scene of Barcelona, Spain. (a) SPOT P, 10 m resolution; (b) SPOT XS1, 20 m resolution; (c) synthetic image (P+XS method), 10 m resolution; (d) synthetic image (ARSIS-RWM method), 10 m resolution.

This comparison of histograms is a fairly good estimator of image quality, and is easy to handle. However, the effect of the spatial resolution upon the statistical properties of an image should not be neglected. Several published studies demonstrate the non-preservation of statistical distribution with the change in spatial resolution. This non-preservation depends upon the observed type of landscape. The more energetic the high frequencies at scale *h*, the more dissimilar the statistical distributions at scales *h* and *l*. That means that we should not try to identify the statistical properties of a

synthetic product to those of the original image. Therefore, any protocol based upon the comparison of statistical quantities (e.g., histogram, cumulative distribution, entropy etc.) is not valid.

Another approach found in the Earth observation domain is to compare land-use maps obtained after spectral (and possibly textural) classification of the fused products. This classification approach is valuable because land-use mapping is often the goal of satellite image processing. These maps are compared either to the map obtained from original low-resolution data (e.g., SPOT XS), or to ground truth. In the first case, the same assumption as above is made, that is that some statistical properties are preserved through the increase in resolution. More generally, classification greatly reduces the content in information; this reduction decreases the discrepancies between fusion methods. In the classification process, pixel spectral values are aggregated with their spectral neighbors. Hence, a small difference between the synthesized and the actual spectra at a given pixel may have an impact on classification ranging from null to significant. Furthermore, the results of the classification depend upon the type of landscape, its diversity, its heterogeneity, the time of observation, the optical properties of the atmosphere, the sensor system itself (including the viewing geometry), the type of classifier (supervised, unsupervised), and the classifier itself. Hence, this approach may not reflect the overall performance of a fusion method and should be avoided.

HOW TO CREATE A REFERENCE IMAGE?

Several authors have proposed an approach to create a reference image. It calls upon a change in scales and is as follows:

- two sets of images A_l and B_{kv} are created from the original sets of images A_h and B_{kl}. The image A_h is degraded to the low resolution l (A_l) and the images B_{kl} to the very low resolution v (B_{kv}) with $v=l(l/h)$. If $l=2h$, then $v=2l$;
- the fusion method is applied to the two sets of images, resulting into two sets of synthesized images B^*_{kh} at resolution h and B^*_{kl} at resolution l;
- the original images B_{kl} serve as references. A comparison is performed between B_{kl} and B^*_{kl} by the means of visual analysis and analysis of the similarities and discrepancies;
- finally, the quality observed for the fused products B^*_{kl} is assumed to be close to the quality that would be observed if a reference at resolution h were present.

Such an approach alleviates the lack of "truth" images B_{kh}. This raises a question. How can the assessment of quality of the synthetic images be made at the highest resolution h based upon that made at the lowest

resolution *l*? In other words, how can one extrapolate the quality assessment made at the lowest resolution to the highest resolution?

Intuitively, one thinks that, except for objects having a size much larger than the resolution, the error should increase with the resolution, since the complexity of a scene usually increases as the resolution is getting better and better. That is, one may expect the error made at the highest resolution *h* to be greater than that at the lowest resolution *l*. However, several recent works have demonstrated the influence of the resolution on the quantification of parameters extracted from satellite imagery. Many works dealt with clouds (here the parameter is the cloud coverage), or address the problem of resolution in weather prediction and climate models. Others study how the values of a geographical parameter (e.g., the number and surface of lakes or agricultural lots in a region) vary as a function of the resolution. Some mathematical models have been constructed to explain such these changes in rather simple cases. All these studies demonstrate that the quality of the assessment of a parameter is an unpredictable function of the resolution. It is a very complex function of the relative power of the high frequencies and of the very high frequencies, *i.e.* objects that are unresolved at the resolution *h*, and of the distribution of these objects within the pixel. The multi-modality aspect adds to this complexity.

It follows that the quality of the synthetic images at the highest resolution *h* cannot be predicted from the assessments made with synthetic images at the lowest resolution *l*. However, we may rely on the results of several assessments performed at Ecole des Mines de Paris. They show that there is no clear relationship between the quality parameters obtained for the fused products B^*_{kh} and B^*_{kl}, or between B^*_{kl} and B^*_{kv}, as expected. Nevertheless, it has been often found that the quality was best at the resolution *h* (respectively *l*) relative to the resolution *l* (respectively *v*), and also that the ranking of a method relative to the others was the same at these resolutions. It does not prove that estimates should be better at the resolution *h* than at the resolution *l*. However, it seems reasonable to assume that the quality of the synthetic images at the highest resolution *h* is close to that at the lowest resolution *l*.

A GENERAL PROTOCOL FOR QUALITY ASSESSMENT

A protocol has been worked out, which is accepted by several professional organizations. It is simple to implement. It may become the standard approach agreed upon by all the producers of fused products whose scopes are in the frame of this discussion. It permits to alleviate the need for a reference image if not available and offers a complete checking of the three

properties[1]. It can be used in any case, whether a reference image is available or not, and for evaluating products as well as methods. The general frame is as follows:

- the fusion method is applied to the original sets of images A_h and B_{kl}. It results into a new set of synthesized images B^*_{kh} at resolution h;
- testing the first property: *any synthetic image B^*_{kh}, once degraded to its original resolution l, should be as identical as possible to the original image B^*_{kh}*. To achieve this, the synthetic image B^*_{kh} is spatially degraded to an approximate solution $(B^*_{kh})_l$ of B_{kl}. If the first property is true, then $(B^*_{kh})_l$ is very close to B_{kl}. The difference between both images is computed on a per-pixel basis. The fused products together with the difference image are visually compared to the original images B_{kl} in order to detect trends of error, if any. These trends may be related to the objects in the scene. Then some statistical quantities are computed to quantitatively express the similarities and discrepancies between both images;
- testing the second property: *any synthetic image B^*_{kh} should be as identical as possible to the image B_{kh} that the corresponding sensor would observe with the highest resolution h*. The second and third properties refer to B_{kh}, an image that would be sensed if the sensor had a better resolution h. This image is the reference image and is not always available; otherwise, all the above-cited methods would not have been developed. If a reference B_{kh} is available, the comparison is performed between B_{kh} and B^*_{kh}, using the same means, techniques and statistical parameters as for the first property. If a reference B_{kh} is not available, a change in scale is performed as described in the previous section for creating a reference. The images to compare are now the original images B_{kl} and the fused products B^*_{kl} . The comparison is made exactly in the same way than in the case of the availability of a reference. It is assumed that the quality attained for this reference at the resolution l is similar to that that would be attained at the resolution h;
- testing the third property: *the multispectral set of synthetic images B^*_h should be as identical as possible to the multispectral set of images B_h that the corresponding sensor would observe with the highest resolution h*. As above, if the set of images B_h is not available, the comparison is performed between the sets B_l and B^*_l. As for all the properties, the comparison is made by the means of visual analysis and computation of similarities and discrepancies.

---

[1] L. Wald, T. Ranchin and M. Mangolini. *Fusion of satellite images of different spatial resolutions: assessing the quality of resulting images*. Photogrammetric Engineering & Remote Sensing, 63, 6, 691-699, 1997.

Depending upon the objectives of the assessment and of the available resources, the task of visual analysis will be more or less sophisticated and the computer analysis of the similarities and discrepancies will be more or less extensive. An example of experimentation for the assessment of several fusion methods is given in next Chapter.

THE IMPORTANCE OF THE SELECTION OF THE TEST IMAGES

The type of landscape or objects present within the image used to assess the quality of a synthesizing method has a strong influence upon the results. Obviously, if the objects of the scene are spatially homogeneous for scales ranging between *h* and *l*, any sound method will provide good results. In this case, the benefit of the fusion is questionable since interpolation methods and even duplication will lead to satisfactory results.

Whatever the method, the more predictable the changes in signal with the scale, the better the quality of the final product. Hence, scenes whose objects are self-similar for scales between *h* and *l* should be avoided for test cases, since they do not enhance the properties of a method.

Taking Earth observation as an example, over areas such as the ocean or large agricultural lots, which appear very homogeneous at, say, 20 m resolution, the error made in assuming that these areas are still homogeneous at, say, 10 m resolution, is small. On the contrary, urban areas or small agricultural lots are among the most difficult cases because they exhibit a large number of interwoven objects having different scales.

The particular case of the SPOT images of the city of Barcelona was examined. Barcelona is a large city located in northeast Spain, on the Mediterranean coast. Its harbor is the busiest in Spain. The scene is mostly comprised of urban districts, highways and railroads. It also exhibits small agricultural lots and mountainous areas covered by typical Mediterranean vegetation. The images are shown in next Chapter where several fusion methods are compared (Figs 9.1 and 9.2). Such an urban area has been selected for illustrating the comparison because it is certainly the most difficult type of landscape to deal with according to our knowledge. Urban areas often point out the qualities and drawbacks of algorithms because of the high variability of information in space and spectral band, induced by the diversity of features in both size and nature.

It was found that all information (100 percent), expressed as variance, of the homogeneous part covering the Mediterranean Sea, is borne by structures larger than 40 m for each of the three modalities. On the contrary, for the urban area, half the information (50 percent) is borne by structures having sizes less than 40 m. Urban areas do not possess self-similarity properties, though some parameters, such as the growth of city limits, can be approximated by fractal functions. In other words, structures observed at,

say, 10 m resolution, cannot be accurately predicted from their observations at lower resolution, say, 20 m. This is well-known by experienced image interpreters, and is also sustained by published mathematical evidence. The benefit of an image of a higher spatial resolution is the greatest in these cases. Hence, it is recommended that test images should mainly include such areas. Such cases also offer a large diversity of spectral signatures, which is helpful in judging the ability of a method to synthesize the spectral signatures during the change in spatial resolution.

The spectral heterogeneity of a scene may be characterized by the spectral diversity of the set *B* of images relative to the maximum possible number of spectra S_{max}. A heterogeneity parameter *he* can be defined:

$$he = S / S_{max} \qquad [8.1]$$

where *S* is the number of spectra observed in the set *B*. If *NP* is the number of pixels of the images, then $S_{max} = NP$ and

$$he = S / NP \quad and \quad 0 £ he £ 1 \qquad [8.2]$$

The larger *he*, the more spectrally heterogeneous the scene.

A threshold cannot be given for *he*, separating suitable test cases for inappropriate ones. Actually, this parameter is not robust enough. Assume a scene that offers the same number of spectra when its sizes are slightly reduced. It results in increasing *he* but the difficulty in synthesizing images remains the same.

To avoid this problem, we define another quantity *ho*:

$$ho = 10^4 / S \qquad [8.3]$$

The larger *ho*, the more spectrally homogeneous the scene. Hence *ho* characterizes the spectral homogeneity of the set *B*. The larger the number of modalities, the smaller *ho*. For the case of the SPOT images with three modalities, the author found *ho* values less than 1.4 and most often comprised between 0.2 and 0.4 for urban areas. These areas are considered as the most difficult test cases. For the case of Marseille discussed in Chapter 7 (Fig. 7.8), *ho=0.1*. One of the most difficult cases encountered by the author is the color image (R, G, B) of the baboon. This image is well-known in the community of researchers in image processing. It displays the very colored face of a mandrill and most of the information is contained in the very high frequencies. In this case, *ho=0.04*, *i.e.* the spectral homogeneity is dramatically low. *ho* decreases as the number of modality increases. A value of *ho=0.01* was found for the case of a set of images taken in four spectral bands by the satellite SPOT-4. The landscape was made of several villages close each to the other surrounded by small agricultural lots and other vegetation patches exhibiting high frequencies.

This quantity *ho* can be used to discriminate between test cases that permit to assess the properties of a fusion method and others. The smaller *ho*, the more difficult the images to synthesize at a better resolution. A threshold of 0.4 can be set up from experience. Appropriate test cases should exhibit *ho* values lower than this threshold.

ASSESSMENT BY A PANEL OF INVESTIGATORS

The visual analysis is a key to quality assessment. The objective comparison of the visual quality of multiple images is a difficult and lengthy task to handle. The human visual system is not equally sensitive to various types of distortion in an image. The perceived image quality is strongly dependent upon the observer and upon the thematic application. Standard protocols have been defined, in the field of television and image compression or Earth observation by airborne or space-borne instruments.

Several investigators are gathered together to perform such an assessment. Several sets of fused products are shown to these investigators, who judge some well-defined aspects of the images with respect to well-established criteria. Then their notations are weighted and further processed to obtain a mean opinion score defining the quality of the result. When it comes to the assessment of the quality of a set of multispectral images, the mass of data becomes very large. This dramatically increases the difficulty in computing a quantitative picture quality scale.

Such operations are relevant to the general problem of the assessment of the satisfaction of the customers regarding a given product. Similar experimentations are currently performed for industrial products. Conceptually, the assessment of fused products or of fusion methods is not different. Similar techniques for the selection of panels of users may be used, similar criteria may be employed, and similar mathematical procedures may be applied for the screening of the individual responses and the analysis of the results.

The panel should comprise as much as possible investigators. These investigators may be either trained persons or unaware persons depending on the purpose of the test. The larger the panel, the better since statistics will perform better on a large panel. However, a large panel is more costly in many aspects than a smaller one. The investigators should view images and perform the requested analysis in the same conditions: same type of color monitor, same monitor calibration, same distance of viewing, same surrounding illumination etc. Such assessments operations are very heavy to manage and accordingly, they cannot be performed on a routine basis.

The protocol of such experimentations is more or less the same and is as follows. A set of specifications is established regarding the quality of images. This set of specifications comprises a set of criteria to be respected

by the ideal product. Examples of criteria in the case of visual interpretation of images of scenes, natural or artificial, may be:

- colors should be as close as possible from colors perceived by the human eyes;
- objects of size T_0 or more should be detectable;
- objects of size T_1 ($T_1>T_0$) or more should be identifiable;
- objects of size T_2 ($T_2>T_1$) or more should be subjected to analysis.

A committee within the U.S. Government has established criteria for the interpretability of multispectral imagery in Earth observation, which may serve as references. Examples of such criteria are given in Table 8.1.

Then one product is selected among a series of standard products. This product is called the reference product. Its performances with respect to the above-mentioned criteria serve as references against which are compared the subjective valuation of the panel of users. If the score of a fused product is better than that of the reference product, the fused product is said to be better or to offer better performances than the reference product. Here, a standard product may be the images $B_{khInterp}$ resulting from an interpolation of the images B_{kl} from the resolution l to the resolution h. Another standard product may be a fused product produced by a well-known method (e.g., a projection and substitution method).

Then a panel of investigators is selected. The investigators assess each fused product versus the defined criteria with respect to the reference product. For each criterion, each investigator gives a note. The scale often comprises five notes: much worse performances, worse performances, similar performances, better performances, much better performances. It may comprise more notes, e.g. ranging from 0 to 10 or 0 to 100. When the references are loosely defined or even absent, the scale is often reduced to four notes:

- not satisfying (not relevant, not performant, not efficient), weak;
- not much satisfying (not much relevant, not much performant, not much efficient), rather weak;
- rather satisfying (rather relevant, rather performant, rather efficient), rather strong;
- very satisfying (very relevant, very performant, very efficient), very strong.

Once the human analysis performed, the individual notations are screened. Apparent inconsistencies of the answers (e.g., an increase in resolution should likely lead to an increase in the quality of the identification of objects) are looked for. Biased answers are rejected as well as those resulting from misunderstanding of the instructions, criteria, objectives of

the task or protocol. Cross-analyses are performed on the set of scores to discover irregularities. Finally the individual notations are weighted and mean scores are obtained. They qualify specific aspects of the fused product and its overall quality.

MS IIRS Level 1
Distinguish between urban and rural areas.
Identify a large wetland (greater than 100 acres).
Delineate coastal shoreline.
Detect major highway and rail bridges over water.
Delineate extent of snow or ice cover.
MS IIRS Level 2
Detect multilane highways.
Determine water current direction as indicated by color differences.
Detect timber clear-cutting.
Delineate extent of cultivated land.
Identify riverside flood plains.
MS IIRS Level 3
Detect vegetation/soil moisture differences along a linear feature (suggesting the presence of a fenceline).
Identify major street patterns in urban areas.
Identify shoreline indications of predominant water currents.
Distinguish among residential, commercial, and industrial areas within an urban area.
Detect reservoir depletion.
MS IIRS Level 4
Detect recently constructed weapon positions based on the presence of revetments, berms, and ground scarring in vegetated areas.
Distinguish between two-lane improved and unimproved roads.
Detect indications of natural surface airstrip maintenance or improvements (e.g., runway extension, grading, resurfacing, etc.).
Detect landslide or rockslide large enough to obstruct a single-lane road.
Identify areas suitable for use as light fixed-wing aircraft (e.g., Cessna, Piper Cub, or Beechcraft), landing strips.

Table 8.1. Example of criteria related to interpretability of multispectral images taken from the US Government[2]

[2] *Multispectral imagery interpretability rating scale. Reference Guide.* Image Resolution Assessment and Reporting Standards (IRARS) Committee, U.S. Government. February 1995.

GROUND SAMPLE DISTANCE - RESOLUTION OF THE FUSED PRODUCT

Image resolution has a significant effect on interpretability of images. It can be defined as a ground sample distance (GSD), that is the smallest distance that can be measured accurately by the analysts. The fused product intends to simulate what should be observed with a sensor having the best spatial resolution and one may expect the ground sample distance measured in the fused product to be greater than the claimed resolution *h*.

In the course of the analysis, the investigators are asked to compare the GSD they measure on the fused product to that measured on the reference product. An effective ground sample distance (EGSD) is thus defined.

Experiments made for the US Department of Defense[3] by the means of panels of image analysts show that perceived image quality is proportional to the logarithm of the GSD. The effective ground sample distance can be roughly predicted as a function of the spatial resolutions *h* and *l* of the high and low resolution images:

$$EGSD = l - 0.94(l-h) \qquad [8.4]$$

where *EGSD*, *h* and *l* are expressed in meters. Another formulation was proposed

$$EGSD = (1.103\ h) - (0.004\ h^2) + (0.001\ l^2) + 0.37 \qquad [8.5]$$

Equation 8.4 better fits the observations made. The relative gain in resolution is constant (equal to 0.94 times the difference *l-h*) for the resolutions that have been studied (*h* and *l* less than 30 m). Table 8.2 gives some values of *EGSD* computed from Equation 8.4 for several couples of resolutions (*h*, *l*).

Table 8.2 shows that the effective distance EGSD is close to the spatial resolution *h*. The smaller the ratio *l/h*, the closer the EGSD to *h*. This similarity between *h* and *EGSD* demonstrates the benefits of the fusion of images.

The values in Table 8.2 are indicative. The effective distance EGSD depends upon the fusion method employed to construct the products and of the properties of the sensors themselves, including the modulation transfer function, which may impact on the quality of the fused products, depending upon the methods.

---

[3] J. Vrabel. *Multispectral imagery advanced band sharpening study*. Photogrammetric Engineering & Remote Sensing, 66, 1, 73-79, 2000.

l (m)	*h (m)*	*EGSD (m)*
30	10	11.2
20	10	10.6
20	5	5.9
4	2	2.1
4	1	1.2
2	1	1.1

Table 8.2 Predicted effective ground sample distance (EGSD).

COMPUTER-DERIVED MEASURES OF PERFORMANCES

The general protocol is based upon visual analyses of the fused products B^*_{kh} (respectively B^*_{kl}) with respect to the original images B_{kh} (respectively B_{kl}) and upon the computation of the difference between the fused product and the original images on a per-pixel basis. Statistical quantities help in summarizing the similarities and discrepancies between the sets of images. Such measures of performances estimated from these differences offer the benefits of quantitative values and the advantage of being automated in the production lines.

QUANTITATIVE ASSESSMENT FOR THE FIRST PROPERTY

An important point here is the way the synthetic image B^*_{kh} is degraded to $(B^*_{kh})_l$. Some wavelet transforms have the ability to separate scales well, that is, to separate structures of small size from larger ones and, therefore, to simulate what would be observed by a lower resolution sensor. Many authors use an averaging operator on a window of 3 by 3 pixels or more. Such an operator does not have this ability in scale separation and is not as appropriate here. Other filtering operators should be used, some of them simulating a given modulation transfer function (MTF) of a sensor.

A comparison was made at École des Mines de Paris (T. Ranchin, personal communication) on a few scenes using some operators, such as a sine cardinal (sinc) kernel truncated by a Hanning apodisation function of size 13 by 13 pixels, a truncated Shannon function, a bi-cubic spline, a pyramid-shaped weighted average, and the wavelet transforms of Daubechies (1988, regularity of 2, 10 and 20). It showed relative discrepancies between the results on the order of a very few per cent. In conclusion, there is an influence of the filtering operator upon the results, but it can be kept very small provided the operator is appropriate enough.

The quantities that are computed from the differences between the two sets of images are similar to the first and second sets of criteria described under the second property below.

QUANTITATIVE ASSESSMENT FOR THE SECOND PROPERTY

The synthetic image B^*_{kh} (respectively B^*_{kl}) is compared to the reference image B_{kh} (respectively B_{kl}) by means of some criteria described below. The numerical comparison should be made preferably in physical units and in relative values. Thus, different tests made over different scenes may be compared. A difference is computed between B_{kh} and B^*_{kh} (respectively B_{kl} and B^*_{kl}). After visual inspection, the difference image is reduced to a few statistical parameters, which summarize it. There are a large number of candidate parameters. We have computed many for several tens of cases. We have retained some whose definitions are well-known to engineers and researchers and which clearly characterize the advantages and disadvantages of a method.

Two sets of criteria are proposed to quantitatively summarize the performance of a method in synthesizing an image in one spectral band. The first set of criteria provides a global view of the discrepancies between the original image B_{kh} and the synthetic one B^*_{kh}. (respectively B_{kl} and B^*_{kl}). It contains:

- the bias, as well as its value relative to the mean value of the original image. Recall that the bias is the difference between the means of the original image and of the synthetic image. Ideally, the bias should be null;
- the difference in variances (variance of the original image minus variance of the synthetic image), as well as its value relative to the variance of the original image. This difference expresses the quantity of information added or lost during the enhancement of the spatial resolution. For a method providing too many innovations (in the sense of information theory), *i.e.*, "inventing" too much information, the difference will be negative because the variance of the synthetic image will be larger than the original variance. In the opposite case, the difference will be positive. In information theory, the entropy describes the quantity of information. However, we selected the variance difference because most researchers, engineers and practitioners are much more familiar with variance, and entropy and variance act quite similarly for our purpose. Ideally, the variance difference should be null;
- the correlation coefficient between the original and synthetic images. It shows the similarity in small size structures between the original and synthetic images. It should be as close as possible to 1;
- the standard deviation of the difference image, as well as its value

relative to the mean of the original image. It globally indicates the level of error at any pixel. Ideally, it should be null.

The error at pixel level may be more detailed. The absolute value of the difference and the absolute relative error are computed at each pixel. The absolute relative error is the absolute value of the difference between the original and synthetic values, divided by the original value. Then the histogram of the absolute values of the difference and the histogram of these relative errors are computed. Both can be seen as probability density functions. Therefore, we can compute the probability of having at a pixel an error or a relative error (in absolute value) less than a given threshold.

This probability denotes the error made at pixel level, and hence indicates the capability of a method to synthesize the small size structures. The closer to 100 percent the probability for a given error threshold, the better the synthesis. The ideal value is a probability of 100 percent for a null error, relative or not. Here, for reasons of computer precision, the lowest threshold "no relative error or null error" should be set to a very small value instead of zero. Values such as 0.001 or 0.001 percent can be used.

QUANTITATIVE ASSESSMENT OF THE MULTISPECTRAL QUALITY (THIRD PROPERTY)

Visual inspection may be made through color composites of, for example, the first three principal components of the set of images. Both color composites should agree visually. Most methods for color composites are using dynamical adjustment for color coding. If the sets of images are different, even slightly, then the color coding will be different for both composites and no comparison will be possible.

Practically, we recommend the following approach. For each modality or spectral band k, the reference images B_k and the fused images B^*_k are juxtaposed into a single computer file. Here the set of reference images is that used to test the second property (*i.e.* the set B_h or B_l). The principal components analysis as well as the color coding are performed on this set of juxtaposed files. If the number of modalities is less than or equal to three, there is no need to perform a principal components analysis. A color composite is computed by the means of the first three components, which usually contain most of the information. This color composite is split in order to retrieve the composite of the reference images on the one hand and the composite of the fused products on the other hand. If more than one fused product is to be assessed for the same scene, the concatenation (juxtaposition) should be performed with all sets of fused products. This approach guarantees that the color composites are comparable.

The color composites are displayed, simultaneously or alternatively, onto the screen and are compared to the reference composite and to the others.

The advantage of this visual assessment is that it does show trend in errors, if any, possibly related to features in the scene. The drawback of it is that it is a subjective assessment and also that this assessment may be limited either by physiological factors (e.g., color contrast perception by humans), or by technical factors (e.g., when a large number of modalities or spectral bands are present). In the latter case, and if the scene offers a large variety of objects, the color re-coding of the first three principal components reduces dramatically the differences between the sets of images B and B^*, particularly if these differences are random, *i.e.*, not related to specific features in the scene or to a spectral band or modality.

A quantitative assessment can be made using the following three additional sets of criteria in order to quantify the performance of a method to synthesize the spectral signatures during the change in spatial resolution.

The third set (numbered after the two sets described above for the second property) deals with the information correlation between the different spectral images taken two at a time. This dependence can be expressed by the correlation coefficients, with the ideal values being given by the set of reference images B. This is done for every pair among the N available modalities and the image A. As an example, for the case of the modality k, the correlation coefficient between each pair (B_k, B_j, $j=1...N$) and (B_k, A) is computed and compared to the correlation coefficient for each pair (B^*_k, B^*_j, $j=1...N$) and (B^*_k, A). The correlation coefficients found for the fused products should be as close as possible to the correlation coefficients found for the reference images.

The fourth set of criteria partly quantifies the synthesis of the actual multi-modality or multispectral n-tuplets by a method, where n-tuplet means the vector composed by each of the N modalities or spectral bands at a pixel. It comprises the number of different n-tuplets (*i.e.*, the number of spectra) observed in the reference set B and in the synthesized set B^*, as well as the difference between these numbers. A positive difference means that the synthesized images do not present enough n-tuplets; a negative difference means too many spectral innovations.

The previous criteria do not guarantee that the synthesized n-tuplets are the same as in the reference set B. The fifth and final set of criteria assesses the performance in synthesizing the actual n-tuplets. It deals with the most frequent n-tuplets, because they are predominant in multispectral classification. For a given threshold in frequency, only the n-tuplets having a frequency (relative number of pixels) greater than this threshold are used. The threshold is set to e.g., 0.01 percent, 0.05 percent, 0.1 percent, and 0.5 percent, successively. The greater the threshold, the lower the number of n-tuplets, but the greater the number of pixels exhibiting one of these n-tuplets. For each of the n-tuplets, the difference is computed between the

reference frequency and the one observed in the synthesized images. These differences are summarized by the following quantities:

- the number of actual *n*-tuplets, the number of coincident *n*-tuplets in the synthesized images, and the difference between these numbers, expressed in absolute and relative terms;
- the number of pixels in these *n*-tuplets, in absolute and relative terms;
- and the difference between the above number of pixels for the reference and synthesized sets of images, in absolute and relative terms.

This protocol has been applied to several cases. Its capabilities in characterizing the performances of methods and the quality of fused products have been demonstrated. The statistical parameters have proven their high value to summarize the similarities and discrepancies, still conveying enough details so that one may see at a glance the major merits and drawbacks of a method or of a fused product.

A GLOBAL ERROR PARAMETER FOR DESCRIBING THE QUALITY

There is a further need for a simple characterization of the quality of the product of the fusion process, which can be associated to each product and qualifies it. It would greatly help producers to select methods and improve their production lines, and customers to make their choice among products and to assess the impact of this quality on further processing.

The protocol discussed in this Chapter computes the differences between the synthesized images and the actual ones. These differences are summarized by various statistical quantities, which characterize the performance in synthesizing images in a given modality and the multi-modality signature, and especially the most frequent spectra. Published works often use statistical quantities, such as root mean square errors $RMSE(B_k)$. On the contrary, biases and mean values are given seldom.

These quantities as discussed before are very useful to fully understand the performances and properties of a method. However, experience shows that there are too many figures, which are of no help to the customers. There is a need for a quantity, which gives a quick insight of the quality. What we are looking for, is a number simple to understand which is a good indicator of the overall error of the fused product. The closer to 0 this number, the better the product. This quantity should fill three requirements:

First requirement. It should be independent of units, and accordingly of calibration coefficients and instrument gain. Customers seldom consider calibration coefficients. Some fusion methods can be applied to unitless quantities or to radiances. Consequently, the quality parameter should be independent of units.

Second requirement. This quantity should be independent of the number of spectral bands under consideration. This is a *sine qua non* condition to compare results obtained in various conditions.

Third requirement. This quantity should be independent of the scales *h* and *l*. This permits to compare results obtained in different cases, with different resolutions.

The following quantity was proposed to globally characterize the quality of the fused product. It was called total error and is given by:

$$Total\ error = \sum_{k=1}^{N} RMSE(B_k) \qquad [8.6]$$

It is actually the sum over the *N* modalities of the root mean square errors (RMSE) for each modality *k*. The RMSE is that computed by the means of the reference set of images used for the testing of the second property. (*i.e.* B_{kh} or B_{kl}). It is defined as

$$RMSE(B_k) = \frac{1}{NP}\sqrt{\sum_{i=1}^{NP}(B_k(i) - B*_k(i))^2} \qquad [8.7]$$

where *i* is the current pixel and *NP* is the number of pixels. It is also equal to:

$$RMSE(B_k) = \sqrt{(bias)^2 + (standard\ \ deviation)^2} \qquad [8.8]$$

This total error does not obey any of the three requirements. In particular it is sensitive to the changes from numerical counts to radiances. Another error was proposed[4] in order to be able to compare errors obtained from different methods, different cases and different sensors. Let M_k be the mean value for the original spectral image B_k. Let *M* be the mean radiance of the *N* images B_k:

$$M = (1/N) \sum_{k=1}^{N} M_k \qquad [8.9]$$

The relative average spectral error RASE is expressed in percent and characterizes the average performance of a method in the considered spectral bands:

---

[4] T. Ranchin, and L. Wald. *Fusion of high spatial and spectral resolution images: the ARSIS concept and its implementation*. Photogrammetric Engineering & Remote Sensing, 66(1), 49-61, 2000.

$$RASE = \frac{100}{M}\sqrt{\frac{1}{N}\sum_{k=1}^{N} RMSE(B_k)^2} \qquad [8.10]$$

The RASE mostly obeys the first and second requirements. Shortcomings arise in the case of uncalibrated images in different modalities with very different dynamics in gray levels. Further, the RASE does not obey the third requirement.

From this experience, another quantity is proposed. It is called ERGAS, after its name in French "*erreur relative globale adimensionnelle de synthèse*" that means relative adimensional global error in synthesis.

$$ERGAS = 100\ \frac{h}{l}\sqrt{\frac{1}{N}\sum_{k=1}^{N}\left[RMSE(B_k)^2 \Big/ (M_k)^2\right]} \qquad [8.11]$$

It is more robust than the RASE with respect to calibration and changes of units. It also obeys the second requirement. The ratio *h/l* takes into account the various resolutions. For the same error ERGAS, the mean value of the relative $RMSE(B_k)$ increases as the ratio *h/l* decreases, since it is equal to:

$$\mathrm{RMSE}(B_k) = \sqrt{\frac{1}{N}\sum_{k=1}^{N}\left[RMSE(B_k)^2 \Big/ (M_k)^2\right]} \qquad [8.12]$$

For example, if *h/l=1/2* and *ERGAS=3*, the mean value of the relative $RMSE(B_k)$ is equal to 6 percent. If *h/l=1/4*, this mean value is equal to 12 percent for the same ERGAS. This recognizes the increase in difficulty when synthesizing images with large differences in resolutions *h* and *l*.

These various quantities were computed for several cases (see following Chapter). These cases comprise the application of various fusion methods on different sets of images acquired in various modalities with different scales *h* and *l*. The quality of each fused product was assessed as described in the previous pages. A global note was given to each product: bad or good.

The total error decreases as the RMSE for each modality *k* decreases. It is very sensitive to changes in units and in number of modalities. There is no evident relationship between the total error and the global note of quality. The total error cannot represent in a simple way the overall quality.

The relative average spectral error RASE behaves better. It offers a better tendency to decrease as the quality increases. It is independent of units provided they are the same for all bands. It is also independent of the number of bands provided the range of values for each band is constant.

However, like before, there is evident relationship between the error RASE and the global note of quality.

The error ERGAS exhibits a strong tendency to decrease as the quality increases. Thus, it is a good indicator of the quality. It behaves correctly whatever the number of bands is because it uses for each band the RMSE relative to the mean of the band. This definition makes also this quantity independent of the calibration or changes in units, allowing even changes from band to band.

Figure 8.2 displays the error ERGAS computed for several cases. These cases have been sorted out in two categories: bad or good. The labeling was made by the persons providing the cases to the author. Though based upon the protocol above-mentioned and numerical parameters, it has obviously a subjective aspect in the absence of an accepted global error parameter.

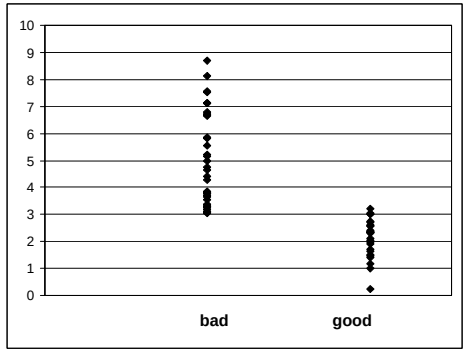


Figure 8.2. The error ERGAS for several cases of fusion

A striking feature in Figure 8.2 is the presence of a threshold. Cases of "good quality" exhibit values less than 3, or slightly greater, while the error ERGAS is larger than 3 for cases of "bad quality".

The existence of this threshold means that the error ERGAS is a good candidate for being the desired global error parameter. A fused product of good quality should exhibit an error ERGAS less than 3. This threshold corresponds to a mean value of the relative RMSE of 6 percent if *h/l=1/2* and 12 percent if *h/l=1/4* (Equation 8.12).

The error ERGAS provides a quick and accurate insight of the overall quality of a fused product. It behaves better than the other quality parameters. Since the error ERGAS reflects the conclusions of the different authors relative to the methods, it may serve to broadly assess the quality of a method. Very similar values of the error ERGAS are found for different cases, which have been declared satisfactory by their authors.

A threshold of satisfaction may be set to *ERGAS=3* for a product. Below 3, the global error is small and the product is of good quality. Well above 3, the global error is large and the product is of lower quality. The quality decreases as the error ERGAS increases.

Further investigations on the error ERGAS, or an equivalent error, would make possible in a near future for producers of fused products to deliver a standardized assessment of the quality of their products. This would allow them to better design and improve their production chains, and would allow customers to better select the products and improve their efficiency.

9. ANALYSIS AND COMPARISON OF THE DIFFERENT METHODS

This Chapter presents a comparison of the most frequently used fusion methods for synthesizing multi-modality images with an improved spatial resolution. This comparison is performed by comparing the results attained by the various methods seen in Chapter 7. The quality is assessed by the means of the protocol described in Chapter 8. Several aspects are assessed: visual, performances in synthesizing individual spectral images and multispectral sets. These aspects are the most important with respect to the subsequent application of classification techniques on the synthesized multi-modalities ensemble.

Other aspects may have been considered, such as spatial gradients, forms and structures, both in each spectral band and in the multispectral set. Such aspects and the corresponding criteria are of high importance in several applications such as the automatic recognition of objects, features, networks and so on. They have not been considered here. Hints about the performances of each method vis-à-vis these more specific aspects may be drawn from the present discussion.

Several comparisons of the fusion methods presented in Chapter 7 have been published. It is not always easy to draw firm conclusions from these studies because their conditions differ greatly. However, several of them followed the same protocol for quality assessment and the conditions of experimentation are well known. In addition, many efforts were spent at the Ecole des Mines de Paris to assess the quality of various methods in several different cases: various spectral bands, various spatial resolutions, various areas (though mostly urban areas), and various sensors. The various results can be gathered together, thus creating a large set of experiments, enlarging the basis of expertise and increasing the knowledge about these fusion methods.

The conclusions drawn by such comparisons may lead to a selection of a particular method. Once implemented in an operational system, this method may reveal itself inappropriate or of lesser accuracy than expected. In this Chapter, the comparison of the methods is extended to some operational considerations. It is sometimes, if not often, difficult to obtain images A_h and B_{kl} acquired at the same time. The second part of this Chapter is devolved to the examination of the influence of the time lag between the dates of acquisitions of the images on the quality of the images synthesized by the various methods.

Two cases illustrate the findings of this Chapter. Both are taken in the domain of Earth observation. Nevertheless, the conclusions drawn in this Chapter are fully general.

THE METHODS UNDER COMPARISON

Eight methods were selected. They are relevant to the three groups discussed in Chapter 7.

Projection and substitution group. The IHS (Intensity, Hue, and Saturation) method and the PCA (Principal Component Analysis) method were tested. The IHS method was only used in case of three images in the set *B*.

Relative spectral contribution group. The Brovey transform and the CNES P+XS method were tested. It should be noted that the Brovey transform does not well represent this group because there is no adjustment of the mean value. Nevertheless, this method and its related "color normalized" method are often used, especially in the military domain. Their implementation often comprises an adjustment of the cumulative histograms of all images under concern A_h and B_{kl}. This creates spectral distortions that prevent their products from respecting any of the three properties. Nevertheless, the visual quality is often good.

As usual the Brovey transform has been applied to the whole set *B*, including the modalities *k* that are not in the spectral ranges covered by the set *A*. Artifacts are created in the correlation between the image A_h and the images B^*_{kh} synthesized in these modalities. It also induces additional spectral distortions. The P+XS method was applied as described in Chapter 7. Nearest neighbor resampling was applied in case of modality *k* outside the spectral ranges of the set *A*.

ARSIS concept group. The High-Pass Filtering (HPF) method and three methods making use of wavelet transform: Model 1, Model 2 and RWM were selected. For the last three methods, the mathematical tools are those described in Chapter 5, except for the non-dyadic cases where filter-banks close to wavelets were employed.

Additionally, an interpolation was used, in order to assess the benefits of the fusion and to check whether it is worth to implement or use one of these fusion methods relative to a much simpler procedure, for which there is no fusion at all. The interpolation method was based on the nearest neighbor technique (duplication technique), or bicubic function or spline function.

THE PROTOCOL FOR ASSESSMENT

The protocol described in previous Chapter is followed to assess the quality of the results of the different methods. More than thirty sets of images were processed at the École des Mines de Paris using several methods among the eight above-listed. Comparisons were made between the various products and conclusions were drawn. In addition, results from published works were incorporated.

Most images were acquired by space-borne sensors. Spatial resolution ranges from 1 to 120 m. Systems under concern are the SPOT 1 to 4 systems, the Landsat 5 system and the Ikonos system. Airborne images were also used; the spatial resolution ranges from 0.8 to 10 m. Most of the observed landscapes were urban areas. A few color images (R, G, B), such as the famous baboon image, were also processed. In that case, one of the channels (e.g., R) was selected as the high resolution image A_h and the others were degraded to twice the original pixel size.

Modalities cover the optical spectrum, from blue to thermal infrared. The ratio of resolutions h/l ranges from 1/2 to 1/12.

To assess the first property, the image B^*_{kh} synthesized at the resolution h for the modality k were filtered for high frequencies before resampling to degrade the resolution down to the resolution l: $(B^*_{kh})_l$. They were then compared to the original images B_{kl}. The filtering function was a sine cardinal (sinc) kernel truncated by a Hanning apodisation function of size 13 times 13 pixels at the resolution h.

To test the second and third properties, the A_h and B_{kl} images were degraded to the resolutions l (image A_l) and $v = l\ (l/h)$ (image B_{kv}), respectively. Then, the images B^*_{kl} were synthesized at the resolution l for comparison with the original images B_{kl}. The degradation was performed as for the first property.

The comparison between the synthesized images $(B^*_{kh})_l$ and the original images B_{kh}, or between B^*_{kl} and B_{kl}, was achieved by a visual inspection on the one hand, and by performing a difference pixel per pixel on the other hand. The discrepancies were analyzed and synthesized in five sets of criteria, which deal respectively with:

- each modality k in a global way;
- the statistical distribution of errors at pixel level for each modality k;
- information correlation between the different modalities;
- the multispectral aspect, that is the error in synthesizing spectral signatures (multi-modality signature);
- the synthesis of the most frequent spectral signatures.

THE ILLUSTRATION CASE

Images taken by the satellite SPOT illustrate the comparison. This case permits the application of all the methods. It is difficult enough and general enough to enhance the properties of each method.

The images were acquired on 11 September 1990, over the city of Barcelona, which is a large city located in northeast Spain, on the Mediterranean coast. Its harbor is the busiest in Spain. The sub-scene used for the comparison is mostly comprised of urban districts, highways and railroads. It also exhibits small agricultural lots and mountainous areas covered by typical Mediterranean vegetation.

Figure 9.1 displays an extract showing the western, newest districts of the city. The upper left part shows the panchromatic image SPOT P acquired at a spatial resolution of 10 m.

A highway crosses the area from the northeast to the southwest. South of it, is a stadium with a gymnasium. North of the highway, a series of parallel elements can be seen; they are close to a very large building. Actually, these elements are made of vegetation or bear vegetation, as it appears in the color composite in Figure 9.2 where vegetation is in red. Without the set *XS* of multispectral images, they would have been mistaken as small-elongated buildings. This demonstrates the usefulness of color in image analysis on the one hand and the benefit of fusion for the synthesis of images on the other hand. In the upper left corner is the foot of the hill covered by sparse vegetation. Large avenues can be seen. Other streets are discernable.

The image *XS1* is displayed in the upper right part of Figure 9.1. It has been magnified by a factor of two, since its original resolution is 20 m. Lower left is the synthetic image obtained by the P+XS method ($XS*1_{P+XS}$). The image $XS*1_{RWM}$ synthesized by the ARSIS-RWM method is in the lower right part.

The image *XS1* exhibits fewer details than the image *P*. It appears more blurred. The correlation between both images is very high. Nevertheless, one may note some differences in contrast between objects of large size in both images, due to the change in resolution and the difference in spectral bands.

The parameter *ho* of spectral homogeneity has a value of 0.02. As an average, each triplet (spectrum) in this set *B* is borne by 5.7 pixels. These values are very low and mean that this set *B* is spectrally heterogeneous; it has a large number of triplets and the spatial distribution of these triplets shows high spatial frequencies. It is a difficult case for synthesis.

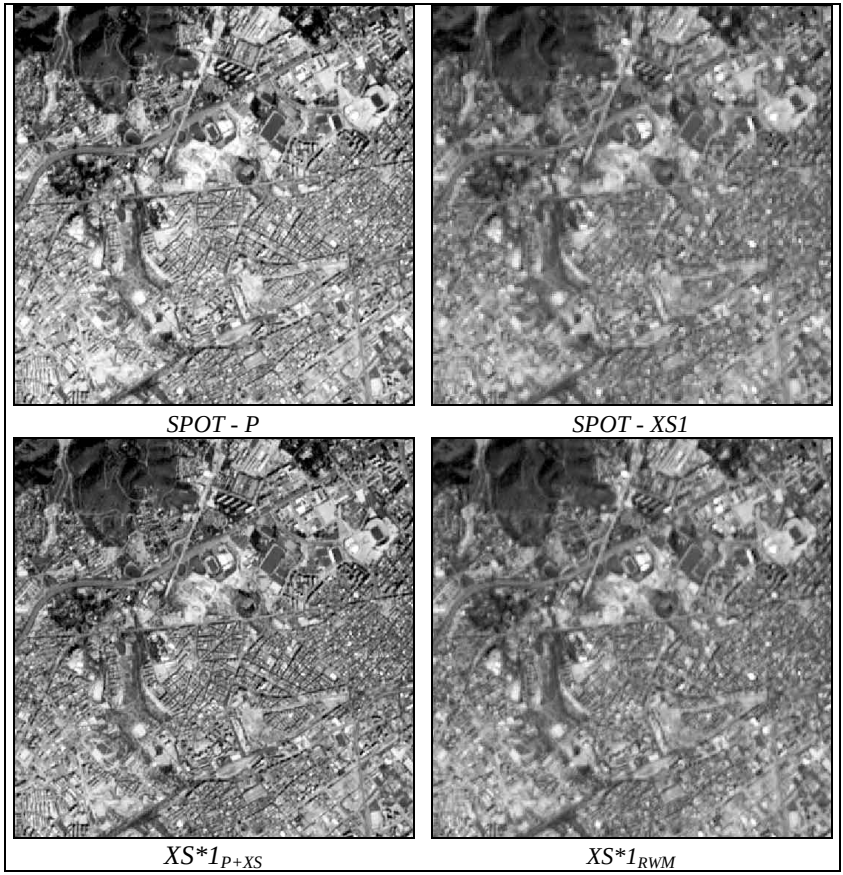


Figure 9.1. SPOT images of the city of Barcelona. ã CNES SPOT-Image 1990. Upper left: panchromatic image (resolution: 10 m). Upper right: image XS1 (green-yellow). Lower left: synthetic image $XS*1_{P+XS}$*. Lower right: synthetic image* $XS*1_{RWM}$*.*

Table 9.1 gives some statistics and numbers for the set *B*. The mean values and the standard deviations are given in radiances and in gray levels. The calibration factor permits to convert gray levels into radiances and reciprocally. For a spectral band *k*, the radiance R_k is linked to the digital count (gray level) DC_k by the calibration factor a_k:

$$R_k = DC_k / a_k$$

The mean radiances in the three images *XSk* are similar. However, the standard deviations are different; that of the image *XS3* is low compared to the others. It means that the dynamics of the images are very different. This

causes a problem to some methods, especially the IHS and Brovey methods. As assumed by the P+XS method, the sum of the mean values of the images *XS1* and *XS2* expressed in radiances is twice the mean value of the image *P*. This is not true at all in gray levels.

This Table also reports on the correlation between the images B_{kl} (*XSk*) and A_l (*P*) for each modality *k*. The image *P* is highly correlated to the images *XS1* and *XS2* and only weakly to the image *XS3*. This conflicts with the constraints of success for the methods within the projection and substitution group.

	XS1	***XS2***	***XS3***	***P***
Mean	58	48	55	53
Standard-deviation	12	15	9	15
Calibration coefficient	1.2181	1.22545	1.29753	1.39198
Mean value in gray level	71	59	71	74
Correlation coefficient with P	0.97	0.97	0.35	1.00

Table 9.1. Mean radiances, standard deviations, and calibration coefficients of original images (in $W.m^{-2}.sr^{-1}.mm^{-1}$). Mean values in gray levels. Correlation coefficient between the original spectral bands and the image P resampled at 20 m.

When comparing images, one must pay attention to the contrast table (look-up table) because it acts as a filter (together possibly with the printer) between the information and the human observer. In the case of the SPOT system, the radiances observed in the bands P, XS1, and XS2 are similar for a spectrally neutral target. In the particular case shown in Figure 9.1, the calibration factors are very similar for the bands P and XS1, and, thus, so are the digital counts. It follows that the same look-up table can be applied to each image in Figure 9.1 and that they can be visually compared.

Beyond demonstrating the interest of merging images having different spectral and spatial resolutions, the visual inspection clearly shows the major properties of the two fusion methods. Details are highly visible in the image $XS*1_{P+XS}$. This image is at times sharper than the image *P*: local contrasts are too much reinforced. The extreme values are also reinforced: the white areas are whiter, compared to the image *P*, and the dark areas are darker. This image is convenient to interpret but it is so similar to the image *P* that one may feel that the synthetic image $XS*1_{P+XS}$ does not correctly convey the spectral information that would be observed in the spectral band XS at a resolution of 10 m.

In the image $XS*1_{RWM}$, on the contrary, the local contrasts are maybe too smooth. The high frequencies are not sharp enough; the actual spatial resolution is likely greater than 10 m. Gray tones are very similar to the image *XS1*, which denotes a good synthesis of the spectral content when improving the spatial resolution.

COMPARISON OF THE METHODS

The comparison is carried in two major phases: the visual analysis and the quantitative analysis. Each phase comprises the analysis of the synthesized product with respect to the three properties. The visual phase comprises another step, which is the visual assessment of the product $B*_{kh}$.

VISUAL ANALYSIS

The visual assessment of the product $B*_{kh}$ aims at checking that the fused product is in conformity with what is expected. The actual image B_{kh} is not available for comparison but someone of experience may judge the major drawbacks and benefits of a fused image (see e.g., the discussion about Fig. 9.1). In order to do that, one may use the original images A_h and B_{kl}, or $B^{interp}{}_{kh}$ to guide the analysis.

Each following step deals with one of the three properties. For the first property, images $(B*_{kh})_l$ are compared to the corresponding actual images B_{kl}. For the second and third properties, images under concern are $B*_{kl}$ and B_{kl}. For these three properties, the set of images B_{kl} is the reference and the synthetic products should resemble these images as much as possible.

Firstly, each synthesized image $B*_k$ (*i.e.*, $(B*_{kh})_l$ or $B*_{kl}$) is visually screened with various look-up tables to explore its properties. Then the image $B*_{kh}$ is scrutinized together with the images A_h and $B^{interp}{}_{kh}$ to refine the findings. The look-up table is adjusted as mentioned above for the case in Figure 9.1. This permits a comparison between any synthesized image and the original images *A* and B_k and between all synthesized images for a given modality *k*. The same operation is performed with the images $B*_{kl}$, A_l and B_{kl}.

Finally, color composites are built using three images $B*_k$ (*i.e.*, $(B*_{kh})_l$ or $B*_{kl}$). Principal component analysis may be performed to construct the three components entering the color compositing. These color composites are analyzed separately and then are compared to the corresponding color composites made from the images $B^{interp}{}_{khl}$ or from the actual images B_{kl}.

For the comparison, one should allocate the same color code to the same spectrum in both sets of images *B* and *B**. Here, dynamic allocation of color codes was performed using median-cut or similar technologies. The operation should apply to the ensemble of the sets *B* and *B**, and not to each set individually.

A last comparison is made between the original set B and all the sets B^* synthesized by each method. Figures 9.2 and 9.3 display these color composites B^*_l at the resolution 20 m for the methods IHS, PCA, P+XS and ARSIS-HPF, Model 1 and RWM. The color composite of the original images XS is shown in both images in the upper left corner to facilitate comparisons. Also displayed is the color composite made from the interpolated images B^*_{lDup}. The color coding is the same for all images: comparisons of colors can be made. These color composites are useful to assess the third property. Each color composite should be identical in terms of colors and details to the actual color composite in the upper left corner.

The contrast-adjusted images B^*_k (*i.e.*, $(B^*_{kh})_l$ or B^*_{kl}) obtained by the various methods are visually fairly close and of satisfactory quality, except for the HPF and the duplication images. Details, or high frequencies structures, are more or less enhanced according to the fusion method. The dynamics in gray levels or radiances are also satisfactory; some methods induce bias, others reinforce the extreme values.

Of course, the images resulting from the duplication or the interpolation techniques exhibit fewer high frequencies than the other images. There is no fusion at all, and there is no innovation in terms of high frequencies, in dynamics, or in spectra. The color composite B^*_{lDup} (Fig. 9.2, lower left) offers less details than the actual color composite B_l (Fig. 9.2, upper left). The contours are less sharp. The colors are very similar, the tones are the same but there is a lack in intensity and saturation compared to the original. While the original images B_{kl} are defining the ideal values for the products $(B^*_{kh})_l$ and B^*_{kl}, the interpolated images $(B^*_{khDup})_l$ and B^*_{klDup} give the bottom line for the quality of a fused product. If a fusion method offers lower quality than the interpolation, then fusion is useless or the method is inappropriate.

The images synthesized by the IHS and PCA methods look nicely. They present sharp details, which are coming from the image A_h. Close examinations demonstrate that such enhancements become a drawback in the case of objects whose signal in the modality k is uncorrelated or anti-correlated to that in the image A.

Adjusting the contrast table for each image accommodates for linear changes in statistical distribution, and especially mean and variance. For the Brovey and IHS methods and at times the PCA method, these parameters are strongly modified relative to the original B_{kl} images. Hence, when the look-up table is adjusted to accommodate all the fusion products and the original images, such drawbacks clearly appear. The dynamics of the gray levels, and hence the spectra, of the products synthesized by these methods are very different from that of B_{kl} or B^{interp}_{kh}.

This clearly appears in Figure 9.2. The images B^*_{klIHS} synthesized by the IHS method exhibit a bias similar in the three bands compared to the original images B_{kl}. This would result into a lower intensity in the color composite (upper right). The variances of the synthesized images are much lower than the variances of the original images. The dynamics in each colored band (R, G, B) is low, there is not much variability in color tones and high frequencies are missing.

On the contrary, the color composite B^*_{lPCA} synthesized by the PCA method (middle left) is visually close to the original composite. Details are there, at times too sharp. Colors are very similar globally but large differences may appear when looking closely to groups of pixels.

Images synthesized by methods belonging to the projection and substitution group do not respect any of the three properties, except if the correlations between the images B_{kl} and A_l are very high, for all modalities.

The images B^*_{Brovey} and B^*_{P+XS} are visually satisfactory. High frequencies are present and the contours are sharp. This is the main reason why such techniques (relative spectral contribution) are usually employed when the visual interpretation of fused products is an important topic. The contours in these images may be at times sharper than the images A_h or A_l themselves. The extreme parts of the statistical distribution of radiances or gray levels may be more frequent than in the original distributions: the dark areas appear darker, and the white ones whiter than in the original images.

The Brovey method adapts the statistical distributions of gray levels of the images B_{kl} and A_h so that they are similar. This induces very large differences between the statistical distributions of values of the synthesized images and the actual ones. This prevents the Brovey method from respecting any of the three properties. In the specific case shown in Figure 9.2, the bias is so large that the color composite made with the color coding of the actual composite exhibits a very few tones and is mostly dark.

The comparison of the color composite made from the images B^*_{klP+XS} with the actual one does not reveal drawbacks. They are very close one to each other in details and in colors. Actually, some noticeable differences may be found in some parts, but they are few. They are due to the changes made in the statistical distributions of the values.

The images synthesized by the HPF method are usually of poor quality for the visual aspects. They contain too many high frequencies: the contours are enforced in an excessive manner. The first property is not always respected, though it should be by principle. The two others are much less respected because there is too much high frequencies innovations in the synthesized images. This is fully illustrated in Figure 9.3 (upper right). Too many artifacts are introduced in the image. The details are so enhanced that the

image is useless for visual interpretation. Globally, the colors are similar to those of the actual composite (upper left): this illustrates the respect of the first property. However, the differences in high frequencies are very large; close examination of pixels, or groups of pixels, shows large differences between the two color composites.

The images synthesized by the other methods within the ARSIS concept are satisfactory for the visual aspects, as illustrated in Figure 9.2 (lower half). These methods are inherently built to respect the first property, with reservations regarding the degradation process as discussed in previous Chapter, and therefore the synthesized images are equal to the actual images for the medium and low frequencies.

The introduction of high frequencies and its quality depends upon the Inter-Modality Model. The contours in the images synthesized by the Model 1 are at times too sharp: the high frequencies are those found in the image A_h. As for the Model 2 and RWM, the high frequencies are at times not sharp enough; the actual spatial resolution is likely greater than the resolution h. The statistical distributions of values of the synthesized images and the actual ones are close: there is no bias and the variances are fairly similar.

The synthesis of the spectra in the course of the improvement of the spatial resolution is usually satisfactory. The color composites are very close to the actual ones. In Figure 9.3, the composite made using the Model 1 exhibits more details than that made using the model RWM. As for the colors, the differences are small and quantitative assessments of the differences are necessary to distinguish between these methods. The methods within the ARSIS concept perform quite well, except the HPF method.

Whatever the method, good quality can be reached only if the images B_{kl} and A_l are well geometrically aligned. A difference of one pixel in co-registration leads to unsatisfactory results. Actually bad co-registration is the major cause of unsatisfactory results for the methods within the ARSIS concept group. For the others, the effect of the bad co-registration is the same but the major cause of bad quality lies in their very construction.

As a whole, one may conclude that most fused products are of better quality than are the interpolated images and that fusion is worthwhile. This is not always the case for several methods, including the IHS, Brovey and HPF methods.

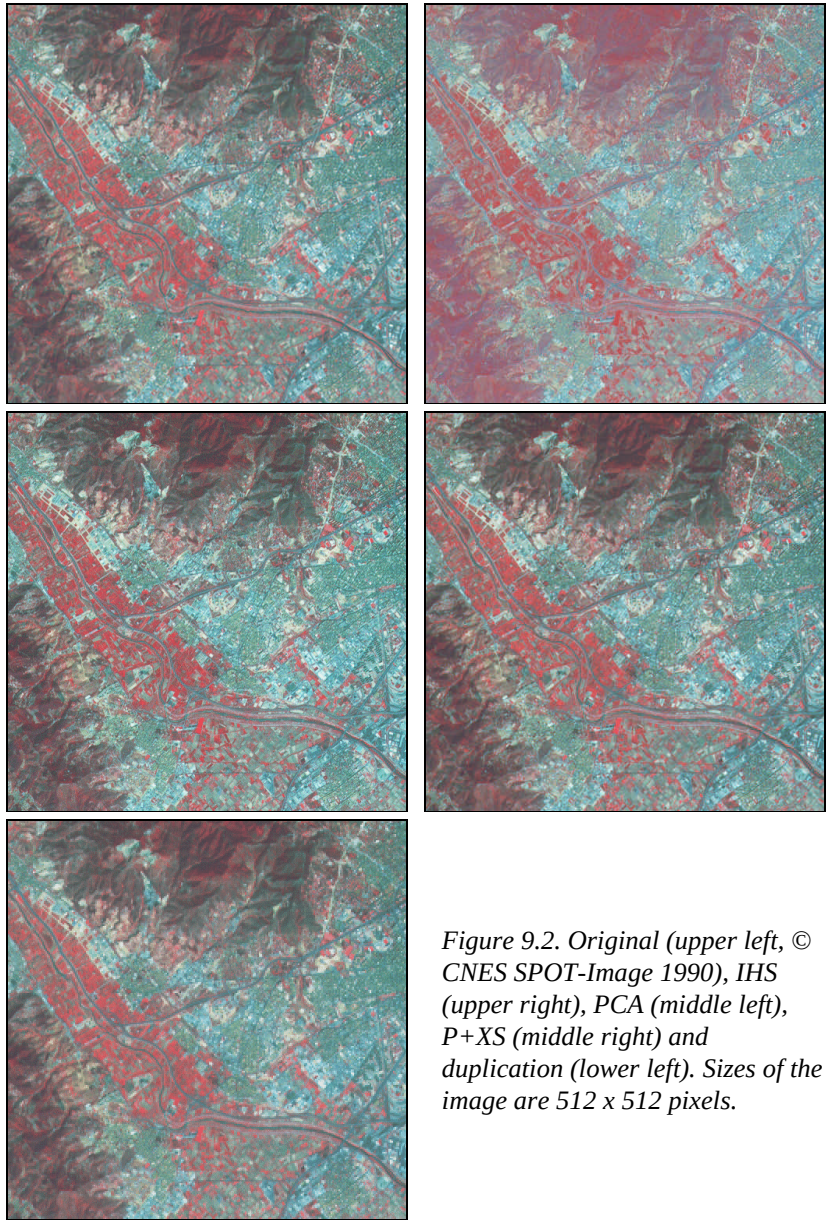


Figure 9.2. Original (upper left, © CNES SPOT-Image 1990), IHS (upper right), PCA (middle left), P+XS (middle right) and duplication (lower left). Sizes of the image are 512 x 512 pixels.

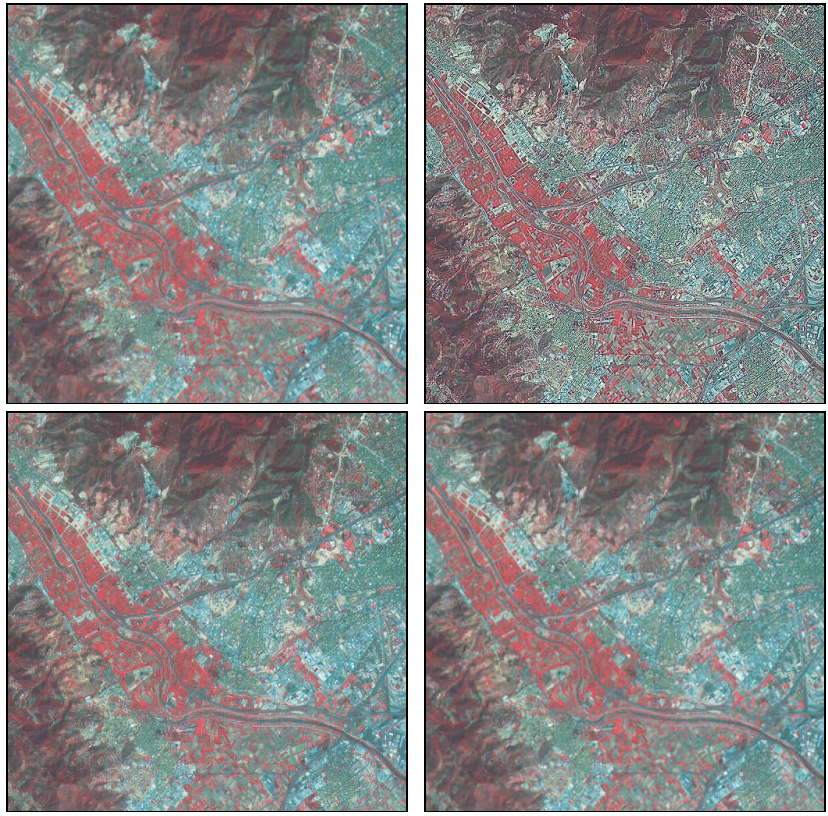


Figure 9.3. Original (upper left, © CNES SPOT-Image 1990), HPF (upper right), Model 1 (lower left) and RWM (lower right).

QUANTITATIVE ASSESSMENT OF THE FIRST PROPERTY

The details of the quantitative comparison further demonstrate that the IHS, PCA, Brovey and P+XS methods do clearly not satisfy the first property. In these methods, the synthesis of the image B^*_{kh} is influenced by the high resolution image A_h and the other images B_{kl} for the modalities $j \neq k$. This influence is irrespective of the size of the structures, that is that the large structures observed in these images A_h and B_{jl} are partly included in the synthesized image B_{kh}. A mathematical analysis of these methods clearly shows that the influence of A_h and of the images B_{jl} on the synthesized image B^*_{kh} does not disappear when reducing the resolution from h to l.

These comments can be extended to any method belonging to the projection and substitution group or to the relative spectral contribution group, including the generalized methods.

The methods built within the ARSIS concept are by essence built to satisfy this first property, with reservations regarding the degradation process as already said. However, the HPF method does not always satisfy this property, mostly because the Laplacian filtering is inappropriate in most cases to extract the high frequencies and only them.

These quantitative findings substantiate the conclusions already drawn from the visual analysis.

QUANTITATIVE ASSESSMENT OF THE SECOND PROPERTY

Compared to the first property, it was found that the testing of the second and third properties better enhances the qualities and drawbacks of a method. This is why an emphasis is put on these two properties. For the second property, statistics are computed, which summarize the differences between the original B_{kl} images and the synthesized B^*_{kl} images. They provide a global view of the quality of a method to synthesize each individual image B_k. Some of these parameters are reported in Tables 9.2 and 9.3 for the illustrating case. Also reported are the values for the pure interpolation (here duplication in this specific case).

Some methods perform better than others do; only a few provide satisfactory results. The IHS method often exhibits a noticeable bias (the difference between the mean values of the images B_{kl} and B^*_{kl}). This bias may be partly overcome by *a priori* equalization of the dynamics of the images B_{kl} and A_l. This would also reduce the differences in variance, and more generally would provide better results if the correlation between the images B_{kl} and A_l were large. This equalization step is made at the expenses of the physical significance of the images. This remark also holds for the PCA and the HPF methods. The IHS method introduce either too much high frequency signal in the synthesized image (the variance is too low), or on the contrary not enough. This depends upon the scene and upon the mutual correlation between bands and the variance in each band. The introduction of high frequencies will be either too large or too low, and sometimes satisfactory. This is true for the other parameters under examination for this second property.

In the case illustrated in Table 9.2, the bias between the actual images *XSk* and the synthesized images XS^*k_{IHS} is large. The standard deviation of the differences is acceptable. The IHS method does not introduce enough high frequency information. The difference in variance is positive and large (Table 9.3), especially for the image *XS3*, which is only weakly correlated to the high resolution image *P*, while the two other images are highly

correlated. The images *XS*1* and *XS*2* are well correlated with the actual images, but this is not the case of the images *XS*3*. This is in full agreement with the analysis of the equations of the method.

	XS1		***XS2***		***XS3***	
	Bias	***Standard-deviation***	***Bias***	***Standard-deviation***	***Bias***	***Standard-deviation***
Duplication	0 %	7 %	0 %	9 %	0 %	7 %
IHS	- 10 %	8 %	- 10 %	8 %	- 10 %	11 %
PCA	- 4 %	7 %	- 6 %	10 %	1 %	6 %
Brovey	64 %	10 %	64 %	17 %	65 %	13 %
P+XS	1 %	7 %	1 %	6 %	0 %	7 %
HPF	1 %	36 %	0 %	43 %	0 %	38 %
Model 1	0 %	4 %	0 %	5 %	0 %	8 %
Model 2	0 %	4 %	0 %	5 %	0 %	7 %
RWM	0 %	3 %	0 %	4 %	0 %	5 %

Table 9.2. Some statistics on the differences between the original and synthesized images for the three bands. The bias and the standard deviation of the differences are relative to the mean value of the image XSk. The relative root mean square error (RMSE) is equal to the square root of the quadratic sum of the relative bias and the relative standard deviation.

	XS1		***XS2***		***XS3***	
	Diff. in variance	***Correl. coeff.***	***Diff. in variance***	***Correl. coeff.***	***Diff. in variance***	***Correl. coeff.***
Duplication	7 %	0.94	5 %	0.96	11 %	0.91
IHS	22%	0.92	14%	0.96	55 %	0.78
PCA	- 47 %	0.98	- 51 %	0.98	8 %	0.92
Brovey	70 %	0.97	77 %	0.98	81 %	0.69
P+XS	- 35 %	0.97	- 19 %	0.98	11 %	0.91
HPF	- 420 %	0.66	- 267 %	0.71	- 57 %	0.48
Model 1	- 4 %	0.98	- 5 %	0.99	- 17 %	0.89
Model 2	- 3 %	0.99	- 3 %	0.99	- 5 %	0.92
RWM	5 %	0.99	3 %	0.99	9 %	0.95

*Table 9.3. Some statistics on the differences between the original and synthesized images for the three bands. The difference between the actual variance and the estimate is relative to the actual variance. The correlation coefficient is computed between the actual image XSk and the estimate XS*k.*

The same comments hold for the PCA method. The PCA method performs slightly better than the IHS method as a whole. The bias is usually less than

that of the IHS products. The PCA method also behaves better for the modalities that are not well correlated with the images A_l. Nevertheless, it is far from being satisfactory.

A large bias is usually found for the fused products synthesized by the Brovey method. This is due to the very construction of the synthesized image $B^*_{klBrovey}$, which, briefly, written, is equal to the image B_{kl}, multiplied by the ratio of n times the band A_l and the sum of the n images B_{jl} for the modalities j^1k. Since the method does not request the computation to be made in radiances, a difference in mean between the images B_{kl}, B_{jl} and A_l may induce a strong bias for all synthesized images. The other methods of the same group impose the equality of the mean values (*i.e.* a null bias). The relative error at pixel level in reconstructing the original image is usually large (see e.g., the standard deviation in Table 9.2). The equations of the method also imply that the variance of a synthesized image B^*_{kl} is a combination of the variances of all other images $B_{jl(l/h)}$, and of the image A_l. It follows that the variance of the B^*_{kl} image strongly differs from that of the original image B_{kl}. In Table 9.3, the variance of B^*_{kl} is too small by a relative amount of 70 - 80 percent. The correlation between the actual images B_{kl} and the synthesized images B^*_{kl} is high as far as the correlation between the B_{kl} and A_l images is high. This is illustrated in Table 9.3 by the comparison between the results attained for the images *XS*1* and *XS*2* on the one hand, and the image *XS*3* on the other hand.

The P+XS method is unbiased by construction, as are the methods of the spectral relative contribution group (but the Brovey method). The standard deviation of the differences is usually acceptable (see e.g., Table 9.2). As already mentioned, it introduces too much signal from the A_l image into the B^*_{kl} images. This translates in a large amount of additional variance compared to the actual images (see Table 9.3 for images *XS1* and *XS2*). This method reduces to duplication for the images acquired in spectral bands not covered by the high-resolution band. Accordingly, the variance of these images is too low: there is no fusion and no addition of signal from another sources (see Table 9.3 for image *XS3*).

The HPF method is rather disappointing. All the contours are enforced but excessively. It induces a low bias, but the standard deviation of the differences is very large (approximately 40 percent in Table 9.2). The amount of energy associated to the high frequencies injected by this filtering technique is too large, and the amount of excessive variance is huge (up to 420 percent in Table 9.3). The correlation between the synthesized and original images is very low for all modalities and clearly indicates the weak similarities between the actual images and the synthesized ones.

The best results are attained by the methods using the wavelet transform. All methods offer approximately the same level of quality. The bias is zero and the standard deviation is usually small. Almost all pixels exhibit a relative error less than 10 - 20 percent in the synthesis. The fused products are usually close to the ideal values. The quality is at times not acceptable. A fine analysis reveals that it comes mostly from inaccurate geometrical alignment and then from the inappropriateness of the models that are unable to represent the relationship between the high frequencies (more exactly the wavelet coefficients) of the various modalities and the change of this relationship with the change in spatial resolution.

The Model 1 exhibits lower quality for images that are not correlated with the high resolution band A_l. In this model, the high frequencies of the A_l image, expressed in wavelet coefficients, are added to the B_{kl} image, possibly after histogram equalization. The low correlation coefficient between the B_{kl} and A_l images denotes a poor similarity in small size structures. Accordingly, the A_l wavelet coefficients do not represent the actual corresponding B_{kl} wavelet coefficients. It results that the synthesized variance is larger than the actual one (Table 9.3 for image *XS3*) and that the correlation coefficient between the actual images and the synthesized images are not high enough. Finally, it should be noted that the results are better for the bands spectrally covered by A_l. The two other models, Model 2 and RWM, are capable of producing satisfactory results for all bands.

QUANTITATIVE ASSESSMENT OF THE THIRD PROPERTY

The performances of each method in synthesizing the multispectral information, *i.e.* the set B_l, have been studied. The differences between the sets B_l and B^*_l are quantitatively analyzed and the findings substantiate the visual analysis of the color composites.

All methods, but those making use of the wavelet transform, increase the correlation between the images B^*_{kl} and A_l, compared to the existing correlation between the images B_{kl} and A_l. This provides a first indication that the multispectral character of the synthetic images provided by these methods may be only partly verified.

The Brovey method is not able to synthesize in an acceptable way the multispectral character. It flattens out the spectral diversity of a scene: the number of spectra found in the set B^*_l is much lower than the actual number. On the contrary, the HPF method synthesizes too many spectra (about twice more). This is due to the large amount of energy found in the high frequencies of the synthesized images, as already noted. As expected, the interpolation method exhibits fewer spectra than the original (approximately half) because of the high spectral heterogeneity of the scene under study associated with large amount of energy in high frequencies. Large changes occur in the statistical distribution of spectral signatures

when changing the spatial resolution. The IHS method performs from bad to good, depending upon the case. The other methods perform from correctly to very satisfactory (Models 2 and RWM).

Of major importance are the performances of each method in synthesizing the most frequent actual spectra. Consider the spectra that have a frequency of at least 0.01 percent relative to the total number of pixels. The total number of pixels that are bearing these most frequent spectra is a large amount of the total number of pixels in the image: these spectra are predominant in the multispectral character of the set B. Hence synthesizing them accurately is of primary importance in classification purposes or true color visualization. The sets of images exhibiting large spectral heterogeneity combined with most of the energy in the high frequencies do not possess predominant spectra. This is the case of the set of images (R, G, B) of the baboon, whose multispectral character is very difficult to synthesize.

Table 9.4 illustrates these performances for the case of Barcelona. Each spectrum (here, a triplet) under consideration is borne by at least 26 pixels in the image (0.01 percent of the total number of pixels). The total of pixels that are bearing these predominant triplets amounts here to 22 percent of the total number of pixels in the image. These most frequent, or predominant, spectra are those causing the colors coding in the color composites shown in Figures 9.2 and 9.3.

For each of these spectra, one looks whether it is present or not in the set of synthesized images B^*_l. Then the number of pixels bearing this spectrum in the synthesized set of images is compared to the corresponding number in the original set of images B_l. The differences are summed up for all the spectra, giving the difference with original. A difference equal to zero means that all the predominant spectra are the same than in original images and that they are borne by the same number of pixels.

Very bad performances are observed for the IHS method. It is usually unable to retrieve most of the predominant spectra. In the case of Barcelona (Table 9.4), the set B^*_l of synthesized images only exhibit less than half of the most frequent triplets (43 percent). The performances are even worse for the number of pixels bearing these triplets. About 90 percent of the pixels are missing. It means that the predominant triplets are not correctly synthesized by the IHS method and even for those it retrieves, they are not correctly allotted to the pixels: this would induce large errors in cartography after classification. If all predominant triplets were representing one unique geographical feature, the surface retrieved from the synthetic set B^* would be reduced by 90 percent compared to the actual one.

The performances of the Brovey method are bound to the non-respect of the mean values in its construction. This creates large bias in all images B^*_{kl},

which impedes the Brovey method from retrieving any of the most frequent spectra. If the dynamics in gray levels of all images $B_{kl(l/h)}$ and A_l are adjusted prior to the fusion, this creates spectral distortions and there is no chance to retrieve the predominant spectra. This method should be avoided if the multispectral character of the set of images B is of importance.

	Number of predominant triplets	***Difference with original (ideal: 0) (in percent)***	***Number of pixels***	***Difference with original (ideal: 0) (in percent)***
Original	1 675	—	60 372	—
Duplication	1 675	0 (0 %)	61 916	-1 544 (-3 %)
IHS	721	954 (57 %)	6 961	53 411 (88 %)
PCA	1 673	2 (0 %)	52 186	8 186 (14 %)
Brovey	0	1 675 (100 %)	0	60 372 (100 %)
P+XS	1 671	4 (0 %)	35 864	24 508 (41 %)
HPF	1 675	0 (0 %)	28 849	31 523 (52 %)
Model 1	1 675	0 (0 %)	53 876	1 996 (3 %)
Model 2	1 675	0 (0 %)	60 002	370 (1 %)
RWM	1 675	0 (0 %)	60 195	177 (0 %)

Table 9.4. Some statistics on the performances in the synthesis of the multispectral character. Number of the triplets borne by at least 26 pixels (0.01 percent of the total number of pixels) for each method and the difference with the actual value. Number of pixels bearing these triplets for each method and the difference with the actual value.

The other methods provide results that are more satisfactory. This means that each of these methods is capable of synthesizing the predominant spectral signatures. If an unsupervised classification is made by using only the spectral signatures, without additional morphological features or other information, these methods will provide more or less the same number of classes close to the actual one. The best performances are found for the interpolation methods and those belonging to the ARSIS concept group. This can be explained by their respect of the first property. Because the most frequent spectra are borne by many pixels, which are forming regions, the frequencies observed for the corresponding pixels are ranging from high to medium or even low. Thus, a possible inaccuracy in the synthesis of the high frequencies in the images B^*_{kl} does not prevent the synthesized images to exhibit these most frequent spectra since the latter are also present at lower frequencies, *i.e.* in the set of images $B_{l(l/h)}$.

However, while the predominant spectra are correctly synthesized, the corresponding number of pixels differs from original for most of the methods. The HPF method retrieves a few of the total number of pixels belonging to these predominant classes because of the too large amount of energy it introduces in the high frequencies. In the case of Barcelona, only

half of the pixels are found. If cartography is at stake, it follows that the resulting map may be inaccurate, except if, by chance, class aggregation processes overcome this drawback. The P+XS method is also inaccurate: too many pixels are missed. This comes from the enhancement of the high frequencies, as for the HPF method but with a much more limited extent.

The performances of the PCA method are varying. It may perform better than the above-mentioned methods but is usually of low accuracy. For the reasons mentioned above, the interpolation method obtains good results, better than many methods. This is fully illustrated in the case shown in Table 9.4. This seriously poses the question of the expected benefits of the fusion of images in some cases where classification techniques will provide results wherein high frequencies are of low importance.

The Model 1 usually obtains good results, while the RWM and the Model 2 achieve the best results of all methods. These results are most often excellent. All the spectra are exactly retrieved and the numbers of retrieved pixels carrying one of these spectra in the original and synthesized images are close to each other. In the example in Table 9.4, less than 1 percent of the total number of pixels is missing. This ensures on the one hand a good classification, and on the other hand a good accuracy in mapping from this classification.

GLOBAL ERROR IN THE ILLUSTRATION CASE

The global errors summarize the various parameters seen above in the quantified assessment of quality. Three of them have been seen in Chapter 8 and are reported in Table 9.5 for the illustration case of Barcelona. These global errors are the sum of the root mean square errors for each modality, the RASE (relative average spectral error) and the ERGAS (relative adimensional global error in synthesis, in French "*erreur relative globale adimensionnelle de synthèse*").

	Sum of RMSE	***RASE***	***ERGAS***
Duplication	12.2	7.5	3.8
IHS	21.6	13.4	6.7
PCA	13.6	8.6	4.4
Brovey	105.2	65.1	32.5
P+XS	10.7	6.7	3.3
HPF	62.6	38.6	19.5
Model 1	9.3	6.0	3.0
Model 2	8.1	5.2	2.6
RWM	6.5	4.1	2.0

Table 9.5. Global errors for each method for the illustration case

The values of these global errors are in agreement with the discussion above. Recall that an error ERGAS less than 3 denotes a good quality. The three global errors give the same ranking for the methods. The greatest global errors are found for the Brovey method. Then come the HPF method, the IHS and PCA methods and the duplication method. The P+XS comes next and is close to good quality. The other methods within the ARSIS concept group provide good quality results. They are close to each other, the best results being attained by the model RWM.

In Figure 9.4 are reported the errors ERGAS found for the various fused products obtained by the application of the different methods to the different cases. The errors for the methods belonging to the projection and substitution group and for the Brovey method have not been reported here; only were kept the most performing methods. The errors for the Model 1 cases have not been reported either, because these cases are too few.

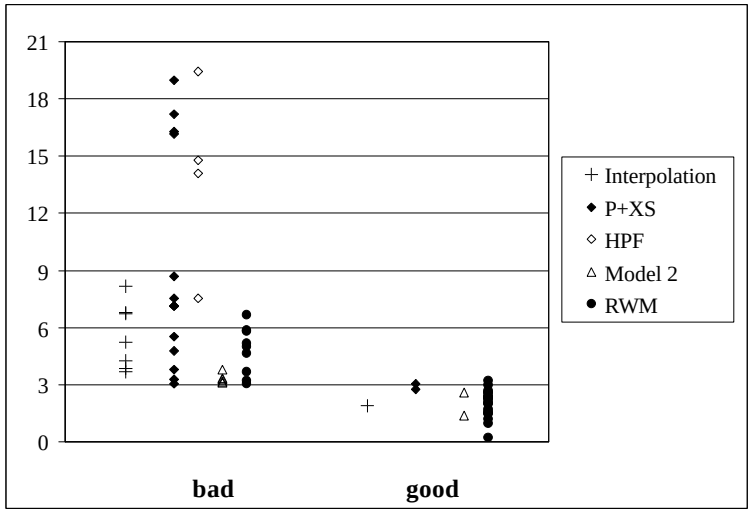


Figure 9.4. Relative global adimensional error ERGAS for the various cases and methods.

As expected, most images produced by interpolation methods are of low quality: the ERGAS is greater than 3 in all cases but one. These methods are not calling upon fusion at all and their results are the baseline against which the fused products are to be compared to demonstrate the benefits of the fusion process. The errors ERGAS associated to the products of the HPF method are always greater than 3 and are larger than those reached by the interpolation methods. There is no benefit at all in using such a method. The

same is true for the projection and substitution methods (not shown in Figure).

The P+XS products exhibit ERGAS that are often greater than 3. They are similar to those of the interpolation methods. This comes from the enhancement of the high frequencies in the P+XS products, which induces large root mean square errors. However, the visual quality of the P+XS products should not be neglected, especially when compared to the interpolated products, provided the sets of images *A* and *B* are contemporary.

The best results are produced by the methods using the wavelet transform. Many cases exhibit an ERGAS less than 3 for the ARSIS-Model 2 products. In the other cases, it is close to 3. The results attained by the ARSIS-Model RWM are better, though at times values much greater than 3 can be reached denoting an unacceptable quality. A fine *a posteriori* analysis of such cases reveals that the geometrical alignment performed by the providers of images is not accurate enough. Evidently a local shift of one pixel or more between the images A_l and B_{kl} has a strong influence upon the synthesis of the high frequencies and affects all methods, except the interpolation methods of course. Other cases of bad quality evidence the inaccuracy or inappropriateness of the models in the ARSIS concept to convert the wavelet coefficients representing the high frequencies of one modality (A_h) into another one (B_{kl}). The test case of the baboon is such a case.

CONCLUSIONS ON THE METHODS

From the previous section, the methods were ranked according to their results with respect to the three properties listed in Chapter 7; they are briefly discussed from the worst to the best. This ranking is based on an average.

Brovey method, color normalized method

The Brovey transform is not relevant at all, mostly because there is a strong bias error due to its very construction. There is no constraint on the mean values. The same conclusions hold for the "color normalized" method. There is also a strong spectral distortion induced by the equations of the method. The constraint on the mean values may be set up through the appropriate adjustment of the gray levels of the images A_l and B_{kl}. The bias would disappear but the spectral distortion would increase. Such methods will never reproduce the spectral content in an accurate way, except in rare cases.

HPF method

As a possible implementation of the ARSIS concept, better results were expected from the HPF method. This method often leads to disappointing results. Too much variance is introduced in the synthesis and this leads to an excessive enforcement of contours as well as to a low correlation coefficient between synthesized and original, or actual, images. The quality of the synthesis of the predominant spectra is bad: though usually all these spectra are retrieved, a large amount of the pixels carrying these spectra are missing.

IHS method

The IHS method often produces nice-looking results but not always. The results are of poor quality: the bias is usually high and the correlation coefficient between the original, or actual, and synthesized images is low. Furthermore, it strongly distorts the spectral content of the synthesized images and the synthesis of the multispectral character of the set B with the change in resolution is of low quality.

PCA method

The PCA method also produces nice-looking results. It can apply in a more general fashion compared to the IHS method; it is not limited to three modalities as inputs. It usually performs better. The bias is small, but too much structures of the high resolution image A_h are injected into the low resolution images B_{kl}. The synthesis of the predominant spectra is often acceptable. Accordingly, it may be recommended instead of the IHS.

These projection and substitution methods deliver products of inconstant quality. Using the ERGAS error, this quality is often bad, if not always. This study and several other authors find that the results achieved by such methods are inferior to the results obtained by the relative spectral contribution methods.

Interpolation method

The interpolation methods provide fairly good results though they do not call at all on the high resolution image A_h. Of course, these products do not exhibit as much details as the others do. However, if the similarities with the actual observations or the multispectral properties of the fused products are at stake, one may legitimately prefer an interpolation technique to the above mentioned methods, especially when considering the extra resources requested.

The performances of the interpolation methods constitute the baseline against which the performances of the other methods should be compared. If the synthesizing method performs better than the interpolation methods,

the fusion is worthwhile. Otherwise, it is not and this regards the Brovey, HPF, IHS and PCA methods.

P+XS method

Belonging to the relative spectral contribution group of methods, the P+XS method performs better than the projection and substitution methods. Though it is limited in itself to the processing of the images provided by the SPOT system, it is a very good example of the generalized relative spectral contribution methods. Results are usually not satisfactory. Of course, it performs like duplication or interpolation for the modalities that are not encompassed by the modalities of set of images *A*. For the other modalities the results are not good. It introduces too many high frequency signals in the synthesized images. It is sensitive to the time lag between the two sets of images *A* and *B*. Finally, the frequencies of occurrence of the predominant spectra are badly synthesized. However, the effective visual enhancement performed by the P+XS method may be recognized.

Model 1, Model 2 and RWM methods

The three methods using the ARSIS concept with wavelet transform provide similar results, which are of good quality and fairly close to the ideal values. ARSIS Model 1 (identity) does not perform so well for the modalities, which are not encompassed by the set of high resolution images *A*.

ARSIS Model 2 and RWM methods perform the best. They achieve good quality products. The quality of the synthesis of the predominant spectra is usually impressive; it depends upon the complexity of the scene with respect to the spectral heterogeneity in combination with the high frequencies. Another striking feature compared to the other methods is that they are capable of achieving good results for all modalities, including those which are not encompassed by the set of high resolution images *A*, to a certain extent of course. All published comparisons show that the ARSIS concept, combined with the wavelet transform and the multiresolution analysis leads to the best presently achievable results.

The conclusion is that only a very few methods achieve satisfactory results (Model 2, RWM). The fusion process is very often worthwhile if one of these two methods is employed. However, the quality reached by these methods is not always satisfactory. Further investigations are needed to improve these two methods or to design new ones that perform better. There are two ways of improvement. One deals with the modeling of the content of the information within a modality. Several tools exist for the multiresolution analysis and for the modeling of the high frequencies in the time-frequency domain. They have different properties and some may be more adapted than others, resulting in a better quality of the synthesized images. The second way is expected to bring definite improvements. The

modeling of the inter-modality behavior of the small-size structures (high frequencies) is central in the ARSIS concept. The models presently available are rather straightforward. Though they already produce satisfactory results, better than other methods, efforts should be made to improve them and finally provide better synthesized images. They are mostly based upon statistical adjustment of some properties representing the signal dynamics. Physical laws should be taken into account in these models. In addition, further work should verify that these two methods and others of equivalent quality, or better, could enter a production system delivering fused products with a controlled quality.

INFLUENCE OF THE TIME LAG BETWEEN THE TWO SETS OF IMAGES ON THE QUALITY OF THE FUSED PRODUCTS

The images A_h and B_{kl} may not have been acquired at the same time. Hence, changes may be observed between the image A and the image B_{kl} that are only due to this time lag. These changes should be taken into account by the synthesis method in order to avoid to introduce artifacts in the synthesis of the image B^*_{kh}. This section discusses the performances of the methods with respect to that problem. The analytical analysis is illustrated by a specific case. Another illustration was also provided in Chapter 7 for the IHS method.

ANALYTICAL ANALYSIS

Assume two instants t and t_0. Assume that images B_{kl} of the set B have been acquired at time t_0 and the image A_h at instant t. Assume that one may write:

$$A_h(t) = A_h(t_0) + DA_h \qquad [9.1]$$

By reporting this equation in the equations of Chapter 7, the influence of the time lag on the synthetic images B^*_{kh} produced by a method can be assessed. This influence will be characterized by the difference:

$$DB^*_{kh} = B^*_{kh}(t_0, \textit{with } A_h(t_0)) - B^*_{kh}(t_0, \textit{with } A_h(t)) \qquad [9.2]$$

where $B^*_{kh}(t_0, \textit{with } A_h(t))$ denotes the image synthesized from the images B_{kl} at time t_0 and A_h at time t.

For the generalized relative spectral contribution method, this difference is

$$\Delta B^*_{kh} = \frac{B^{\text{interp}}{}_{kh}(t_0)\,\Delta A_h}{\sum_{j=1}^{N} a_j\, B^{\text{interp}}{}_{jh}(t_0)} \qquad [9.3]$$

This equation shows that the time lag has an influence on all scales. The difference is proportional to the difference DA_h and thus may be very

important. The same comments hold for the projection and substitution methods.

The influence is a bit more complicated to express for the methods within the ARSIS concept. It depends upon the number of scales used for the multiresolution analysis and the complexity of the inter-modality model to infer the missing wavelet coefficients. Anyhow, the influence is limited to the scales that are involved in the fabrication of the model. In the simplest case (the Model 1 and the HPF method), the influence is limited to the range of the highest frequencies *[1/l, 1/h]* under the form of a difference in the wavelet coefficients $C_{B*k}{}^{Z}{}_{h\text{-}l}$:

$$DC_{B*k}{}^{Z}{}_{h\text{-}l} = C_{DA}{}^{Z}{}_{h\text{-}l} \qquad [9.4]$$

where $C_{DA}{}^{Z}{}_{h\text{-}l}$ are the wavelet coefficients for the scales between h and l. Thus, the time lag will introduce high frequencies artifacts in the synthesized images. The frequencies that are less than *1/l* will remain unchanged (first property). Other models may take care of such artifacts and decrease their influence.

The conclusion of this analysis is that the influence of the time lag on the fused products strongly depends upon the performances of the method under concern with respect to the first property. This property is "any synthetic image $B*_{kh}$ once degraded to its original resolution l (image $(B*_{kh})_l$), should be as identical as possible to the original image B_{kl}". The greater the similarity between B_{kl} and $(B*_{kh})_l$, the lower the influence of the time lag. Except the HPF method, the methods belonging to the ARSIS concept group respect this first property better than the others and thus offer results of better quality with respect to the time lag.

EXAMPLE OF THE THREE GORGES DAM

The Three Gorges Project in China is the largest water conservancy project ever built in the world. Figure 9.5 is a color composite image made from images taken by the satellite SPOT. The Yangtze River is flowing from left (West) to right; it appears in blue-green. The river is enclosed in steep relief (Fig. 9.6). The dam is constructed in the elbow of the river. The northern bank is equipped with a pass for large ships; this pass is clearly visible in Figure 9.5.

The reservoir is of a canyon and river-like reservoir with a total length of about 600 km and average width of 1.1 km (Fig. 9.6). It is less than twice the width of natural alluvial channel and the storage capacity of the reservoir is 39.3 billion of cubic meters with the normal pool level (NPL) at 175 m.

The Three Gorges Project is a multi-purpose hydro-development project, producing comprehensive benefits mainly in flood control, power

generation and navigation improvement. The Yangtze River is the major axis of circulation for goods and people in this region and has an essential role in the economy of central China. It experiences dramatic flooding. In summer 1998, 4,000 persons were killed and around 230 millions inhabitants were affected.

The preparation of the Three Gorges building site started in 1993. The detection of geological problems as faults, landslides and rockfalls, that can affect the riverbanks, is of tremendous importance for evaluating environmental and human impacts of the future reservoir. The geological survey has been entrusted to the Chengdu University, which selected remote sensing as a mean for studying the upstream geological impact of this project.

Satellite images from the SPOT system were used to achieve the geological survey. The detection of geological hazards is performed by a photo-interpretation of color composites of such images. Photo-interpreters are able to locate the active faults and the landslides. The geologists working on the dam site were very satisfied with the possibility of analyzing fused products with a resolution of 10 m instead of the original images XS at 20 m. The benefits were on the accuracy of the detected lineaments (lower rate of false detection) and on comfort (easy to do, less time consuming).

Yang *et al.* explored the influence of the time lag between the days of acquisition of the images SPOT P and XS for different fusion algorithms[1]. Their findings are reproduced here with their permission to illustrate the analytical analysis.

The set of data available is comprised of three panchromatic images SPOT P from 1990, 1997 and 1998 and a multispectral image SPOT XS from 1998. The comparison of the three panchromatic images (Fig. 9.7) shows that over these years, significant changes have taken place on both sides of the Yangtze River at the dam site.

[1] W. Yang, F. Cauneau, J.-P. Paris and T. Ranchin. *Influence of landscape changes on the results of the fusion of P and XS images by different methods.* In Proceedings of the third conference "Fusion of Earth data: merging point measurements, raster maps and remotely sensed images", Sophia Antipolis, France, January 26-28, 2000, Thierry Ranchin and Lucien Wald Editors, published by SEE/URISCA, Nice, France, pp 47-56, 2000.

Figure 9.5. Color composite of the synthetic images XS^*_{RWM} (resolution 10 m). The original images XS were taken in November 1998.

Figure 9.6. Artist view of the Three Gorges Project (note the steep relief)

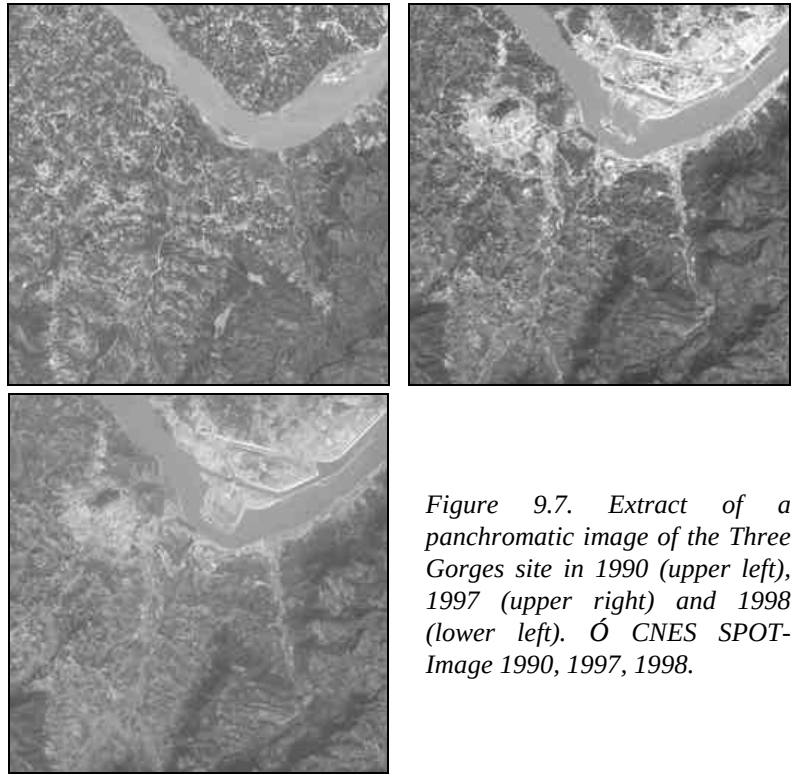


Figure 9.7. Extract of a panchromatic image of the Three Gorges site in 1990 (upper left), 1997 (upper right) and 1998 (lower left). Ó CNES SPOT-Image 1990, 1997, 1998.

One may follow the construction of the dam in Figure 9.7. The upper left image SPOT P was taken in 1990 before the beginning of the construction in 1994. The upper right image was taken in 1997. Compared to the image of 1990, the south bank of the elbow moved slightly southwards. Large human settlements are visible in the left part (clear tones). On the north bank, the pass for ships is almost completed. An artificial island was created in the stream. A bridge and several roads were built. In 1998 (lower left image), the pass is completed and the artificial island has grown.

Each of these images *P* is fused with a set of spectral images *XS* acquired in 1998 at the same time than the image P_{1998}. Figure 9.8 displays the image *XS1* (green-yellow band) at a spatial resolution of 20 m. The structures are very similar to those observed in the image P_{1998}. They differ from those observed in the image P_{1997} especially in the mainstream of the river. Finally, they strongly differ from those in the image P_{1990}. Because of the very large changes observed over the years, this area and these sets of images are well suited to a case study of the influence of landscape changes

on the results of the algorithms. Fusing images with such differences in structures is a challenge.

Figure 9.8. Original XS 1 image of the Three Gorges Dam, China, acquired in November 10, 1998 (resolution 20 m). Ó CNES SPOT-Image 1998.

Three algorithms were scrutinized: the high-pass filtering (HPF) algorithm, the model RWM within the ARSIS concept (ARSIS-RWM) and the P+XS algorithm. They were applied to the three following sets of images:

- images *XS* acquired in 1998 (playing the role of the images B_{kl}) and *P* acquired in 1990 (P_{1990}) (playing the role of the image A_h);
- images *XS* acquired in 1998 and *P* acquired in 1997 (P_{1997});
- images *XS* acquired in 1998 and *P* acquired in 1998 (P_{1998}).

Nine (three sets of images times three methods) synthesized multispectral images at 10 m were obtained. Actually, the IHS algorithm was also used. However, regarding the objectives of the study, the images synthesized by the IHS algorithm offered more or less the same characteristics than the fused images produced by the P+XS algorithm. For the sake of the simplicity, the images output by the IHS algorithm are not discussed here.

In a first phase, the influence of the time lag between the images *P* and *XS* was assessed for each method separately. It was assumed that the images *XS** synthesized at a resolution of 10 m using the panchromatic image taken in 1998 best represent the reality for 1998. For each method, they served as a reference; the other images synthesized using either the image P_{1990} taken in 1990 or the image P_{1997} taken in 1997 were compared to that reference. This comparison was made by skilled geologists. For each method, conclusions were drawn regarding the influence of the time lag with respect to their objectives in geological interpretation. Then comparisons between methods were made.

In a second phase the quality of each set *B** was assessed following the protocol discussed in previous Chapter. To test the second and third properties, all the images *P* were degraded to 20 m and all images *XS* to 40 m. This protocol produces two sets of images for each method. Comparison was made between the original sets of images *B* and the synthesized sets *B** for each method. Quality parameters were computed as discussed in previous Chapter.

The visual inspection over the site of the dam of the resulting images for the two first sets of images, achieved with the HPF and the P+XS methods, demonstrated the failure of these methods when the time acquisition of the images are different. On the contrary, the algorithm based on the ARSIS concept is able to fulfil the objectives of the end-users.

This is illustrated by the case of the band XS1 (green-yellow). For each method, the images synthesized with the help of the images P_{1990}, P_{1997} and P_{1998} are presented.

The HPF method

Figure 9.9 displays the images *XS*1* synthesized by the HPF method at a spatial resolution of 10 m. It focuses on the dam area, where most of the noticeable changes occur. The upper left image is the synthetic image $XS*1_{HPF\ 1990}$ obtained from the images *XS1* and P_{1990}. The upper right image is obtained using P_{1997}, and the lower left using P_{1998}.

The image synthesized from the images *XS1* and P_{1998} is considered as the best achievable result, since both were acquired simultaneously. The high frequencies in this image (lower left) are very similar to those in P_{1998} and are enhanced too much: the image $XS*1_{HPF\ 1998}$ is too sharp. The set of images $XS*_{HPF\ 1998}$*(20 m)* synthesized at the resolution 20 m exhibits an error ERGAS of 7.6 when compared to the original set *XS* (Table 9.1). This confirms the low quality of the synthesized set $XS*_{HPF}$ found by the geologists.

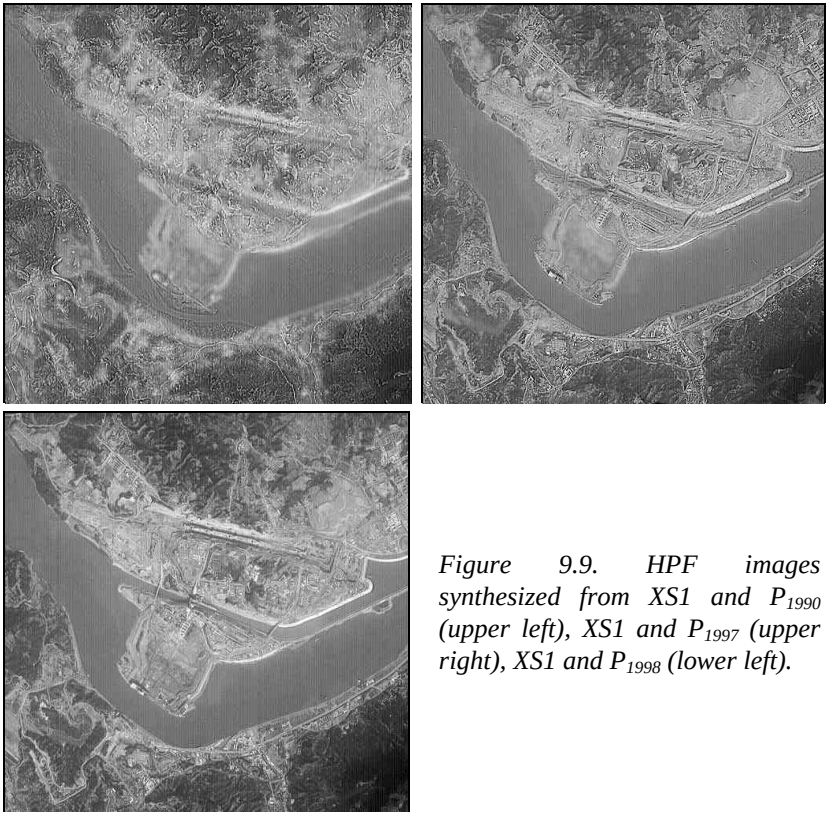


Figure 9.9. HPF images synthesized from XS1 and P_{1990} (upper left), XS1 and P_{1997} (upper right), XS1 and P_{1998} (lower left).

The image $XS*1_{HPF\ 1990}$ synthesized at the resolution of 10 m by the means of P_{1990} exhibits large structures that are similar to those of *XS1*. The first property is respected at least for the low and medium frequencies. However, many high frequency artifacts were introduced, which render this image useless for analysis. One may notice in the mainstream in the elbow the trace of the south bank that moved a bit southwards between 1990 and 1997 (Fig. 9.7).

The information added to the image *XS1* is extracted from the image *P* by application of a Laplacian filter. The modification of the landscape between 1990 and 1998 was tremendously important and the introduction of a non-contemporary information gives the impression of a superimposition of two images.

The image $XS*1_{HPF\ 1997}$ is of better quality. It is still suffering from the addition of high frequency artifacts: it cannot be used for geological interpretation. These findings are sustained by the error ERGAS, which is

equal to 28.2 for 1990 and 26.9 for 1997 (Table 9.1). The quality of the HPF product is already low when the images *XS* and *P* were acquired simultaneously and decreases notably as the time lag increases.

Case	*Method*	*ERGAS*
XS and P_{1990}	HPF	14.8
	P+XS	8.7
	ARSIS-RWM	1.5
XS and P_{1997}	HPF	14.1
	P+XS	7.5
	ARSIS-RWM	1.6
XS and P_{1998}	HPF	7.6
	P+XS	3.8
	ARSIS-RWM	1.5

Table 9.1. Relative global adimensional error ERGAS for the various cases when synthesizing images XS at a resolution of 20 m.*

The P+XS method

Figure 9.10 displays the images *XS*1* synthesized by the P+XS method at a spatial resolution of 10 m. The upper left image is the synthetic image $XS*1_{P+XS\ 1990}$ obtained from the images *XS1* and P_{1990}. The upper right image is obtained using P_{1997}, and the lower left using P_{1998}.

The best achievable result (lower left) is obtained for contemporary images. It is of medium quality. The P+XS method usually enhances the high frequencies; this image is not well suited for interpretation.

The striking feature is provided by the display of the image $XS*1_{P+XS\ 1990}$ (upper left), whose structures are identical to those of P_{1990}. The dam that is visible in the original image *XS1* disappeared during the synthesis! This is a dramatic illustration of the Equation 9.3, as is also the image $XS*1_{P+XS\ 1997}$ (Fig. 9.10, upper right). Both images cannot be used for the analysis of the geological features.

The sets of images $XS*_{P+XS}(20\ m)$ synthesized at a resolution of 20 m exhibit errors ERGAS of respectively 18.5, 16.1 and 3.8 for respectively P_{1990}, P_{1997} and P_{1998}. This confirms the medium to low quality of the fused products. It also demonstrates the dramatic sensitivity of such algorithms to the time lag.

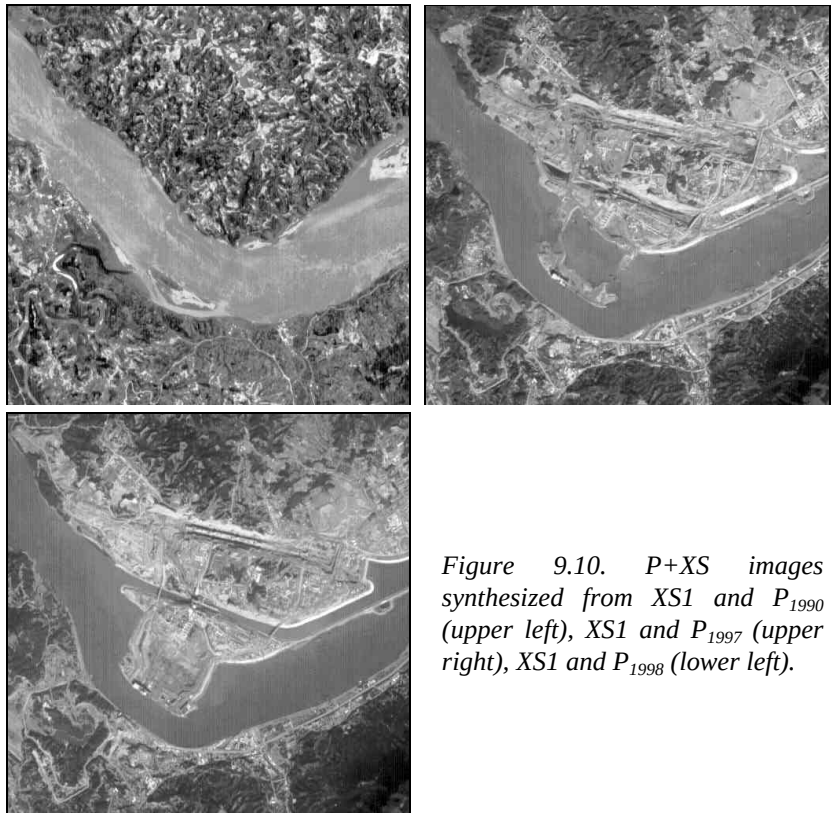


Figure 9.10. P+XS images synthesized from XS1 and P_{1990} (upper left), XS1 and P_{1997} (upper right), XS1 and P_{1998} (lower left).

The ARSIS RWM method

Figure 9.11 displays the images *XS*1* synthesized by the ARSIS-RWM method at a spatial resolution of 10 m. The upper left image is the synthetic image $XS*1_{RWM\ 1990}$ obtained from the images *XS1* and P_{1990}. The upper right image is obtained using P_{1997}, and the lower left using P_{1998}.

The best achievable result $XS*1_{RWM\ 1998}$ is of high quality. Geologists were very satisfied with this product and used it to achieve the geological interpretation of the site. Actually, the three images look very similar though differences appear when analyzing the images on a computer screen. The influence of the time lag is kept very low by the method RWM, which is capable to deal with non-contemporary images. The stability of the results is confirmed by the error ERGAS, which is close to 1.5 in the three cases. Of course, detailed analyses demonstrate that the best results were achieved with contemporary images, for which benefits of the fusion are at their most.

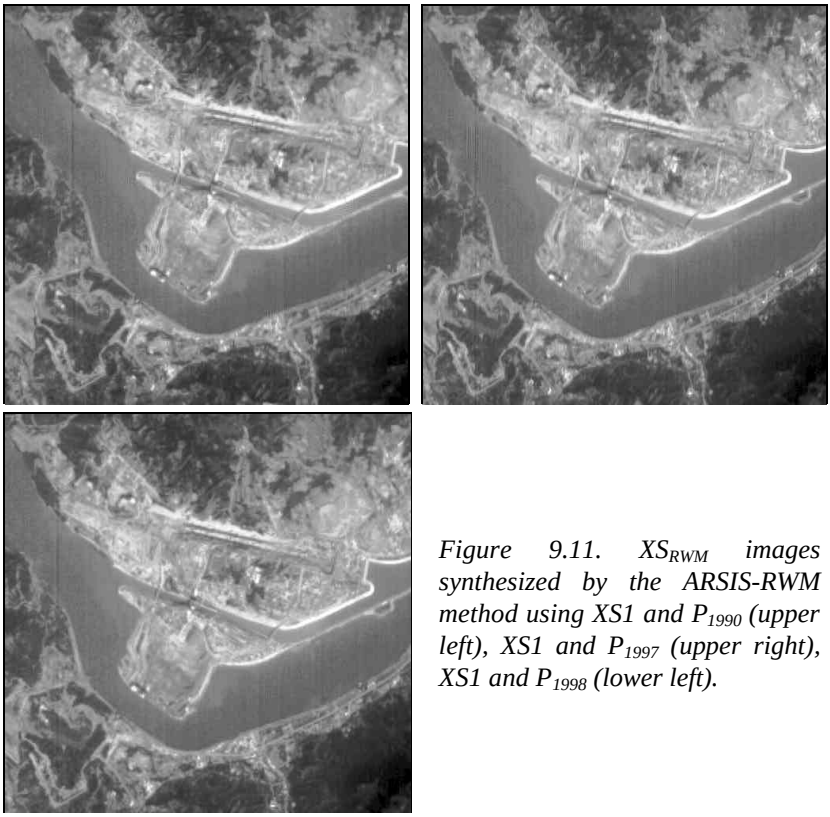


Figure 9.11. XS_{RWM} images synthesized by the ARSIS-RWM method using XS1 and P_{1990} (upper left), XS1 and P_{1997} (upper right), XS1 and P_{1998} (lower left).

Whatever the couples of images *P* and *XS* used, the images synthesized by the ARSIS-RWM method were acknowledged by skilled geologists as given the most exploitable results for interpretation of geological features. The synthesized images are still exploitable for further processing such as classification processes or interpretation of geological features and faults.

Data fusion is a formal framework in which are expressed the means and tools for the alliance of data originating from different sources. It means an approach to information extraction spontaneously adopted in several domains before this was expressed as "data fusion". This approach is based upon the synergy offered by the various sources. Applications are numerous, from biology to civil aviation.

This book clearly establishes the fundamentals (particularly definitions and architectures) in data fusion. It can be read with profit by anyone interested in data fusion, whatever his domain of expertise, and should be valuable to engineers, scientists and practitioners.

The second part of the book is devoted to methods for the fusion of images. It offers an in-depth presentation of standard and advanced methods for the fusion of multi-modality images. The emphasis is put on images having different spatial resolutions, but the book is not limited to this case. Given several sets of images acquired by disparate sensors, the problems treated are to create new sets of images of reduced dimensionality, in order to either better visualize the original sets of images as a comprehensive ensemble of information, or to synthesize images with a better spatial resolution.

Author

L. Wald graduated in Theoretical Physics (France-1977). After his PhD (Paris-1980) on the applications of remote sensing to oceanography, he obtained his Doctorat d'Etat ès Sciences (1985). Since 1991, he is a Professor at Ecole des Mines de Paris, where he is currently the Head of the Remote Sensing Group, and is focusing his own research in applied mathematics, meteorology and oceanography. He obtained the Autometrics Award (1998) and the Erdas Award (2001) for articles on data fusion. His career in information technologies has been rewarded in 1996 by the famous French Blondel Medal.

DATA FUSION

Price : 40 Euros

ECOLE DES MINES DE PARIS
Les Presses