



HAL
open science

Supervised Morphology for Structure Tensor-Valued Images Based on Symmetric Divergence Kernels

Santiago Velasco-Forero, Jesus Angulo

► **To cite this version:**

Santiago Velasco-Forero, Jesus Angulo. Supervised Morphology for Structure Tensor-Valued Images Based on Symmetric Divergence Kernels. First International Conference on Geometric Science of Information (GSI'2013), Aug 2013, Paris, France. pp.543 - 550, <10.1007/978-3-642-40020-9_60>. <hal-01536382>

HAL Id: hal-01536382

<https://minesparis-psl.hal.science/hal-01536382v1>

Submitted on 11 Jun 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

Supervised morphology for structure tensor-valued images based on symmetric divergence kernels

Santiago Velasco-Forero¹ and Jesús Angulo²

¹ ITWM - Fraunhofer Institute, Kaiserslautern, Germany

² CMM-Centre de Morphologie Mathématique,

Mathématiques et Systèmes, MINES ParisTech, France

velascoforero@itwm.fraunhofer.de; jesus.angulo@mines-paristech.fr

Abstract. Mathematical morphology is a nonlinear image processing methodology based on computing min/max operators in local neighbourhoods. In the case of tensor-valued images, the space of SPD matrices should be endowed with a partial ordering and a complete lattice structure. Structure tensor describes robustly the local orientation and anisotropy of image features. Formulation of mathematical morphology operators dealing with structure tensor images is relevant for texture filtering and segmentation. This paper introduces tensor-valued mathematical morphology based on a supervised partial ordering, where the ordering mapping is formulated by means of positive definite kernels and solved by machine learning algorithms. More precisely, we focus on symmetric divergences for SPD matrices and associated kernels.

Keywords: matrix-valued image, tensor mathematical morphology, supervised morphology, structure tensor image processing

1 Introduction

An extensive development of morphological operators has been started with the revolutionary work of Matheron [10] and Serra [13] more than four decades ago. In the long way of the evolution of mathematical morphology, from binary image to vector-valued images, several monographs and articles have illustrated the requirement of a *partial ordering* for an adequate application of the adjunction-based morphological transformations [8,15,1,18]. However, nowadays the world of digital images includes also unconventional acquisition methods as Diffusion Tensor MRI, which produce matrix-valued images. We wish to consider the situation where the available data (pixel information) are symmetric positive definite (SPD) matrices. In that case, we denoted by \mathbf{I} the image defined as a function:

$$\mathbf{I}: \begin{cases} \mathbf{E} \rightarrow \mathcal{P}(d) \\ x \rightarrow \mathbf{X} \end{cases} \quad (1)$$

where $\mathbf{E} \in \mathbb{Z}^2$ for 2D images, is the support space of pixels x and $\mathbf{X} \in \mathcal{P}(d)$ denotes a matrix in the set of SPD matrices of dimension $d \times d$. Genuine matrix-valued concepts with channel interaction are available for nonlinear regularisation methods and related diffusion filters [19], median filtering [20], morphological filters [6] and regularisation [11]. However, we would like to take in consid-

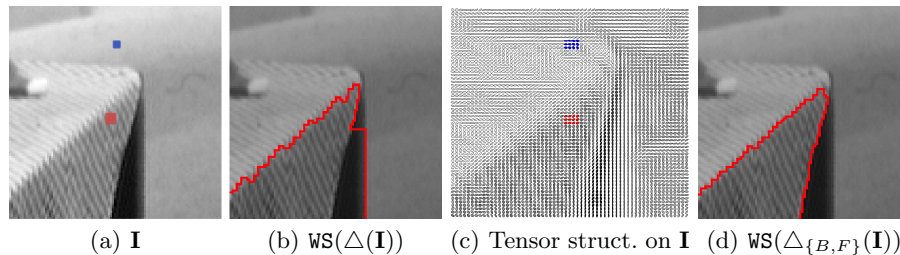


Fig. 1. (a) Red and blue squares show the pixels of sets F and B respectively. Image (c) is the representation of the structure tensor of the grey scale image (a). Pixels in F and B are coloured in red and blue. Image (b) is the marked based watershed on classical gradient, and (d) is the marked based watershed on supervised gradient.

eration some *a priori* information about important objects on the image. In our context, two set of SPD matrices, $B = \{\mathbf{B}_1, \dots, \mathbf{B}_{|B|}\}$ and $F = \{\mathbf{F}_1, \dots, \mathbf{F}_{|F|}\}$ containing local orientation information of pixel associated with background and foreground of the image are available.

Figure 1 illustrates the practical motivation behind our approach. We would like to perform a segmentation on a grey scale image from two sets of markers: one for the background and other for foreground. The classical morphological approach performs marker based watershed on the gradient of the Fig. 1(a). Watershed detects the most contrasted gradient between F and B , as it is illustrated in Fig. 1(b). However, it does not includes any information about the texture of the objects in the image. We characterises the “texture” by a local orientation/anisotropy of gradient, which can vary arbitrarily over the entire image. That means, that every point in the image is associated with a dominant local orientation or degree of anisotropy of the flow pattern. One way of visualising oriented textures is to think about the image intensity surface as being comprised of ellipses, whose direction and height can vary continuously as it is illustrated in Fig. 1(c). This is the rationale behind the notion of structure tensor. We refer keen readers to [4] for a comprehensive review of structure tensors on image processing. We note that structure tensors, defined by the tensor product of gradient vectors, are elements of $\mathcal{P}(d)$. Given the texture information at each pixel extracted from the structure tensor, a main question arises: How that texture information can be included to perform segmentation, smoothing, interpolation and so on?

The main difficulty is that tensor space is not a vector space with the usual operations. In fact, it is the space of SPD matrices form a convex half-cone in the vector space of matrices. In this paper, we limit ourselves to follow the idea introduced in [18] for designing morphological transformation on vector-valued images. Roughly, the idea in [18], can be summarised as follows: To use supervised morphology methods on image $\mathbf{I} : \mathbf{E} \rightarrow \mathcal{P}(d)$, we require the definition of a positive definite kernel (PDK) on $\mathcal{P}(d)$ is required. A mapping $\mathbf{K} : \mathcal{P}(d) \times \mathcal{P}(d) \rightarrow \mathbb{R}$, so that $\mathbf{K}(\mathbf{A}, \mathbf{B}) = \mathbf{K}(\mathbf{B}, \mathbf{A})$ and $\sum_{i=1}^n \sum_{j=1}^n \mathbf{K}(\mathbf{A}_i, \mathbf{A}_j) c_i c_j \geq 0$ for all finite sequence of matrices $\mathbf{A}_1, \dots, \mathbf{A}_n \in \mathcal{P}(d)$ and all choices of real numbers c_1, \dots, c_n . From \mathbf{K} a supervised classification problem between B and F is solved and the evaluation function can be utilised as a partial ordering to perform mathematical morphology transformation. Following the same methodology, a morphological transformation can be designed for structure tensor-valued images by building a PDK on $\mathcal{P}(d)$. The usual approach of computing kernels is the application of univariate kernel on the Euclidean Distance. However, since the space of positive semidefinite matrices is a non-Euclidean space, it is more natural to consider alternatives distances.

2 From Bregman matrix divergences to PDK on $\mathcal{P}(d)$

2.1 Matrix spaces

Let $\mathcal{M}(d)$ be the set of $d \times d$ real matrices and $GL(d)$ be its subset containing only non-singular matrices. $GL(d)$ is a Lie group, i.e., a group which is a differentiable manifold and for which the operations of group multiplication and inverse are smooth. In $\mathcal{M}(d)$ we shall use the Euclidean (or Frobenius) inner product defined by $\langle \mathbf{A}, \mathbf{B} \rangle_F = \text{tr}(\mathbf{A}^T \mathbf{B})$, where $\text{tr}(\cdot)$ stands for the trace and the superscript T denotes the transpose. Additionally, let $\mathcal{S}(d) = \{\mathbf{A} \in \mathcal{M}(d), \mathbf{A}^T = \mathbf{A}\}$ be the space of all $d \times d$ symmetric matrices, and $\mathcal{P}(d) = \{\mathbf{A} \in \mathcal{S}(d), \mathbf{A} > 0\}$ be the set of all $d \times d$ positive-definite symmetric matrices. Here, $\mathbf{A} > 0$ means that the quadratic form $\mathbf{x}^T \mathbf{A} \mathbf{x} > 0$ for all $\mathbf{x} \in \mathbb{R}^d, \mathbf{x} \neq \mathbf{0}$. It is well known that $\mathcal{P}(d)$ is an open convex cone, i.e., if \mathbf{A} and \mathbf{B} are in $\mathcal{P}(d)$, so is $\mathbf{A} + c\mathbf{B}$ for any $c > 0$.

2.2 Bregman matrix divergences

A possibility to measure the nearness between two matrices are the *Bregman matrix divergences* [9], which are generalisations of Bregman vector divergences. Let ϕ be a real-valued strictly convex function defined over set $\text{dom}(\phi) \subset \mathbb{R}^d$ such that ϕ is differentiable on the relative interior of $\text{dom}(\phi)$. The *Bregman vector divergence* [3] with respect to ϕ is defined as:

$$\text{div}_\phi(\mathbf{x}, \mathbf{y}) = \phi(\mathbf{x}) - \phi(\mathbf{y}) - (\mathbf{x} - \mathbf{y})^T \nabla \phi(\mathbf{y}) \quad (2)$$

For instance, if $\phi(\mathbf{x}) = \mathbf{x}^T \mathbf{x}$ then $\text{div}_\phi(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|_2^2$. If $\phi(\mathbf{x}) = \sum_i x_i \log x_i - x_i$ the $\text{div}_\phi(\mathbf{x}, \mathbf{y}) = \sum_i (x_i \log \frac{x_i}{y_i} - x_i + y_i)$, corresponds to the unnormalised relative entropy. Bregman divergences generalise many properties of squared

loss and relative entropy. See for more details in [12]. We can naturally extend this definition to real, symmetric $d \times d$ matrices, in $\mathcal{S}(d)$. Given a strictly convex, differentiable function $\phi : \mathcal{S}(d) \rightarrow \mathbb{R}$, the *Bregman matrix divergence* is defined to be:

$$\text{div}_\phi(\mathbf{A}, \mathbf{B}) = \phi(\mathbf{A}) - \phi(\mathbf{B}) - \text{tr}((\nabla\phi(\mathbf{B}))^T(\mathbf{A} - \mathbf{B})), \quad (3)$$

Particular cases of Bregman divergences for matrices are the following:

- if $\phi_{\text{Frob}}(\mathbf{A}) = \|\mathbf{A}\|_F^2$ leads to the well-know Frobenius norm.
- if $\phi_{\text{Neuman}}(\mathbf{A}) = \text{tr}(\mathbf{A} \log \mathbf{A} - \mathbf{A})$, where $\log \mathbf{A}$ is the matrix logarithm³, the $\nabla\phi_{\text{Neuman}}(\mathbf{B}) = (\log \mathbf{B})^T$ and the resulting Bregman divergence is:

$$\text{div}_{\text{Neuman}}(\mathbf{A}, \mathbf{B}) = \text{tr}(\mathbf{A} \log \mathbf{A} - \mathbf{A} \log \mathbf{B} - \mathbf{A} + \mathbf{B}) \quad (4)$$

- Another important matrix divergence arises by taking the Burg entropy of the eigenvalues, that is, $\phi_{\text{Burg}}(\mathbf{A}) = -\log \det(\mathbf{A})$, so $\nabla\phi_{\text{Burg}}(\mathbf{B}) = -(\mathbf{B}^{-T})$, where we obtain:

$$\text{div}_{\text{Burg}}(\mathbf{A}, \mathbf{B}) = \text{tr}(\mathbf{A}\mathbf{B}^{-1}) - \log \det(\mathbf{A}\mathbf{B}^{-1}) - d. \quad (5)$$

Expression (5) is also known as *Stein's loss*[17] or the *LogDet-Divergence* [9,16].

The most obvious computational benefit of using the divergences arise from the fact that they are defined over positive definite matrices. Because of this, our algorithms not need to explicitly constrain our learned matrices to be positive definite. However, Bregman divergences are non negative and definite, but almost always asymmetric.

2.3 Symmetric divergence and associated kernels

Despite the broad applicability of Bregman divergence, their asymmetry is something undesirable. This drawback prompted researchers to consider symmetric divergences [7,16], among which the most popular is the *Jensen-Shannon divergence*

$$S_\phi(\mathbf{A}, \mathbf{B}) = \text{div}_\phi\left(\mathbf{A}, \frac{\mathbf{A} + \mathbf{B}}{2}\right) + \text{div}_\phi\left(\frac{\mathbf{A} + \mathbf{B}}{2}, \mathbf{B}\right) \quad (6)$$

Applying (6) in the case of (5), we obtain the *symmetric stein divergence* [16] defined as follows

$$S_{\text{stein}}(\mathbf{A}, \mathbf{B}) = \log \det\left(\frac{\mathbf{A} + \mathbf{B}}{2}\right) - \frac{1}{2} \log \det(\mathbf{A}\mathbf{B}) \quad (7)$$

The PDK is obtained from the Stein symmetric divergence using an important result from [16].

³ If $\mathbf{A} = \mathbf{V}\Lambda\mathbf{V}^T \in \mathcal{P}(d)$ is the eigendecomposition, then $\log \mathbf{A} = \mathbf{V} \log \Lambda \mathbf{V}^T$, where $\log \Lambda$ is the diagonal matrix whose entries contain the logarithm of the eigenvalues. The matrix exponential can be defined analogously.

Theorem 1. [16] Define $\text{div}_{\text{stein}}(\mathbf{A}, \mathbf{B}) = \sqrt{S_{\text{stein}}(\mathbf{A}, \mathbf{B})}$. Then, $\text{div}_{\text{stein}}$ is a metric on $\mathcal{P}(d)$.

Theorem 2. [16] Let $\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_n$ be real symmetric matrices in $\mathcal{P}(d)$, the following functions

$$\mathbf{K}(\mathbf{A}_i, \mathbf{A}_j) = \exp(-\beta S_{\text{stein}}(\mathbf{A}_i, \mathbf{A}_j)) \quad (8)$$

forms a PDK if and only if β satisfies

$$\beta \in \left\{ \frac{j}{2} : j \in \mathbb{N} | 1 \leq j \leq (d-1) \right\} \cup \{j : j \in \mathbb{R} | j > \frac{1}{2}(d-1)\}$$

3 Supervised morphology on structure tensor-valued images

3.1 Orders on $\mathcal{P}(d)$

Loewner partial order There are also a natural ordering on $\mathcal{S}(d)$, the so-called *Loewner's partial ordering* defined via the cone of positive semidefinite matrices $\mathcal{S}(d)$ by

$$\mathbf{A} \leq_{\text{Loewner}} \mathbf{B} \iff \mathbf{A} - \mathbf{B} \in \mathcal{S}(d), \forall \mathbf{A}, \mathbf{B} \in \mathcal{S}(d) \quad (9)$$

i.e., if and only if $\mathbf{A} - \mathbf{B}$ is positive semidefinite. This definition is equivalent to the statement [14]:

$$\mathbf{A} \leq_{\text{Loewner}} \mathbf{B} \iff \mathbf{x}^T \mathbf{A} \mathbf{x} \leq \mathbf{x}^T \mathbf{B} \mathbf{x} \quad (10)$$

for all non-zero \mathbf{x} such that $\mathbf{x}^T \mathbf{x} = 1$. The application of this order to induce mathematical morphology transformation have been introduced in [5].

Lexicographic spectral order Recently, in [2] was introduced the idea of ordering based on the singular value decomposition, as follows:

$$\mathbf{A} \leq_{\text{LS}} \mathbf{B} \iff \exists j, 1 \leq j \leq d \text{ such that } \lambda_i(\mathbf{A}) = \lambda_i(\mathbf{B}) \forall i < j, \\ \text{and } \lambda_j(\mathbf{A}) < \lambda_j(\mathbf{B}) \text{ for } j \leq d$$

where $\lambda_i(\mathbf{A}), i = 1, \dots, d$ are the ordered eigenvalues of \mathbf{A} .

h-supervised order Following the approach in [18] which takes advantage of supervised classification approaches to define a partial ordering, we introduce a particular case for matrices in $\mathcal{P}(d)$. Given two sets $B = \{\mathbf{B}_1, \dots, \mathbf{B}_{|B|}\}$ and $F = \{\mathbf{F}_1, \dots, \mathbf{F}_{|F|}\}$ such that \mathbf{F}_i and $\mathbf{B}_f \in \mathcal{P}(d)$, for all $i = 1, \dots, |F|$ and $j = 1, \dots, |B|$,

$$\mathbf{A} \leq_{\text{Super}}^{B,F} \mathbf{B} \iff h_{B,F}(\mathbf{A}) \leq h_{B,F}(\mathbf{B}) \quad (11)$$

where the supervised h -mapping, $h : \mathcal{P}(d) \rightarrow \mathbb{R}$, is given by

$$h_{B,F}(\mathbf{X}) = \sum_{i=1}^{|F|} \lambda_i \mathbf{K}(\mathbf{X}, \mathbf{F}_i) - \sum_{j=1}^{|B|} \lambda_j \mathbf{K}(\mathbf{X}, \mathbf{B}_j) \quad (12)$$

and $\mathbf{K}(\cdot, \cdot)$ is a PDK and λ_j are computed as solution of the optimisation problem associated with a supervised classification problem between F and B , see details in [18].

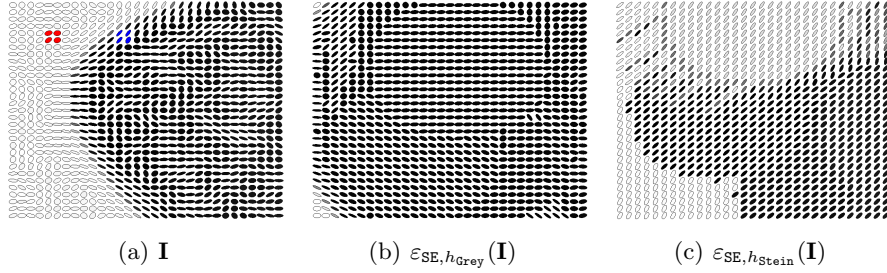


Fig. 2. Background and foreground sets are blue and red ellipses in (a). Note that the erosion by h_{Grey} causes a colour homogenisation in comparison to an orientation uniformity by h_{Stein} . SE is a disk of diameter 20 pixels.

3.2 Morphological operators

Standard morphological operators use the so-called *structuring element*, denoted by SE, to work on images represented by scalar functions $\mathbf{I}(x, y)$ with $(x, y) \in \mathbf{E}$. Grey scale dilation $\delta_{SE}(\cdot)$, resp., erosion $\varepsilon_{SE}(\cdot)$ is defined by:

$$\forall x \in \mathbf{E}, \quad \delta_{SE}(\mathbf{I})(x) = \bigvee_{y \in \tilde{SE}(x)} \mathbf{I}(y), \quad \varepsilon_{SE}(\mathbf{I})(x) = \bigwedge_{y \in SE(x)} \mathbf{I}(y) \quad (13)$$

where $SE(x) \in \mathbf{E}$ denote the spatial neighbourhood induced by the structuring element SE centred at x and \tilde{SE} is the structuring element mirrored in the origin. A structure tensor image is considered as a mapping $\mathbf{I} : \mathbf{E} \rightarrow \mathcal{P}(d)$. Our solution involves a structure of total ordering which allows to compute directly the h -erosion and h -dilation formulated as

$$\begin{aligned} \varepsilon_{SE, h_{B,F}}(\mathbf{I})(x) &= \{\mathbf{I}(y) | h_{B,F}(\mathbf{I})(y) = \bigwedge_{z \in SE_x} h_{B,F}(z)\}, \\ \delta_{SE, h_{B,F}}(\mathbf{I})(x) &= \{\mathbf{I}(y) | h_{B,F}(\mathbf{I})(y) = \bigvee_{z \in \tilde{SE}_x} h_{B,F}(z)\}, \end{aligned}$$

3.3 Application to robust segmentation

A watershed transform, denoted by $\text{WS}(\mathbf{I})$ associate a catch basin to each minimum of the image \mathbf{I} . We note in passing that in practice one often does not apply the watershed transform to the original image, but to its (morphological) gradient [15]. Basically, we apply the watershed transformation in the gradient induced by the h -ordering calculated⁴ by the kernel in $\mathcal{P}(d)$, i.e., $\Delta_{h_{B,F}}(\mathbf{I}) = h_{B,F}(\delta_{\text{SE},h_{B,F}}(\mathbf{I}) - \varepsilon_{\text{SE},h_{B,F}}(\mathbf{I}))$, where SE is typically a unitary structuring element. We notice that the gradient image, $\Delta_{h_{B,F}}(\mathbf{I})$ is scalar function and consequently standard watershed algorithm can be applied on it. In the case of the example illustrated in Fig.3(a), markers in (b) are included in the classical watershed to produce only two regions. Additionally, they are the sets B and F to compute the h -mapping in (12).

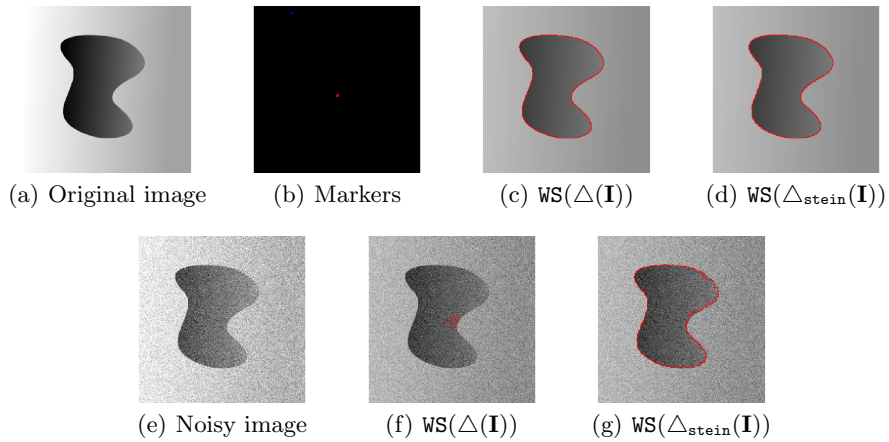


Fig. 3. Note that marked based watershed on the supervised gradient by Stein-Kernel performs similar than classical one for the image (a). Adding random noise (e), we observe that supervised gradient segment correctly the object (g) in opposition to classical approach(e).

References

1. Angulo, J.: Morphological colour operators in totally ordered lattices based on distances: Application to image filtering, enhancement and analysis. *Comput. Vis. Image Underst.* 107(1-2), 56–73 (Jul 2007)

⁴ Two orders are considered: h_{grey} by $\mathbf{K}(x, y) = \exp(-\alpha(x - y)^2)$ and h_{stein} by $\mathbf{K}(x, y) = \exp(-\beta S_{\text{stein}}(\mathbf{A}_x, \mathbf{A}_y))$

2. Angulo, J.: Supremum/infimum and nonlinear averaging of positive definite symmetric matrices. In: Nielsen, F., Bhatia, R. (eds.) *Matrix Information Geometry*. pp. 3–33 (2012)
3. Bregman, L.: The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. *USSR Comp. Mathematics and Mathematical Physics* 7, 200–217 (1967)
4. Brox, T., Weickert, J., Burgeth, B., Mrázek, P.: Nonlinear structure tensors. *Image Vision Comput.* 24(1), 41–55 (Jan 2006), <http://dx.doi.org/10.1016/j.imavis.2005.09.010>
5. Burgeth, B., Papenberg, N., Bruhn, A., Welk, M., Feddern, C., Weickert, J.: Morphology for higher-dimensional tensor data via loewner ordering. In: C. Ronse, L. Najman, E.D. (ed.) *Proceedings of International Symposium on Mathematical Morphology (ISMM 2005)*. vol. 30, pp. 407–416. Springer, Dordrecht (2005)
6. Burgeth, B., Welk, M., Feddern, C., Weickert, J.: Morphological operations on matrix-valued images. In: *The 8 th European Conference on Computer Vision*. pp. 155–167 (2004)
7. Cherian, A., Sra, S., Banerjee, A., Papanikolopoulos, N.: Jensen-bregman logdet divergence with application to efficient similarity search for covariance matrices. *IEEE Transactions on Pattern Analysis & Machine Intelligence* (10 2012)
8. Heijmans, H., Ronse, C.: The algebraic basis of mathematical morphology i. dilations and erosions. *Computer Vision, Graphics, and Image Processing* 50(3), 245 – 295 (1990)
9. Kulis, B., Sustik, M.A., Dhillon, I.S.: Low-rank kernel learning with bregman matrix divergences. *J. Mach. Learn. Res.* 10, 341–376 (Jun 2009)
10. Matheron, G.: *Éléments pour une théorie des milieux poreux*. Masson et Cie (1967)
11. Moakher, M., Zéraï, M.: The riemannian geometry of the space of positive-definite matrices and its application to the regularization of positive-definite matrix-valued data. *Journal of Mathematical Imaging and Vision* 40(2), 171–187 (2011)
12. Nielsen, F., Nock, R.: Sided and symmetrized bregman centroids. *IEEE Trans. Inf. Theor.* 55(6), 2882–2904 (Jun 2009), <http://dx.doi.org/10.1109/TIT.2009.2018176>
13. Serra, J.: *Image Analysis and Mathematical Morphology*. Academic Press, Inc., Orlando, FL, USA (1983)
14. Siotani, M.: Some applications of loewner’s ordering on symmetric matrices. *Annals of the Institute of Statistical Mathematics* 19, 245–259 (1967), <http://dx.doi.org/10.1007/BF02911678>
15. Soille, P.: *Morphological Image Analysis*. Springer-Verlag (1999)
16. Sra, S.: Positive definite matrices and the symmetric stein divergence (2012), preprint: (arXiv:1110.1773)
17. Stein, C.: Inadmissibility of the usual estimator for the mean of a multivariate distribution. In: *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*. vol. 1, pp. 197–206 (1956)
18. Velasco-Forero, S., J. Angulo: Supervised ordering in \mathbb{R}^p : Application to morphological processing of hyperspectral images. *IEEE Transactions on Image Processing* 20(11), 3301–3308 (2011)
19. Weickert, J., Brox, T.: Diffusion and regularization of vector- and matrix-valued images, *Contemp. Math.*, vol. 313, pp. 251–268. Amer. Math. Soc., Providence, RI (2002)
20. Welk, M., Weickert, J., Becker, F., Schnörr, C., Feddern, C., Burgeth, B.: Median and related local filters for tensor-valued images. *Signal Processing* 87(2), 291–308 (2007)