



**HAL**  
open science

## FEEDBACK ON A PUBLICLY DISTRIBUTED IMAGE DATABASE: THE MESSIDOR DATABASE

Etienne Decencière, Xiwei Zhang, Guy Cazuguel, Bruno Laÿ, Béatrice Cochener, Caroline Trone, Philippe Gain, John-Richard Ordóñez-Varela, Pascale Massin, Ali Erginay, et al.

► **To cite this version:**

Etienne Decencière, Xiwei Zhang, Guy Cazuguel, Bruno Laÿ, Béatrice Cochener, et al.. FEEDBACK ON A PUBLICLY DISTRIBUTED IMAGE DATABASE: THE MESSIDOR DATABASE. *Image Analysis & Stereology*, 2014, pp.231-234. 10.5566/ias.1155 . hal-01082570

**HAL Id: hal-01082570**

<https://minesparis-psl.hal.science/hal-01082570v1>

Submitted on 1 Dec 2014

**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.

# FEEDBACK ON A PUBLICLY DISTRIBUTED IMAGE DATABASE: THE MESSIDOR DATABASE

ETIENNE DECENCIÈRE<sup>1</sup>, XIWEI ZHANG<sup>1</sup>, GUY CAZUGUEL<sup>2,8</sup>, BRUNO LAY<sup>3</sup>, BÉATRICE COCHENER<sup>4,8</sup>, CAROLINE TRONE<sup>5</sup>, PHILIPPE GAIN<sup>5</sup>, JOHN-RICHARD ORDÓÑEZ-VARELA<sup>1</sup>, PASCALE MASSIN<sup>6</sup>, ALI ERGINAY<sup>6</sup>, BÉATRICE CHARTON<sup>7</sup> AND JEAN-CLAUDE KLEIN<sup>1</sup>

<sup>1</sup>MINES ParisTech, PSL Research University, Centre for mathematical morphology, Fontainebleau, France, <sup>2</sup>Télécom Bretagne, Institut Mines-Télécom, ITI Department, Brest, France, <sup>3</sup>ADCIS, Saint-Contest, France, <sup>4</sup>Department of Ophthalmology, Brest University Hospital, Brest, France, <sup>5</sup>Department of Ophthalmology, University Hospital of Saint-Etienne, Saint-Etienne, France, <sup>6</sup>Service d'ophtalmologie, Hôpital Lariboisière, AP-HP, Paris, France, <sup>7</sup>Centre de Ressources Informatiques de Haute-Normandie, France, <sup>8</sup>LaTIM - INSERM UMR 1101, SFR ScInBioS (IFR 148), Brest, F-29200 France  
(Submitted)

## ABSTRACT

The Messidor database, which contains hundreds of eye fundus images, has been publicly distributed since 2008. It was created by the Messidor project in order to evaluate automatic lesion segmentation and diabetic retinopathy grading methods. Designing, producing and maintaining such a database entails significant costs. By publicly sharing it, one hopes to bring a valuable resource to the public research community. However, the real interest and benefit of the research community is not easy to quantify. We analyse here the feedback on the Messidor database, after more than 6 years of diffusion. This analysis should apply to other similar research databases.

Keywords: Image database, Messidor, diabetic retinopathy, image processing.

## INTRODUCTION

Public databases are precious tools for researchers. They bring the necessary data to develop and test new methods, and allow for quantitative comparisons between different approaches. The Messidor database is one of such databases. It was created within the Messidor project to evaluate different lesion segmentation methods for color eye fundus images, in the framework of diabetic retinopathy screening and diagnosis. It has been publicly distributed since 2008.

Other databases are available for researchers on retinal image analysis. Concerning the segmentation of retinal blood vessels, the DRIVE (Staal *et al.* (2004)) and STARE (Hoover *et al.* (2000)) databases have become *de facto* standards. DRIONS-DB contains retinal images for optic nerve head segmentation benchmarking (Carmona *et al.* (2008)). Concerning retinal lesions, like microaneurysms and exudates, several databases are available such as DIARETDB0, DIARETDB1 (Kauppi *et al.* (2007)), HEI-MED (Giancardo *et al.* (2012)), ROC (Niemeijer *et al.* (2010)) and e-optha (Decencièrre *et al.* (2013)).

## THE MESSIDOR DATABASE

The Messidor download page <sup>1</sup> gives an appropriate description of the database, which we quote here:

“The 1200 eye fundus color numerical images of the posterior pole for the MESSIDOR database were acquired by 3 ophthalmologic departments using a color video 3CCD camera on a Topcon TRC NW6 non-mydratic retinograph with a 45 degree field of view. The images were captured using 8 bits per color plane at 1440\*960, 2240\*1488 or 2304\*1536 pixels.

800 images were acquired with pupil dilation (one drop of Tropicamide at 0.5%) and 400 without dilation.

The 1200 images are packaged in 3 sets, one per ophthalmologic department. Each set is divided into 4 zipped sub sets containing each 100 images in TIFF format and an Excel file with medical diagnoses for each image.”

Note that, as the description indicates, the database contains a medical diagnosis for each image, but no manual annotations on the images, such as lesions contours or position. This is an important difference with respect to other databases, such as DIARETDB1 and e-optha.

<sup>1</sup><http://messidor.crihan.fr/download-en.php>

The download procedure asks the user to fill in the following fields: E-mail address; First Name; Last Name; Professional Interests; Country and University/Organization. An e-mail is then sent to a member of the Messidor team, who checks the validity of the request, and sends an appropriate link to the submitter. Some requests are not accepted, typically because the fields requested in the download procedure are clearly incorrectly filled. Precise statistics on refused requests are not kept, but we estimate that they represent less than 25% of the total number of requests. They are not taken into account in the statistics below.

It should be noted that Messidor database users are asked to acknowledge the Messidor project partners in their related publications.

## EXPERIENCE FEEDBACK ON MESSIDOR

Most of the statistics on the Messidor database diffusion presented in this section are summarized in Fig. 1.

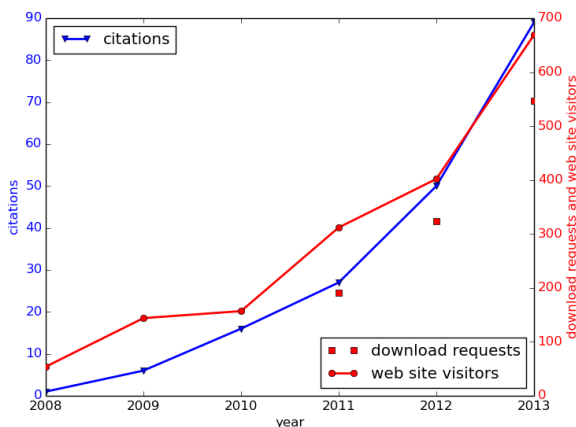


Fig. 1. *Evolution of number of citations, web site visitors and download requests.*

People tend to underestimate support and maintenance costs associated to a publicly distributed database. For instance, given the increasing number of download requests for the Messidor database, processing these requests and related questions requires approximately one hour per week. On top of that, users ask general questions about the database - even if most answers are available in the website. Finally, hosting the database and web pages also takes resources.

Table 1 gives the number of download requests per year between 2011 and 2013, broken into different countries. It can be seen that download requests clearly

increase over time: there have approximately been three time more requests in 2013 than in 2011. This increase comes mainly from less developed countries.

Table 1. *Download requests for the Messidor database, per year. Some countries, where only few requests originated, are not indicated.*

Country	2011	2012	2013
Algeria	7	6	7
Australia	0	2	4
Brazil	3	8	4
Canada	2	1	5
China	10	14	31
Egypt	0	9	19
France	2	5	5
Germany	3	3	4
India	84	141	280
Indonesia	4	12	33
Iran	9	19	19
Iraq	0	0	8
Malaysia	9	12	16
Mexico	3	2	9
Netherlands	3	1	3
Portugal	10	6	4
Spain	6	5	10
UK	7	9	7
USA	7	19	7
Other countries	22	50	72
<b>Total</b>	<b>191</b>	<b>324</b>	<b>547</b>

Another measure on the success of the database can be obtained through access statistics to the corresponding web page (see Table 2). Again, one can see a clear increase in web site access since 2008. The number of visitors is approximately two times higher in 2013 than in 2011. This trend clearly appears in Fig. 1.

Table 2. *Web statistics on the access to <http://messidor.crihan.fr>.*

Year	Visitors	Visits	Pages
2008	54	129	561
2009	144	276	1011
2010	157	301	1761
2011	312	476	2706
2012	402	644	3913
2013	669	968	5595

The link between download requests or web access and the actual contribution to the research domain is not necessarily simple to apprehend. Indeed, people might download the database or consult the web site

for reasons not related to public research. In order to clarify this point, we have looked into the number of citations of the Messidor database in scientific papers. The results are summarized in table 3. Interestingly, it can be seen that the Messidor database has been cited three times more often in 2013 than in 2011 - the same increase as for the number of download requests (see Fig. 1).

Table 3. Citations per year. Values were obtained through Google Scholar using the keywords “Messidor diabetic retinopathy” on June 19, 2014.

Year	Citations
2008	1
2009	6
2010	16
2011	27
2012	50
2013	89
Total	189

Finally, if we pool the results for two of the most cited journals in the field of biomedical image processing, that is Medical Image Analysis and IEEE Transactions of Medical Imaging, we find that, since 2008, 47 papers deal with “diabetic retinopathy”, and among these 10 papers cite the Messidor database.

Note that other databases used in the same domain follow similar trends. DIARETDB1, which has been distributed since 2007, has been cited 295 times (as of June 19, 2014), while HEI-MED, which was established in 2012, 26 times.

## CONCLUSION

The Messidor database has been publicly distributed since 2008. It is of interest mainly for researchers in a relatively specialized domain: retinal image processing, and more specifically computer-assisted diagnosis of diabetic retinopathy. In spite of this, it has gathered a large amount of citations. We have also shown that the number of web site visitors, as well as the number of download requests, seem to be correctly correlated with the number of citations, which provides a simple and convenient method to monitor the success of a database.

The experience gathered by our team on the management of the Messidor database allows us to propose some recommendations for the design of future databases:

- Hosting and managing the database takes resources; this point should be taken into account during the database design, in order to reduce this cost as much as possible.
- The database is typically described on a web page. This description has to be clear and complete, in order to limit the number of requests for additional information (and therefore to reduce the management cost).
- The database managers should ask potential users to acknowledge the database or, better, to cite a relevant paper on the database. This simplifies the evaluation of the success of the database.
- Last but not least, we have shown that an automatic validation procedure seems to be enough to treat download requests.

Moreover, we believe that this study confirms the important role that databases play in medical image processing. In the case of the Messidor database, this is true in spite of the fact that the images contained in the database are progressively getting outdated. Indeed, they were acquired before 2007, and modern fundus cameras offer increasing image resolutions and sensitivities. As far as we know, only two databases have been released in this field after 2010: HEI-MED (for exudate-based macula oedema detection) and e-ophta (microaneurysms and exudates segmentation). This stresses the importance of new databases, corresponding to the current clinical practice.

## ACKNOWLEDGEMENTS

The Messidor project was funded by the French Techno-Vision program. The Messidor database and site have been hosted by Crihan (<http://www.crihan.fr>) since 2008. Researchers from the Center for Mathematical Morphology, including Adnan Rashid and Estelle Parra-Denis, have offered support to the Messidor database users and processed their download requests for several years.

## REFERENCES

- Carmona EJ, Rincón M, García-Feijoó J, Martínez-de-la Casa JM (2008). Identification of the optic nerve head with genetic algorithms. *Artificial Intelligence in Medicine* 43:243–59.
- Decencièrre E, Cazuguel G, Zhang X, Thibault G, Klein JC, Meyer F, Marcotegui B, Quéllec G, Lamard M, Danno R, Elie D, Massin P, Viktor Z, Erginay A, Laÿ B, Chabouis A (2013). *TeleOphta*:

- Machine learning and image processing methods for teleophthalmology. *IRBM* 34:196–203.
- Giancardo L, Meriaudeau F, Karnowski TP, Li Y, Garg S, Tobin Jr. KW, Chaum E (2012). Exudate-based diabetic macular edema detection in fundus images using publicly available datasets. *Medical Image Analysis* 16:216–26.
- Hoover A, Kouznetsova V, Goldbaum M (2000). Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response. *IEEE Transactions on Medical Imaging* 19:203–10.
- Kauppi T, Kalesnykiene V, Kamarainen JK, Lensu L, Sorri I, Raninen A, Voutilainen R, Uusitalo H, Kälviäinen H, Pietilä J (2007). The DIARETDB1 diabetic retinopathy database and evaluation protocol. In: *British Machine Vision Conference*.
- Niemeijer M, van Ginneken B, Cree M, Mizutani A, Quellec G, Sanchez C, Zhang B, Hornero R, Lamard M, Muramatsu C, Wu X, Cazuguel G, You J, Mayo A, Li Q, Hatanaka Y, Cochener B, Roux C, Karray F, Garcia M, Fujita H, Abramoff M (2010). Retinopathy online challenge: Automatic detection of microaneurysms in digital color fundus photographs. *IEEE Transactions on Medical Imaging* 29:185–95.
- Staal J, Abramoff M, Niemeijer M, Viergever M, van Ginneken B (2004). Ridge-based vessel segmentation in color images of the retina. *IEEE Transactions on Medical Imaging* 23:501–9.