Distortion Adaptive Descriptors: Extending Gradient-Based Descriptors to Wide Angle Images

Antonino Furnari[1(✉)], Giovanni Maria Farinella[1], Arcangelo Ranieri Bruna[2], and Sebastiano Battiato[1]

[1] Department of Mathematics and Computer Science, University of Catania, 95125 Catania, Italy
{furnari,gfarinella,battiato}@dmi.unict.it
http://iplab.dmi.unict.it
[2] Advanced System Technology - Computer Vision, STMicroelectronics, 95121 Catania, Italy
arcangelo.bruna@st.com

Abstract. Gradient-based descriptors have proven successful in a wide variety of applications. Their standard implementations usually assume that the input images have been acquired using classic perspective cameras. In practice many real-world systems make use of wide angle cameras which allow to obtain wider Fields of View (FOV) but introduce radial distortion which breaks the rectilinear assumption. The most straightforward way to overcome such a problem is to compensate the distortion by unwarping the original image prior to computing the descriptor. The rectification process, however, is computationally expansive and introduces artefacts which can deceive the subsequent analysis (e.g., feature matching). We propose the Distortion Adaptive Descriptors (DAD), a new paradigm to correctly compute local descriptors directly in the distorted domain. We combine the DAD with existing techniques to correctly estimate the gradient of distorted images and hence derive a set of SIFT and HOG-based descriptors. Experiments show that the DAD paradigm allows to improve the matching ability of the SIFT and HOG descriptors when they are computed directly in the distorted domain.

Keywords: Gradient-based descriptors · Wide angle images · Gradient estimation · SIFT · HOG

1 Introduction

Most Computer Vision applications assume that the input images have been acquired using cameras employing a perspective projection model. The perspective projection has the convenient property that straight lines in the real world are mapped to straight lines in the image, producing a representation of the scene which is coherent with our perception. Unfortunately, the perspective projection

V. Murino and E. Puppo (Eds.): ICIAP 2015, Part II, LNCS 9280, pp. 205–215, 2015.
DOI: 10.1007/978-3-319-23234-8_20

is not suitable to build wide angle cameras covering large Fields Of View (FOV) up to 180°. Those cameras indeed require different designs which are grouped into two main categories: dioptric [2] and catadioptric [4,5]. Due to their ability to acquire a large portion of the scene, wide angle cameras are of great interest in many fields (e.g., automotive, surveillance and robotics [12,13]) as they allow to use a single wide angle camera in place of more perspective ones. However, all wide angle cameras have inherent distortion since it is not possible to form an image of an hemispheric field on a plane without distortion [2]. Under given conditions, the distortion is radially symmetric and can be modelled as an invertible function. When the distortion function is known, the wide angle images can be rectified in order to obtain data compliant with the perspective model. Unfortunately, the rectification process is computationally expansive and the undistorted images contain artefacts that may affect the subsequent computation [6]. To overcome these problems, it would be ideal to process the images directly in the distorted domain (i.e., avoiding rectification) paying attention to the geometry of the deformation affecting the images. Many endeavours towards this direction exist: in [6,8] the Scale Invariant Feature Transform (SIFT) pipeline [9] is modified in order to be used directly on wide angle images, in [7] a direct approach to detect people using omnidirectional cameras is proposed, in [1] it is studied how affine covariant features can be reliably detected on fisheye images, finally in [6,10] methods to estimate geometrically correct gradients of distorted images are investigated.

In this paper we study how gradient based descriptors can be modified in order to be computed directly in the distorted domain. We propose the Distortion Adaptive Descriptors (DAD), a new paradigm for computing local descriptors directly on the distorted images. We combine the DAD paradigm with existing methods for the correct estimation of the gradient of distorted images in order to derive distortion adaptive variants of the SIFT [9] and Histogram of Oriented Gradients (HOG) [11] descriptors. The adaptation of such descriptors to the distorted domain, virtually enables a number of applications in which they have proven to be successful, such as object and people detection [9,11], video stabilization [14], object class recognition [15] and panorama stitching [16]. Experiments show that the DAD variants significantly outperform the regular SIFT and HOG descriptors when they are applied directly in the distorted domain. Moreover, we show that there is still space for improving the gradient estimation techniques. The remainder of the paper is organized as follows: in Section 2 we discuss the distortion model adopted in this paper and briefly review the considered techniques for the gradient estimation of distorted images; in Section 3 we introduce the Distortion Adaptive Descriptors; Section 4 discusses the experimental settings, whereas Section 5 reports the results. Finally Section 6 draws the conclusions.

2 The Distorted Domain

In the rest of the paper we will assume that a distortion function $f : \Re^2 \rightarrow \Re^2$ which maps the point $\mathbf{u} \in \Re^2$ in the rectilinear space to the corresponding

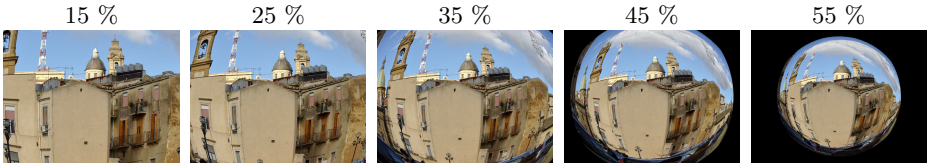


Fig. 1. Different amounts of radial distortion artificially added to a sample image.

point $\mathbf{x} = f(\mathbf{u})$ in the distorted space is known and invertible. Moreover, we will assume that both functions f and f^{-1} can be linearly approximated locally with a small error. These two conditions are easily met in real systems as pointed out in [1]. If we denote the acquired wide angle image by $\hat{I}(\mathbf{x})$, its undistorted counterpart is defined as $I(\mathbf{u}) = \hat{I}(f(\mathbf{u}))$. Since $\hat{I}$ is a discrete function, I is actually reconstructed by interpolation. In this paper we use the division model [6,17] as the distortion model. Its distortion function is defined as:

$$\mathbf{x} = f(\mathbf{u}) = \frac{2\mathbf{u}}{1 + \sqrt{1 - 4 \cdot \xi ||\mathbf{u}||^2}} \tag{1}$$

where the distortion parameter $\xi < 0$ can be used to adjust the amount of distortion undergone by the image and the coordinates of points $\mathbf{x}$ and $\mathbf{u}$ are referred to the centre of the radial distortion. Since the effects of parameter ξ depend on the dimension of the input image, we quantify the amount of distortion independently from the image size as the percentage:

$$d_{\%} = 1 - \frac{\hat{r}_M}{r_M} \tag{2}$$

where r_M represents the maximum radius in the undistorted image (i.e., the distance from the image centre to the corner) and $\hat{r}_M$ represents its distorted counterpart. The parameter ξ is related to the percentage of distortion $d_{\%}$ by the following formula:

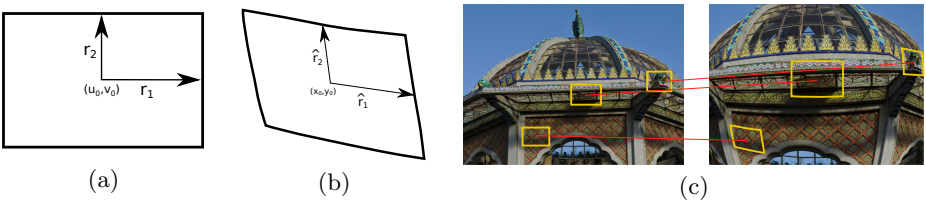
$$\xi = -\frac{d_{\%}}{[r_M(1 - d_{\%})]^2}. \tag{3}$$

As shown through Fig. 1, the main advantage of dealing with the distortion percentage $d_{\%}$ rather than with the distortion parameter ξ is that it gives a perceptually coherent measure of the amount of distortion characterizing an image. Exploiting the local linearity assumption discussed above, a circular neighbourhood of radius r centred at the undistorted point $\mathbf{u}$ can be mapped with a small error to a circular neighbourhood centred at the distorted point $\mathbf{x}$ of radius:

$$\hat{r} = g_{\mathbf{u}}(r) = \frac{2r}{1 + \sqrt{1 - 4\xi r^2}}. \tag{4}$$

2.1 Gradient Estimation of Distorted Images

Since the gradients of an image are very related to the geometry of the acquired scene, employing classic estimation techniques (e.g., Sobel filters) on geometrically distorted images would lead to the estimation of distorted gradients.



(a) (b) (c)

Fig. 2. (a) A rectilinear neighbourhood and (b) its distorted counterpart. (c) Examples of rectilinear neighbourhoods along with their distorted counterparts.

The most common way to compute a geometrically correct gradient is performing a rectification process and applying the classic techniques (e.g., Sobel filters) on the rectified image. Since we would like to be able to work directly in the distorted domain, we consider two more approaches available in the literature. The former consists in correcting the distorted gradients using the Gradient Correction Jacobian (GCJ) matrix computed at the interested distorted locations [6]. The latter estimates the gradients using the Generalized Sobel Filters (GSF), a family of adaptive filters which takes into account the geometry of the distortion [10].

3 Distortion Adaptive Descriptors

In this Section we introduce the Distortion Adaptive Descriptors (DAD). Rather than a new set of descriptors, the DAD constitute a paradigm for correctly computing existing local descriptors directly on the distorted images. For sake of generality we consider a generic descriptor $D(\mathcal{N}, \mathcal{M}(I, \mathcal{N}))$ computed on a rectangular neighbourhood $\mathcal{N}$ using some measurements $\mathcal{M} = \mathcal{M}(I, \mathcal{N})$ performed in the locations of the input image I specified by the neighbourhood $\mathcal{N}$. The measurements can be of any kind and are related to the feature extraction process required by the specific descriptor. In the SIFT descriptor, for instance, the measurements $\mathcal{M}$ are the image gradients estimated at the relevant locations. The rectangular neighbourhood centred at point (u_0, v_0) with radii r_1 and r_2 is naturally defined as the set of points:

$$\mathcal{N}(u_0, v_0, r_1, r_2) = \{(u, v) : |u - u_0| \leq r_1 \wedge |v - v_0| \leq r_2\} \tag{5}$$

Fig. 2 (a) shows an example of rectangular local neighbourhood. When the descriptor has to be computed on a distorted image, the shape of the neighbourhood $\mathcal{N}$ depends on its position in the image. Some examples of such assertion are illustrated in Fig. 2 (c). The rectilinear neighbourhood (5) is easily mapped to its distorted counterpart centred at point (x_0, y_0) with radii $\hat{r}_1$ and $\hat{r}_2$ using the following expression:

$$\begin{aligned}\hat{\mathcal{N}}(x_0, y_0, \hat{r}_1, \hat{r}_2) = \{(x, y) : |f^{-1}(x) - f^{-1}(x_0)| \leq g^{-1}_{x_0, y_0}(\hat{r}_1) \wedge \\ |f^{-1}(y) - f^{-1}(y_0)| \leq g^{-1}_{x_0, y_0}(\hat{r}_2)\}\end{aligned} \tag{6}$$

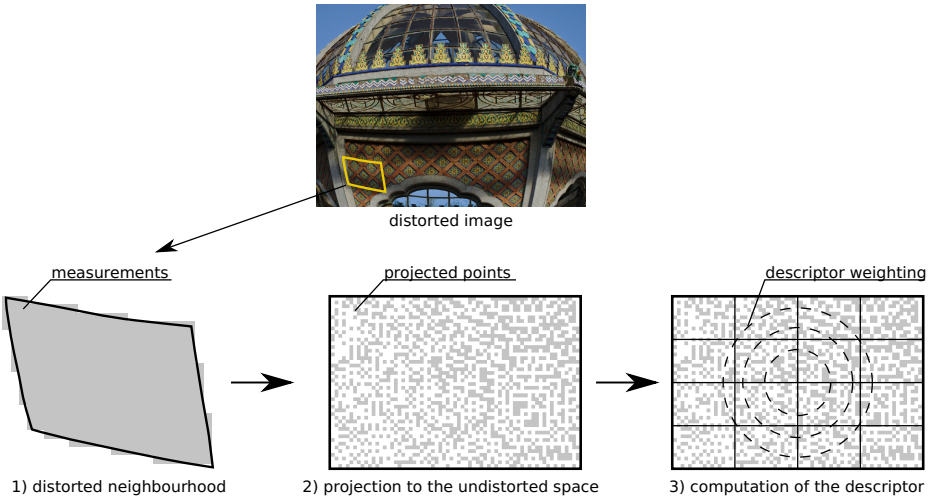


Fig. 3. A scheme of the computation of the Distortion Adaptive Descriptors. 1) The distorted neighbourhood is extracted from the input image. 2) The measurements are projected to the rectilinear space. As it can be noted, this yields to samples of non uniform density. 3) The regular descriptor is computed accounting for the correct arrangement of the measurements in the rectilinear space.

where $(x_0, y_0) = f(u_0, v_0)$ and $\hat{r}_1$ and $\hat{r}_2$ are obtained from r_1 and r_2 using equation (4). Fig. 2 (b) shows an example of distorted neighbourhood. Let be $\hat{\mathcal{M}}(\hat{I}, \hat{\mathcal{N}})$ the geometrically correct measurement performed in the locations of the distorted image $\hat{I}$ specified by the distorted neighbourhood $\hat{\mathcal{N}}$. The Distortion Adaptive Descriptor related to $\mathcal{D}$ is hence defined as:

$$\hat{\mathcal{D}} = D(f^{-1}(\hat{\mathcal{N}}), \hat{\mathcal{M}}(\hat{I}, \hat{\mathcal{N}})) \tag{7}$$

where $f^{-1}(\hat{\mathcal{N}}) = \{f^{-1}(x, y) : (x, y) \in \hat{\mathcal{N}}\}$.

The computation defined above is carried in three key steps: 1) given a point (x_0, y_0) in the distorted space and two radii $\hat{r}_1$, $\hat{r}_2$, the distorted neighbourhood $\hat{\mathcal{N}}(x_0, y_0, \hat{r}_1, \hat{r}_2)$ is considered; 2) all the coordinates of the points in $\hat{\mathcal{N}}$ are projected back to the rectilinear space $(f^{-1}(\hat{\mathcal{N}}))$; 3) the regular descriptor is computed using the geometrically correct measurements $\hat{\mathcal{M}}(\hat{I}, \hat{\mathcal{N}})$ and the projected coordinates $f^{-1}(\hat{\mathcal{N}})$. It should be noted that step 2) is important since it allows the descriptor to weigh the measurements according to their position in the undistorted space. Specifically, the projection leads to samples of non-uniform density which are correctly dislocated in the undistorted circular neighbourhood. The new locations for the considered measurements ensure a correct isotropic spatial weighting. Fig. 3 shows a scheme of the computation of the Distortion Adaptive Descriptors.

Fig. 4. Thumbnails of the images included in the dataset.

4 Experimental Settings

We argue that a combination of the gradient estimation techniques discussed in Section 2.1 and the scheme proposed in Section 3 can improve the matching ability of gradient based local descriptors on distorted images. What we want to evaluate is the invariance of the descriptors with respect to radial distortion, i. e., the ability to produce similar descriptors for two image regions representing the same physical area of the scene despite they are affected by different amounts of distortion. An ideal descriptor, for instance, would give identical results when computed on the matching neighbourhoods shown in Fig. 2 (c). In the following we discuss the experimental settings including the images used for the evaluations, the considered descriptors and the evaluation pipeline.

Wide Angle Images. We work with real-world rectilinear images to which radial distortion is artificially added as done in [1,6,7,10]. Working in this settings is convenient since it allows to control the exact amount of distortion present in the image for evaluation purpose. Moreover, the source rectilinear images can be reliably used to compute the ground truth descriptors using the standard algorithms proposed by their authors. In general, given a rectilinear image I and the percentage of distortion $d_{\%}$, we use the Division Model described in Section 2 to generate the distorted image $\hat{I}(\mathbf{x}) = I(f^{-1}(\mathbf{x}))$. We always consider the centre of distortion as the centre of the distorted image. We use the dataset proposed in [10] which contains 100 high resolution (5204×3472 pixels) images depicting scenes taken from different image categories including indoor, outdoor, natural, handmade, urban, car, pedestrian and street scenes.[1] Fig. 4 shows the thumbnails of the images included in the dataset. In order to obtain distorted images with variable FOV, we map the high resolution images contained in the dataset to distorted images with standard resolution of 1024×768 pixels at variable distortion rates as suggested by [10]. Fig. 1 reports some examples of output images.

Descriptors. We apply the DAD scheme to the SIFT and HOG descriptors using the gradient estimation techniques discussed in Section 2.1 to obtain the measurements. To assess the improvement due to the DAD scheme independently

[1] The dataset is available at the url: http://iplab.dmi.unict.it/icip2015/dataset.zip.

Table 1. The descriptors considered in the experiments.

Acronym	Description
SIFT_{DIST} HOG_{DIST}	Regular SIFT/HOG descriptor computed on the distorted images using the distorted gradients as measurements.
SIFT_{RECT} HOG_{RECT}	Regular SIFT/HOG descriptor computed on the rectified images using the Sobel filters to estimate the gradients.
SIFT_{GCJ} HOG_{GCJ}	Regular SIFT/HOG descriptor computed on the distorted images using the GCJ gradients as measurements.
SIFT_{GSF} HOG_{GSF}	Regular SIFT/HOG descriptor computed on the distorted images using the GSF gradients as measurements.
SIFT_{IDEAL} HOG_{IDEAL}	Regular SIFT/HOG descriptor computed on the distorted images using the ground truth gradients as measurements.
DAD-SIFT_{DIST} DAD-HOG_{DIST}	SIFT/HOG descriptor computed with the DAD scheme on the distorted images using the distorted gradients as measurements.
DAD-SIFT_{GCJ} DAD-HOG_{GCJ}	SIFT/HOG descriptor computed with the DAD scheme on the distorted images using the GCJ gradients as measurements.
DAD-SIFT_{GSF} DAD-HOG_{GSF}	SIFT/HOG descriptor computed with the DAD scheme on the distorted images using the GSF gradients as measurements.
DAD-SIFT_{IDEAL} DAD-HOG_{IDEAL}	SIFT/HOG descriptor computed with the DAD scheme on distorted images using the ground truth gradients as measurements.

from the employed gradient estimation technique, we also consider an ideal estimator by warping the ground truth gradients to the distorted locations. Moreover, we consider the standard SIFT and HOG descriptors computed directly in the distorted domain (without adaptation) combined with the different gradient estimation techniques. Hence we derive the 18 descriptors summarized in Table 1. The SIFT-based descriptors are computed using the implementation provided by the VLFeat library [3], which produces standard 128-dimensional descriptors. For the HOG-based descriptors we consider the variant of HOG proposed in [23] as implemented by the VLFeat library [3]. Moreover, in our settings, the HOG-based descriptors are computed dividing the support region into 4×4 cells and the gradients are computed using 3×3 filters (in place of the non-smoothing $[-1\ 0\ 1]$ and $[-1\ 0\ 1]^T$ filters originally proposed by the authors [11]) in order to allow the gradient estimation techniques to compensate the distortion exploiting neighbourhood information. This configuration returns a 496-dimensional HOG descriptor for input support region of any size.

Evaluation Pipeline. For our evaluations, we measure the matching ability of the considered descriptors when they are densely extracted from the test images. Dense descriptors are appropriate for our analysis since they allow us to draw

conclusions which are independent from any interest point detector. Moreover dense descriptors have proven powerful in a variety of tasks [19–21]. Given the reference-distorted image pair $(I, \hat{I})$, we densely extract square support regions from the reference image at a regular step of 50 pixels. To account for multiscale features, different layers of overlapping support regions are extracted considering radii ranging from 32 to 256 pixels. In this context, a support region is an entity $\mathcal{S}(\mathbf{u}, r)$ made of two elements: a centre $\mathbf{u}$ and a radius r. Each support region $\mathcal{S}$ is mapped to the corresponding support region $\hat{\mathcal{S}}$ in the distorted image using equations (1) and (4) reported in Section 2: $\hat{\mathcal{S}}(f(\mathbf{u}), g(r))$. All projected support regions which are not entirely contained in the distorted image $\hat{I}$ or which projected radius is under 16 pixels are discarded together with their undistorted counterparts. This settings lead to support regions of variable sizes ranging from 32×32 pixels to 512×512 pixels which cover the entire FOV of the distorted images. The number of support regions per image ranges from 887 to 3881 depending on the distortion rate. We refer to the set of reference support regions as $S = \{\mathcal{S}_i\}$ and to the set of projected support regions as $\hat{S} = \{\hat{\mathcal{S}}_i\}$. The reference support regions S are used to compute the standard SIFT and HOG descriptors, while the projected support regions $\hat{S}$ are used to compute the descriptors under evaluation. For instance, let be $\hat{\mathcal{D}}$ one of the SIFT-based descriptors in Table 1, we define the set of reference descriptors as $D = SIFT(S)$ and the set of test descriptors $\hat{D} = \hat{\mathcal{D}}(\hat{S})$. Similar definitions hold for the HOG-based descriptors. To evaluate the matching ability of descriptor $\hat{\mathcal{D}}$, we follow the scheme proposed in [18] to compute 1-precision vs recall curves. According to the authors of [18] two support regions $\mathcal{S}$ and $\hat{\mathcal{S}}$ match if the distance between their descriptors is below a threshold t. Each descriptor from the reference image is compared to each descriptor from the distorted one and the numbers of correct and false matches are counted. The threshold t is varied to obtain the curves. A match between two descriptors is considered correct only if they have been computed on corresponding support regions. For each threshold t, the precision and recall values are computed using the expressions:

$$Precision = \frac{\#correct\ matches}{\#matches} \tag{8}$$

$$Recall = \frac{\#correct\ matches}{\#suppport\ regions}. \tag{9}$$

The 1-precision vs recall curves have a straightforward interpretation: a perfect descriptor would give a recall equal to 1 for any precision. In practice increasing the value of threshold t increases the recall and decreases the precision. The rate at which those values vary with respect to the threshold tells how an algorithm is able to produce distinctive descriptors, which are similar for corresponding regions. As reported in [18], this kind of evaluation is independent from the matching scheme one could adopt (e.g., nearest neighbour with or without rejection of ambiguous matches) and respect the distribution of the descriptors in the space. To complement our analysis we also report threshold vs F-Measure curves. The F-Measure values are computed as reported in [22]:

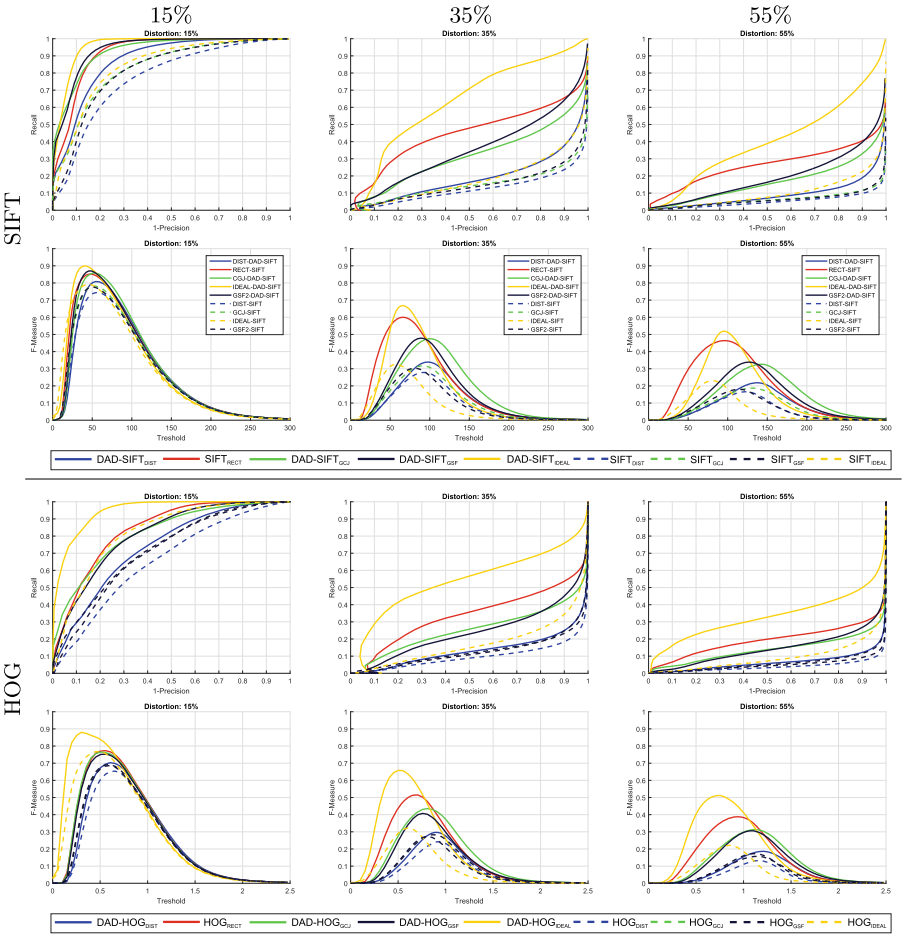


Fig. 5. The 1-precision vs recall curves (rows 1 and 3) and the threshold vs F-Measure curves (rows 2 and 4) for the SIFT-based and the HOG-based descriptors.

$$F_\beta = \frac{(1+\beta^2)Precision \times Recall}{\beta^2 \times Precision + Recall} \tag{10}$$

where $\beta^2 = 0.3$ to weigh precision more than recall. The threshold vs F-Measures curves can be interpreted from a retrieval point of view: a good descriptor allows to get a high number of positives with a small amount of noise. This situation is represented by a F-Measure curve with a high peak for a low threshold value.

5 Results

Fig. 5 shows the 1-precision vs recall and the threshold vs F-Measure curves of the considered descriptors for different amounts of distortion. As it can be noted,

all the DAD variants systematically outperform their non-adaptive counterparts independently from the employed gradient estimation technique. Moreover, using the GCJ and GSF techniques for the measurements allows to improve the performances of all the descriptors over the distorted gradients. Interestingly such techniques, combined with the DAD paradigm, allow to reach the performances obtained through the rectification process for low amounts of distortion (15%). In general, however, the rectification provides better results for higher distortion rates at the cost of the computational time required by the unwarping. The GCJ and GSF techniques have similar performances when used both in the HOG-based and SIFT-based descriptors. The descriptors based on the ground truth gradients always have the best performances, which confirms the power of the DAD paradigm. Moreover, the gap between the performances given by the ground truth gradients and the ones given by the considered gradient estimation techniques suggests that there is still space for improvement for such techniques.

6 Conclusion

We have tackled the problem of improving the matching ability of gradient-based descriptors when they are directly computed on wide angle images. We have proposed the Distortion Adaptive Descriptors, a new paradigm for the correct computation of local descriptors in the distorted domain. Even if the proposed descriptors can be computed directly on the wide angle images, the performances obtained through the rectification process are not matched yet. The results convey that improving the gradient estimation techniques would allow to significantly improve the performances of gradient-based local descriptors on wide angle images. Further works will be devoted to quantitatively assessing the improvement in computational time of the DAD over the rectification-based techniques as well as compare the performances of the DAD against existing solutions (e.g., [6,8]).

This work has been performed in the project PANORAMA, co-funded by grants from Belgium, Italy, France, the Netherlands, the United Kingdom, and the ENIAC Joint Undertaking.

References

1. Furnari, A., Farinella, G.M., Puglisi, G., Bruna, A.R., Battiato, S.: Affine Region Detectors on the Fisheye Domain. In: IEEE International Conference on Image Processing (2014)
2. Miyamoto, K.: Fish eye lens. Journal of the Optical Society of America, 2–3 (1964)
3. Vedaldi, A., Fulkerson, B.: VLFeat: An open and portable library of computer vision algorithms. In: Proceedings of the international conference on Multimedia, pp. 1469–1472, (2010)
4. Puig, L., Guerrero, J.J.: Omnidirectional Vision Systems. Springer (2013)
5. Baker, S., Nayar, S.K.: A theory of catadioptric image formation. In: International Conference on Computer Vision, pp. 35–42 (1998)

6. Lourenço, M., Barreto, J.P., Vasconcelos, F.: sRD-SIFT: keypoint detection and matching in images with radial distortion. IEEE Transactions on Robotics **28**(3), 752–760 (2012)
7. Cinaroglu, I., Bastanlar, Y.: A direct approach for human detection with catadioptric omnidirectional cameras. In: Signal Processing and Communications Applications Conference, pp. 2275–2279 (2014)
8. Cruz-Mota, J., Bogdanova, I., Paquier, B., Bierlaire, M., Thiran, J.: Scale invariant feature transform on the sphere: theory and applications. International Journal of Computer Vision **98**(2), 217–241 (2011)
9. Lowe, D.G.: Distinctive Image Features from Scale-Invariant Keypoints. International Journal of Computer Vision 60, 91–110
10. Furnari, A., Farinella, G.M., Bruna, A.R., Battiato, S.: Generalized Sobel filters for gradient estimation of distorted images. In: The International Conference on Image Processing (submitted 2015)
11. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. Computer Vision and Pattern Recognition **1**, 886–893 (2005)
12. Hughes, C., Glavin, M., Jones, E., Denny, P.: Wideangle camera technology for automotive applications: a review. IET Intelligent Transport Systems **3**(1), 19–31 (2009)
13. Battiato, S., Farinella, G.M., Furnari, A., Puglisi, G., Snijders, A., Spiekstra, J.: A Customized System for Vehicle Tracking and Classification, Expert Systems With Applications (2015)
14. Battiato, S., Gallo, G., Puglisi, G., Scellato, S.: SIFT features tracking for video stabilization. In: International Conference on Image Analysis and Processing, pp. 825–830 (2007)
15. Dorko, G., Schmid, C.: Selection of Scale-Invariant Parts for Object Class Recognition. In: International Conference on Computer Vision, pp. 634–640 (2003)
16. Brown M., Lowe, D.: Recognising Panoramas. In: International Conference on Computer Vision, pp. 1218–1227 (2003)
17. Fitzgibbon, A.W.: Simultaneous linear estimation of multiple view geometry and lens distortion. In: Computer Vision and Pattern Recognition, vol. 1 (2001)
18. Mikolajczyk, K., Schmid, C.: Performance evaluation of local descriptors. Pattern Analysis and Machine Intelligence **27**(10), 1615–30 (2005)
19. Liu, C., Yuen, J., Torralba, A.: SIFT flow: dense correspondence across scenes and its applications. Pattern Analysis and Machine Intelligence 33(5) (2011)
20. Farinella, G.M., Allegra, D., Stanco, F.: A benchmark dataset to study the representations of food images. In: Assistive Computer Vision and Robotics in conjunction with the European Conference on Computer Vision (2011)
21. Lazebnik, S., Schmid, C., Ponce, J.: Beyond bags of features: spatial pyramid matching for recognizing natural scene categories. In: Computer Vision and Pattern Recognition, vol. 2 (2006)
22. Achanta, R., Hemami, S., Estrada, F., Susstrunk, S.: Frequency-tuned salient region detection. In: Computer Vision and Pattern Recognition, pp. 1597–1604 (2009)
23. Felzenszwalb, P.F., Grishick, R.B., McAllester, D., Ramanan, D.: Object detection with discriminatively trained part based models. Pattern Analysis and Machine Intelligence (2009)