ADDING COLOR TO GEODESIC INVARIANT FEATURES

Pier Paolo Campari

Politecnico di Milano, Milan, Italy

pc666272@stud.polimi.it

Matteo Matteucci, Davide Migliore

Politecnico di Milano, Milan, Italy

matteucci@elet.polimi.it, migliore@elet.polimi.it

Keywords: Feature Extraction, Feature Description, Color Pattern Recognition.

Abstract: Geodesic invariant feature have been originally proposed to build a new local feature descriptor invariant not only to affine transformations, but also to general deformations. The aim of this paper is to investigate the possible improvements given by the use of color information in this kind of descriptor. We introduced color information both in geodesic feature construction and description. At feature construction level, we extended the fast marching algorithm to use color information; at description level, we tested several color spaces on real data and we devised the opponent color space as an useful integration to intensity information. The experiments used to validate our theory are based on publicly available data and show the improvement, in precision and recall, with respect to the original intensity based geodesic features. We also compared this kind of features, on affine and non affine transformation, with SIFT, steerable filters, moments invariants, spin images and GIH.

1 INTRODUCTION

In this paper we face the issue of feature description at the base of automatic correspondence matching between images from different views of the same scene or images of the same objects in different poses. In particular we are interested in exploiting color information to enrich feature descriptors invariant with respect to generic transformation/deformations.

In (Ling and Jacobs, 2005), Ling and Jacobs introduced such kind of descriptor, together with the geodesic framework, for deformation invariant features extraction and matching. To cope with generic image deformations, this kind of descriptor is based on a histogram built out of pixel intensities from regions at the same geodesic distance with respect to a given interest point. Geodesic invariant descriptors have been experimentally proved to be covariant with respect to generic deformations, but have little descriptive capabilities with respect to other classical feature descriptors such as SIFT. This weakness is mainly due to the smoothing effect induced by the histogram and the sensitivity to image gradient in geodesic distance calculation.

The aim of our work is to investigate how color information can be used to improve geodesic image descriptors by reducing this sensitivity to the image gradient and extending histogram description. In the original paper Ling and Jacobs (Ling and Jacobs, 2005) use the fast marching algorithm (Sethian, 1999) to expand geodesic borders based on intensity values; in this paper we propose to use the RGB components to consider edges in color components instead of intensity. Color information is also used to complement intensity histogram with an appropriate color description based on opponent color space (van de Weijer and Schmid, 2006).

In the next section we briefly review related works in feature description and evaluation, Section 3 introduces the geodesic invariant framework and the novelties introduced by our work. Section 4 presents the experimental setup and results to assess the improvement given by the use of color information, while a brief discussion about future work is presented in the final section.

2 RELATED WORKS

A considerable amount of work has been done on robust local feature descriptors, studying invariance to orientation, scale, affine transformation, and, recently, also to generic deformation.

In the work of Mikolajczyck and Schmid (Mikolajczyk and Schmid, 2005; Mikolajczyk et al., 2005) a performance evaluation of several of these local descriptors is performed both with respect to viewpoints and lighting conditions. As final result, this work reports that the SIFT descriptor, proposed by Lowe (Lowe, 2004), has the best performance with images of flat scenes and affine transformations. These conclusions have been supported also in the paper by Moreels and Perona (Moreels and Perona, 2005), which generalizes these results to 3D scenes using images of 3D objects viewed under different scales, viewpoints, and lighting conditions.

All the local invariant descriptors, investigated so far, are based on the hypothesis of perspective deformation being properly approximated, locally, by affine transformation; recently, Ling and Jacobs (Ling and Jacobs, 2005) demonstrated that it is possible also to construct descriptors invariant to generic deformation of image subject (e.g., a moving flag). In their proposal, they suggest to treat the intensity image as a surface embedded in 3D space, with the third coordinate being proportional to the intensity values, and then build the descriptor by deformation invariance geodesic distance in this 3D space.

Stimulated by their work, in this paper we want demonstrate that it is possible to improve the performance of the Geodesic Intensity Histogram (GIH) descriptor by introducing color information. Although color seems to be a fundamental clue for object recognition in everyday life only few color invariant descriptors have been proposed in the literature. The work of Van De Weijer and Schmid (van de Weijer and Schmid, 2006) is an important example from this point of view. Their results lead to the encouraging conclusions that a pure color-based approach outperforms a shape-based approach only for colorful objects, while, for the general case, it is possible any way to outperforms a pure shape-based approach using a combination of shape and color.

In our work we verify the results achieved by Van De Weijer and Schmid also for non affine transformations with the idea that shape-based descriptors can fail when dealing with generic deformations and combining it with color can improve recognition rates..

3 COLORING GEODESIC INVARIANT FEATURES

In the geodesic framework, an image can be interpreted as a 2D surface in a 3D space, with the third coordinate being proportional to the pixels intensity value, with an aspect weight $\alpha \to 1$, and the first two coordinates proportional to (x, y) (image pixel coordinates) with weight $1 - \alpha$.

We define a geodesic level curve as the set of points at the same geodesic distance from a given interest point; it is possible to capture the joint distribution of intensity and geodesic distances and summarize it into the so called GIH histogram-based descriptor by sampling pixels with constant geodesic step Δ..

An efficient scheme for the geodesic level curves computation on discrete pixel grids, was provided by Sethian with the name of *fast marching algorithm* (Sethian, 1999). A marching speed $F(x, y)$ is associated to each pixel x, y and the geodesic distance $T(x, y)$ can be estimated solving locally the equation $|\nabla T| F = 1$, where

$$F(x,y) = \frac{1}{f(x,y)} = \frac{1}{\sqrt{(1-\alpha)^2 + \alpha^2 I_x^2 + \alpha^2 I_y^2}}. \quad (1)$$

Although the shape of the resulting region is irregular, it is covariant with deformation and it has shown interesting results for generic continuous deformations. Sometimes worst behaviors may occur in correlation with the presence of isotropic and anisotropic scale transformations, causing a resampling of pattern contours, but for uniform intensity region, the expansion is independent from image gradient, mainly depending on $1 - \alpha$ value.

3.1 Fast Marching Algorithm in RGB Space

The first improvement proposed in this paper aims at modifying region expansion in the fast marching algorithm considering color information[1]. We take into account each RGB channel separately, computing 3 different inverse marching speeds one for each channel:

$$f_r(x,y)^2 = (1-\alpha)^2 + \alpha^2 R_x^2 + \alpha^2 R_y^2, \quad (2)$$

$$f_g(x,y)^2 = (1-\alpha)^2 + \alpha^2 G_x^2 + \alpha^2 G_y^2, \quad (3)$$

$$f_b(x,y)^2 = (1-\alpha)^2 + \alpha^2 B_x^2 + \alpha^2 B_y^2. \quad (4)$$

---

[1]We present in this paper only the RGB implementation for the new fast marching algorithm; we tested also other color spaces with no meaningful improvements thus we decided to use the most efficient one.

Assuming there is no particularly strong photometric transformation (e.g. Lambertian surface), these functions are deformation covariant, as with the geodesic distance on the intensity surface, and, under the same hypothesis, the ordering of $f_r(x,y)$, $f_g(x,y)$ and $f_b(x,y)$ is preserved as well. The geodesic region, in our proposal, is thus computed starting from a new inverse marching speed based on these three color channels:

$$f_{min}(x,y) = \min(f_r(x,y), f_g(x,y), f_b(x,y)), \qquad (5)$$

and the resulting new geodesic distance T is invariant, being the sum of the same invariant stretches. This choice, being somehow counterintuitive, has two rationalities behind it and has been confirmed by experimental results:

- The growth of geodesic region is bounded by all channels and slows down only for high values of $f_{min}(x,y)$. The moving front speed decreases only if contour are thus found on all channels, corresponding to a strong change in luminance or crominance, less sensitive to noise with respect to geometric transformations

- Weak contours found on a single channel have minor influence on the geodesic region and let the front go on. In this way, given a geodesic level corresponding to a quantization interval on the geodesic distance, this will show the salient intensity and color varieties separated by weak contours.

Table 1 shows an example of geodesic distance T expansion on a monodimensional grid of pixel $i \in \{1..n\}$. Pixel updating order is deterministic, in fact the moving front can proceed in only one direction. The updating formula becomes:

$$T(i+1) = T(i) + f_{min}(i+1) \qquad (6)$$

We can identify weak contour on pixels $i = 2$ and $i = 3$ and a strong contour on pixel $i = 6$. Computing geodesic distance in this fashion, it puts in the same interval of T more color varieties, represented by weak contours. For instance, let be the sampling gap $\Delta = 0.1$ and the histogram quantization interval of T to $\delta T = 0.8$. The geodesic level segment (ring in the bidimensional space) starts at the first pixel and stops between pixel $i = 5$ and $i = 6$. Sampling points coming from this segment shows a lot of color variations, the same variations causing the weak contours on pixels $i = 2$ and $i = 3$.

The results obtained by using Equation (5) is twofolds, on one hand we have that strong contours are less likely to be smoothed, if affected by geometric deformations, and, on the other one, a more

Table 1: Evolution of geodesic distance T on a monodimensional grid of pixel $i \in \{1..n\}$., computed using f_r, f_g, f_b.

i	1	2	3	4	5	6	7
f_r	0.1	0.5	0.1	0.1	0.2	0.9	0.1
f_g	0.1	0.1	0.1	0.1	0.1	0.8	0.1
f_b	0.1	0.1	0.4	0.1	0.2	0.8	0.2
T	0.1	0.2	0.3	0.4	0.5	1.3	1.4

deformation-covariant placement of sampling regions is obtained.

3.2 Building the Geodesic Color Descriptor

Low distinctiveness is a disadvantage of the original geodesic descriptor. The GIH descriptor summarizes the geodesic region content as an histogram $H_p(k,m)$ built on two variables: geodesic distance g and normalized intensity I (g is quantized in M intervals, I in K intervals). For each geodesic interval m, a normalization has to be performed such as $\sum_{k=1}^{K} H_p(k,m) = 1$ to compare inner and outer geodesic "rings" with the same weight.

Given a ring of geodesic level curves, corresponding to a quantization interval on the geodesic distance, GIH stores no spatial informations about sample location along the ring thus the resulting numeric vector is less informative than other descriptors such as SIFT (on affine covariant trasformation). We propose to partially overcome this distinctiveness problem, without losing rotation invariance, by adding color information into the descriptor.

To build a richer descriptor, we consider 3 or more dimensions, together with their respective quantization intervals: the geodesic distance g (quantized in M bins), the normalized intensity I (quantized in K bins), and one or more photometric invariants Inv_n (quantized in Q_n bins).

The implemented descriptor keeps the geodesic information associated to color invariants, extending the original $M \times K$ GIH descriptor with N additional $M \times Q_n$ histograms one per invariant Inv_i. The final dimension of the descriptor is thus $M(K + \sum_{n=1}^{N} Q_i)$.

The matching distance between two *Geodesic Histogram & Histograms* descriptor (GHH hereafter) p and q is computed starting from the χ^2 distances between each couple of histograms:

$$d_I = \chi^2(H_p^I, H_q^I) \qquad (7)$$

$$d_{Inv_1} = k_1 \cdot \chi^2(H_p^{Inv_1}, H_q^{Inv_1}) \qquad (8)$$

$$d_{Inv_2} = k_2 \cdot \chi^2(H_p^{Inv_2}, H_q^{Inv_2}) \qquad (9)$$

$$\dots \qquad (10)$$

where k_1 and k_2 are weighting parameters, and than the descriptors distance as a whole is obtained as

$$d(H_p, H_q) = \max(d_I, d_{Inv_1}, d_{Inv_2}, \dots). \qquad (11)$$

This implies that a match between two point is allowed only if all pairs of histograms are fairly similar, up to a given threshold, both in luminance and chrominance.

3.3 Color Invariants Selection

GHH descriptor can be seen as a general construction method for geodesic descriptors including intensity and color information; effective invariant selection, then, should be performed taking into account specific assumptions about image formation and the specific application.

We can distinguish between *zero-order invariants* and *first-order invariants*, i.e., derivative based invariants. For the use with geodesic regions, characterized by generic deformations, the main challenge is to preserve geometric robustness, then we can have focused the work on zero-order invariants computable from common color spaces. Another practical requirement is stability, since known photometric invariant have inherent instabilities.

The measured color information for a camera sensor can be modeled as:

$$C(\vec{x}) = m_b(\vec{x})b_C(\vec{x})e_C + m_i(\vec{x})e_C, \qquad (12)$$

where the first part describes the light which is reflected after interaction with surface albedo b and the second-one is related to the light immediately reflected at the surface, causing specularities. The aim of color space selection is to provide invariance with respect to m_b and m_i, which depend on scene and illumination geometry. In this paper we considered three invariant for color representation:

RGB normalized *RGB* are invariant to the shading modeled by m_b term in absence of specular effects (i.e. $m_i = 0$);

O1,O2 opponent colors, that, for white illuminant, are invariant with respect to specularities;

H,S the hue information is invariant to both m_b and m_i, but this is known to be unstable with low saturation also; weighting each sample by its saturation is only a partial solution.

We have tested these invariants on the real dataset provided in the paper of Mikolajczyck and Schmid (Mikolajczyk and Schmid, 2005) with GHH descriptor, after normalization obtained subtracting
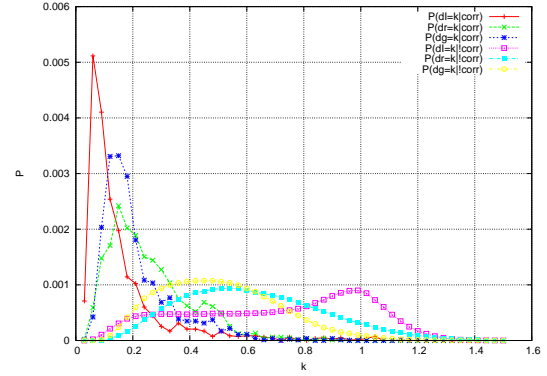

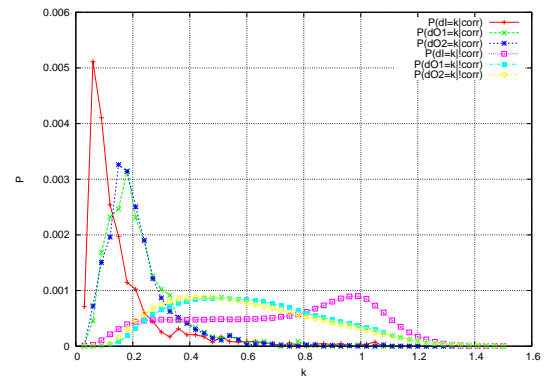Figure 1: Histograms distance distributions for *I*, *r* and *g*.

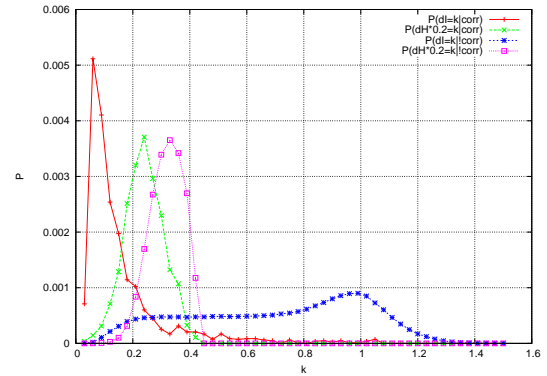Figure 2: Histograms distance distributions for *I*, *O*1 and *O*2.

Figure 3: Histograms distance distributions for *I* and *Hue* (scaled by 0.2 for better illustration).

the mean and dividing by standard deviation. For angular invariant hue, illuminant normalization is previously performed on RGB channels, dividing each channel value by its spatial average. Values within the $(-3\sigma, +3\sigma)$ interval have been segmented in 13 bins and Figures 1, 2 and 3 compare the intensity histograms H^I respectively with normalized r and g histograms H^r H^g, opponent color histograms H^{O1} H^{O2} and hue histograms H^{hue}.

(a) Graffiti image (graffiti 3)

(b) Flag image
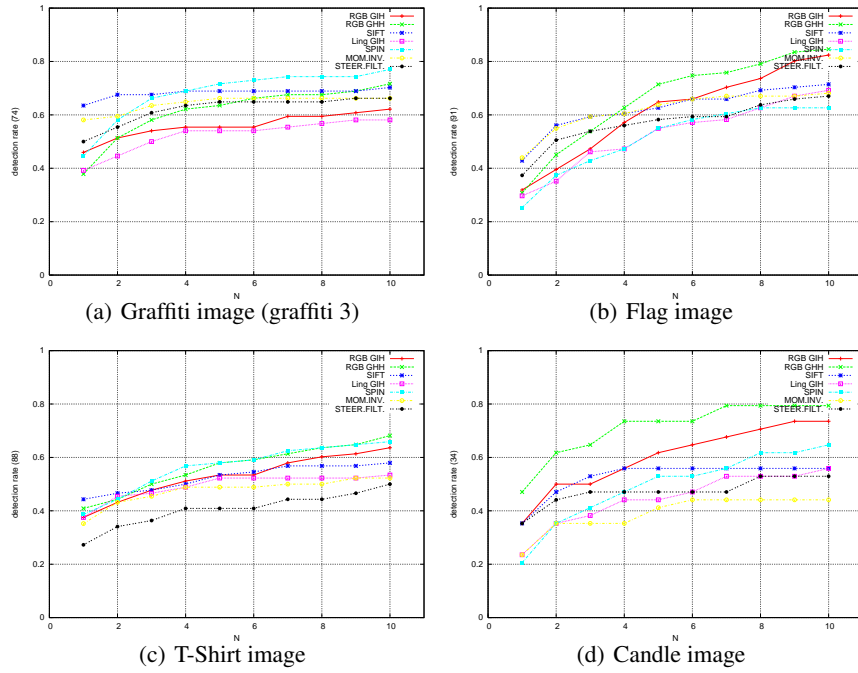
(c) T-Shirt image

(d) Candle image

Figure 4: Performace evaluation on real datasets.

These plots underline the relationship between histograms distance χ^2 and matching correctness in terms of conditional probabilities. Let be (a,b) a couple of features; we estimate the probabilities that the histograms distance $d(a,b)$ falls into a small interval $k-\delta, k$, given that (a,b) is a correspondence

$$P(k-\delta < d(a,b) \leq k | T(a) = b) \qquad (13)$$

and given that (a,b) is not a correspondence

$$P(k-\delta < d(a,b) \leq k | T(a) \neq b) \qquad (14)$$

being T the transformation between the images. In order to make a distinctiveness comparison for intensity and color histograms, the weighting parameters k_i are set to 1.

From the plots it can be noticed that corresponding and non-corresponding features distances for intensity show a higher discrimination power with respect to color information. Nevertheless, opponent colors show a more symmetric and regular behavior (compared to normalized RGB), allowing the assumption $k_{O1} = k_{O2} = k_O$. Mostly due to saturation weight, the hue histograms have low distinctiveness (overlapped distributions). For this reasons, we decided to use opponent color in the following of the paper as color invariants; moreover this analysis suggests a value for the color histograms weighting parameter, that should be adjusted according to the real working conditions. In the case of this dataset $k_O = 0.3$ can be used; this value roughly overlaps correct matches distributions

of d_I, d_{O1} and d_{O2}, allowing histograms comparison with respect to the same threshold d_{th}, and thus improving matching precision.

4 EXPERIMENTAL RESULTS

In this section we present the experimental evaluation done to test the capabilities of the new descriptors proposed in this paper, using as interest point detector the Harris-Affine[2] proposed by Mikolajczyck and Schmid (Mikolajczyk and Schmid, 2004). In particular we compare the performances of our approaches with the descriptors: GIH, GIH with RGB expansion, SIFT (Lowe, 2004), spin images (Lazebnik et al., 2003), moments invariants (Gool et al., 1996), and steerable filters (Freeman and Adelson, 1991).

The characteristics we are interested in evaluating are robustness, i.e., the capability to describe in the same manner two correspondent regions, and precision, i.e., the distinctiveness of the descriptors. We use the criterion proposed by Ling and Jacobs to evaluate the GIH performances. For each pair of images we select the 200 interest points with the higher cornerness and, for each of them, we manually estimate the ground truth matching. Each interest point in the first image is compared with all interest points in

---

[2]Code available at
http://www.robots.ox.ac.uk/vgg/research/affine/

(a) T-shirt images

(b) Candle images

Figure 5: Examples of matches on real images.

the second image according to the distance described in Section 3.2 and we considered the detection rate among the top N matches used to study the performance

$$r = \frac{\#correct\_matches}{\#points\_first\_image}. \quad (15)$$

Experiments have been done on the dataset provided by Ling and Jacobs, on the dataset by Mikolajczyck and Schmid, and on other real images with non affine deformation. Due to space limits we present here only two plots with performances on public available data and we details the results on the data we collected that are more significative for non affine trasformation.

The plots in Figure 4 show that GHH description capability improves the GIH intensity-based description in all the cases and that also the GIH with RGB expansion achieve better performances with respect to classical GIH even if the descriptor is still based only on intensity information. GHH descriptors outperform most of other descriptors when non-affine deformation is present and are able to describe local features also when these are not on a planar surface.

In Figure 5 we report the new (with respect of public available datasets) images used in this paper and the matches found by the use of GHH. To bet-

ter understand the rationale behind GHH we plotted geodesic levels and descriptor histograms for a few interest points on the T-shirt images in Figure 6.

When dealing with affine transformation (See plot in Figure 4(a)) SIFT are still better. We investigated this and devised two possible reasons: the SIFT descriptor is more rich than a histogram based one, and the affine property is not exploited in the geodesic distance calculation. We have coped with the latter issue, but this is out of the scope of this paper, while the former issue is still an ongoing work.

In Figure 4(a) it is also possible to notice that sometimes the SPIN descriptors are better than SIFT, GIH and GHH, this is not a surprising result being the latters based on the "SPIN image" idea: a two-dimensional histogram encoding of the brightness distribution in an affine-normalized patch.

5 DISCUSSION AND FUTURE WORKS

In this paper we introduced a method to extend geodesic invariant feature with color information. Experimental results confirm the promising perfor-
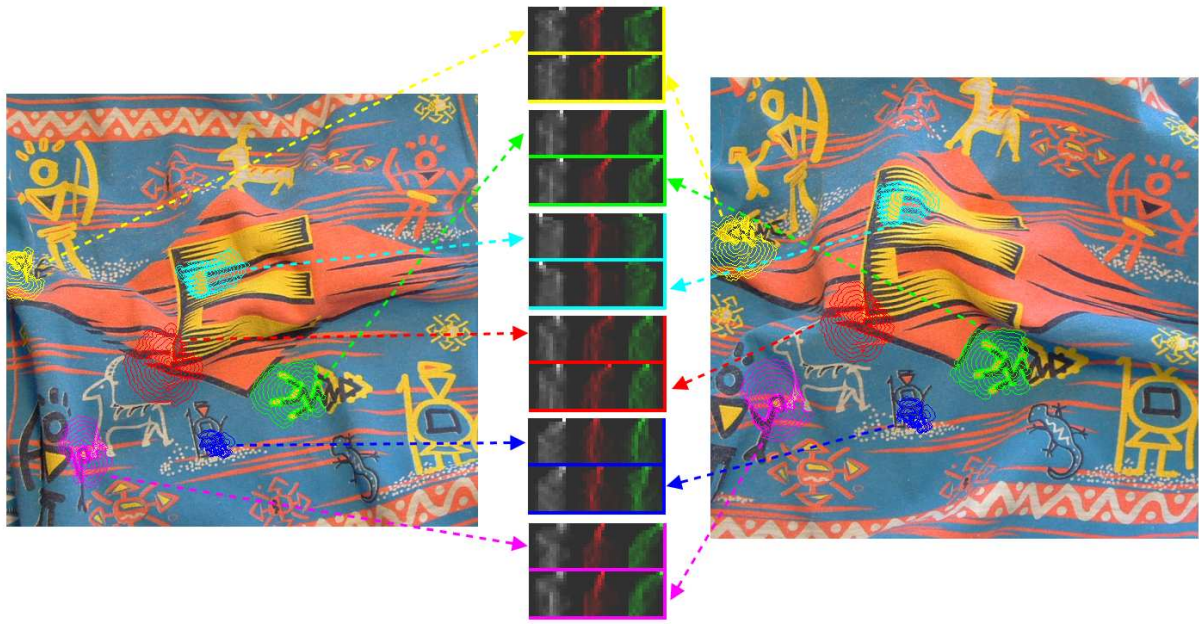
Figure 6: Examples of GHH descriptors on the T-shirt image.

mances of geodesic descriptors and the improvement given by the use of color information.

Althought we succeeded in adding color information to the geodesic framework some open issues still remain: the improvement obtained by GHH on deformed images is paid with a reduced performance with respect to SIFT descriptors on real planar patches. To better understand the reasons for this, we made a deep analysis of geodesic expansion. It turns out that adding the affine information carried by the Harris-affine detector to the expansion mechanism improves the GHH performance on planar patches, but this results are not reported here since somehow reduce the generality of GHH and GHI descriptors. Another reason for SIFT better performance is due to the smoothing effect induced by the histogram representation; we are currently working on SIFT like representation of geodesic levels able to capture local gradient information.

A final note should be made about interest point detectors. In this paper we used the Harris-affine detector, but intuitively it is not optimal for geodesic distance computation since it selects corner points located on singular positions of the image geodesic representation. We have done preliminary work with MSER detectors (Matas et al., 2002), but this needs further investigation since they are located on the baricenter of uniform regions and with non affine deformation this might not be in a repeatable positions. This is a problem since in performance evaluation we do not consider area matching, but points, and this can fail even if we perform the correct match.

ACKNOWLEDGEMENTS

REFERENCES

Freeman, W. and Adelson, E. (1991). The design and use of steerable filters. *IEEE Trans. on PAMI*, 13(9):891–906.

Gool, L. V., Moons, T., and Ungureanu, D. (1996). Affine photometric invariants for planar intensity patterns. *ECCV*, 642-651:1996.

Lazebnik, S., Schmid, C., and Ponce, J. (2003). A sparse texture representation using affine-invariant regions. *CVPR*, 2:319–324.

Ling, H. and Jacobs, D. W. (2005). Deformation invariant image matching. *ICCV*, 1:1466–1473.

Lowe, D. (2004). Distinctive image features from scale invariant keypoints. *IJCV*, 60(2):91–110.

Matas, J., Chum, O., Urban, M., and Pajdla, T. (2002). Robust wide baseline stereo from maximally stable extremal regions. *BMVC*, pages 384–393.

Mikolajczyk, K. and Schmid, C. (2004). Scale and affine invariant interest point detectors. *IJCV*, 60(1):63–86.

Mikolajczyk, K. and Schmid, C. (2005). A performance evaluation of local descriptors. *PAMI*, 27:1615 – 1630.

Mikolajczyk, K., Tuytelaars, T., Schmid, C., Zisserman, A., Matas, J., Schaffalitzky, F., Kadir, T., and Gool, L. V. (2005). A comparison of affine region detectors. *IJCV*, 65(1/2):43–72.

Moreels, P. and Perona, P. (2005). Evaluation of features detectors and descriptors based on 3d objects. *ICCV*, 1:800–807.

Sethian, J. (1999). *Efficient Schemes: Fast Marching Methods*, chapter 8, pages 87–100. Cambridgr University Press.

van de Weijer, J. and Schmid, C. (2006). Coloring local feature extraction. *ECCV*, 2:334–348.