

AN EFFECTIVE LOCAL INVARIANT DESCRIPTOR COMBINING LUMINANCE AND COLOR INFORMATION

Dong Zhang¹, Weiqiang Wang¹, Wen Gao^{1,2}, Shuqiang Jiang²

¹Graduate School of Chinese Academy of Sciences, Beijing, China

²Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China

Email: {dzhang, wqwang, wgao, sqjiang}@jdl.ac.cn

ABSTRACT

Extraction of stable local invariant features is very important in many computer vision applications, such as image matching, object recognition and image retrieval. Most existing local invariant features mainly characterize luminance information, and neglect color information. In this paper, we present a new local invariant descriptor characterizing both of them, which combines three photometric invariant color descriptors with the famous *SIFT* descriptor. To reduce the dimension of the combined high-dimensional invariant feature the principal component analysis (PCA) is used. Our experiments show the proposed local descriptor through combining luminance and color information outperforms the descriptors that only utilize a single category of information, and combining the three color feature representations is more effective than only one.

1. INTRODUCTION

In recent years, a considerable amount of research on local invariant features has shown they are robust to occlusion, background clutter, as well as other content changes, such as orientations, viewpoints and scales. The researchers [1, 2, 3, 4, 5] have successfully demonstrated the strength of local invariant features in many applications, including wide baseline matching, object recognition, image retrieval, etc. The evaluation experiments performed by Mikolajczyk and Schmid [6] show that the ranking of accuracy for the different algorithms is relatively insensitive to the methods employed to locate interest points in images, and depends much on the representation used to model image patches around interest points. Although some effective local descriptors have been presented, most of them use only luminance information and ignore color information.

The research on invariant color features can be found in [7, 8, 9, 10, 11, 12]. Gevers et al. [7] used color invariant gradient information to create a new snake which could greatly diminish the influences of surface orientation, illumination, shadows and highlights. Gouet and Deriche [8]

characterized points of interest by using color differential invariants and generalized the corner detector for gray level to the case of color images. Gevers and Smeulders [9] proposed several color descriptors invariant with respect to specularity, illuminance and lighting geometry. Geusebroek et al. [10] summarized illumination and geometrical invariant properties from the reflectance model based on the Kubelka-Munk theory, and evaluated the discriminative power of several color invariants. Gevers and Smeulders [11] reported a method of image retrieval by combining color and luminance invariant features. But their method was not based on local regions. Recently, Weijer and Schmid [12] proposed a set of color descriptors robust to photometric change and different image qualities, and discussed the combination of local color and luminance information.

In this paper, we present a new local descriptor combining three modified photometric invariant color descriptors with the famous *SIFT* descriptor. To reduce the high dimension of the proposed hybrid feature, the PCA is applied.

The paper is organized as follows. In section 2, we present a new local invariant descriptor combining luminance and color information. Section 3 gives comprehensive experimental results of comparing our image descriptor with some relevant descriptors. Section 4 concludes the paper.

2. OUR LOCAL INVARIANT HYBRID DESCRIPTOR

In this section, we first briefly introduce three photometric color invariants and then present our local hybrid descriptor by combining luminance and color information.

2.1 Three Photometric Color Invariants

First we review the theory about the color model. A camera pixel can be modeled as: [13]

$$C(x) = g_d(x)d(x) + g_s(x)s(x) + i(x) \quad (1)$$

In this model, the first two terms respectively represent body reflection and surface reflection. The last term represents interreflection but can mostly be quite small with respect to other terms and usually ignored. The model of camera pixel

color becomes:

$$C(x) = g_d(x)d + g_s(x)s \quad (2)$$

where $g_d(x)$ is a term that depends on the orientation of the surface, d is the color of the diffuse reflected light, $g_s(x)$ is a term that gives the extent of the specular reflection, and s is the color of the source.

In our hybrid descriptor, three color invariants described in [9] rgb , hue , and $l_1l_2l_3$ are exploited, and they can be proved to have certain invariance based on the model above. The rgb descriptor can be considered invariant with respect to lighting geometry and viewpoints when only body reflection exists. If the color of a pixel is represented as (R, G, B) in the RGB color space, then

$$r = \frac{R}{R+G+B}, g = \frac{G}{R+G+B}, b = \frac{B}{R+G+B}. \quad (3)$$

In the case of a white illuminant, the hue descriptor and $l_1l_2l_3$ descriptor can be derived to be invariant with respect to both the lighting geometry and specularities. They are computed respectively through

$$hue = \arctan\left(\frac{\sqrt{3}(R-G)}{R+G-2B}\right), \quad (4)$$

and

$$\begin{aligned} l_1 &= \frac{(R-G)^2}{(R-G)^2 + (R-B)^2 + (G-B)^2} \\ l_2 &= \frac{(R-B)^2}{(R-G)^2 + (R-B)^2 + (G-B)^2} \\ l_3 &= \frac{(G-B)^2}{(R-G)^2 + (R-B)^2 + (G-B)^2}. \end{aligned} \quad (5)$$

It should be noted that the color invariants rgb and $l_1l_2l_3$ can be represented in a two-dimensional space, due to $r+g+b=1$, and $l_1+l_2+l_3=1$.

2.2. Our Hybrid Descriptor

Our local hybrid descriptor is extracted through three steps.

Firstly, distinctive local regions are identified according to the procedures introduced by [3]. Briefly, an original image $I(x, y)$ is represented as $L(x, y, \sigma)$ under a series of different scales σ , and then local extrema are searched in the scale space $D(x, y, \sigma) = L(x, y, k\sigma) - L(x, y, \sigma)$. These local extrema are further verified, and the corresponding keypoint locations are refined. A distinctive local patch is represented by $R(x, y, s)$, where (x, y) denotes the location of an associated keypoint, and s is its scale.

Secondly, for each local patch $R(x, y, s)$, the *SIFT* descriptor and the three photometric invariant color descriptors are extracted. For the *SIFT* descriptor, dominant orientation is first identified based on a weighted gradient orientation histogram to guarantee rotation invariance. We use the same parameters as [3] to compute the *SIFT* descriptor, a $4 \times 4 \times 8 = 128$ dimensional vector.

In extracting color features, the original image is used. Compared with the *SIFT* descriptor, local image patches for extracting color invariant features are not rectangles but circles in our system. The size of a local patch depends on the scale of the associated keypoint. Assume the size of a local patch with the scale s is p in the first step, the size of the corresponding patch for evaluating the three color invariants in the original image is $s^2 \times p$. Before extraction, a local circular patch is divided into several annular sub-regions as shown in Fig. 1, and then the three color invariant features introduced in section 2.1 are computed respectively for each annular sub-region around the keypoint. Apparently, the modified color descriptors also keep the affine invariance. In evaluating the histograms of the three color invariants, local patches are sampled. Inspired by the *SIFT* descriptor, each sample point is assigned with a weight determined by a Gaussian kernel with the standard variance ψ , so a nearer sample point from a keypoint obtains a bigger weight, i.e., a nearer sub-region is more significant and makes more contribution in the descriptors. In our system, the dimension of the rgb descriptor is $3 \times 9 \times 9$, which corresponds to three sub-regions and nine bins quantized for the r and g components. In the same way, the dimension of the hue descriptor is 3×36 , and the dimension of the $l_1l_2l_3$ descriptor is $3 \times 9 \times 9$. The three color descriptors form a color feature vector $color = (rgb, hue, l_1l_2l_3)$ with 594 entries.

In the last step, the *SIFT* descriptor for luminance information and the three modified color descriptors are combined into a unified descriptor. Before the combination, since luminance feature and color feature have different importance in the process of image matching, the normalized *SIFT* descriptor and color descriptors are weighted by a factor of λ . So the final descriptor vector has the form $(SIFT, \lambda * color)$ and 722 dimensions.



Fig. 1 A local image patch around a keypoint is divided into three sub-regions as above and a nearer region from the keypoint will be assigned a larger weight when extracting color features.

2.3. PCA for a Compact Hybrid descriptor

In keypoints matching, the time cost is very large if the dimension of the associated feature is high. To make our hybrid descriptor has a good tradeoff between descriptive power and matching speed, we apply principal component analysis (PCA) [14] to the component of the color features, i.e., $color = (rgb, hue, l_1l_2l_3)$ to reduce the dimension of our

descriptor. Although the three color descriptors can reinforce each other in characterizing color information, they also carry much correlated information. PCA is used to remove the redundant information among the three color invariants.

To obtain an effective feature space, we collect a set of diverse images and extract about 30000 local patches for training. In our system, we use the PCA technique to compress the original 594-dimensional color feature into a compact 60-dimensional feature. If we use $pcolor$ to denote the compact color descriptor, our final hybrid descriptor becomes $(SIFT, \lambda * pcolor)$, a 188 dimensional feature.

3. EVALUATION EXPERIMENTS

We use the publicly available INRIA dataset [15] to evaluate the performance of our hybrid descriptor. The INRIA dataset has been used for evaluating interest point detectors and various local descriptors.

3.1. Matching Strategy and Metric

In the INRIA dataset, each group of images consists of a reference image and multiple transformed images. We use the Lowe’s algorithm [3] to locate keypoints for each image in a group, and perform point-to-point matching between the reference image and a transformed image in a group. For a keypoint A in the reference image, suppose B and C denote the keypoints in the transformed image that have the maximum similarity and the second maximum similarity compared with A based on our local descriptor. If

$$Sim(D_A, D_B) / Sim(D_A, D_C) > \eta, \quad (6)$$

the keypoints A and B are claimed to be matched, where D_A, D_B and D_C are the descriptors for the keypoints A, B and C respectively, η is a predefined threshold. We use the histogram intersection as the similarity metric, i.e.,

$$Sim(D_x, D_y) = \sum_i \min(D_x^i, D_y^i) \quad (7)$$

where D_x^i and D_y^i denote the i th entries of the descriptors D_x and D_y respectively.

In the INRIA dataset, the homography H from the reference image to a transformed image is given. Thus, the corresponding ground-truth point X in the transformed image for the keypoint A can be determined through $P_X = HP_A$, where P_X and P_A are their coordinates respectively. In our experiments, if the geometry distance between B and X is less than a predefined threshold, the system will declare the match between A and B is correct.

3.2. Comparison Experiments

We use the recall vs. 1-precision as the evaluation criterion. Recall is the ratio of the number of correct matches to the number of the keypoints detected in the reference images and 1-precision is the ratio of the number of false matches to

the number of claimed matches. We set up three groups of experiments to compare the performance of our hybrid descriptor with those of other descriptors in keypoints matching. The weight (λ) of the color features component of our hybrid descriptor is set as 0.2 in our experiments. The abbreviated symbols of various descriptors involved in comparison experiments are summarized in the Table 1.

Table 1: Abbreviated symbols and their meanings

Symbols	Meaning	
(R)HS	hue+SIFT	“R” is used to imply that annular sub-regions are divided before extracting color invariants and the Gaussian weight is used when sampling color invariants.
(R)RS	rgb+SIFT	
(R)LS	$l_1l_2l_3$ +SIFT	
RCS	hue+rgb+ $l_1l_2l_3$ +SIFT	
PRCS	hue+rgb+ $l_1l_2l_3$ +SIFT+PCA	

Firstly, we compare the performance of the SIFT descriptor and the hybrid descriptors only one color invariant descriptor involved. The relevant experimental results are shown in Fig. 2 (a), Fig.2 (b) and Fig.2 (c). It can be observed that RHS, RLS and RRS are obviously better than $SIFT$, HS, LS and RS, so our modification in extracting the three color invariant descriptors through introducing sub-regions and Gaussian weight is effective.

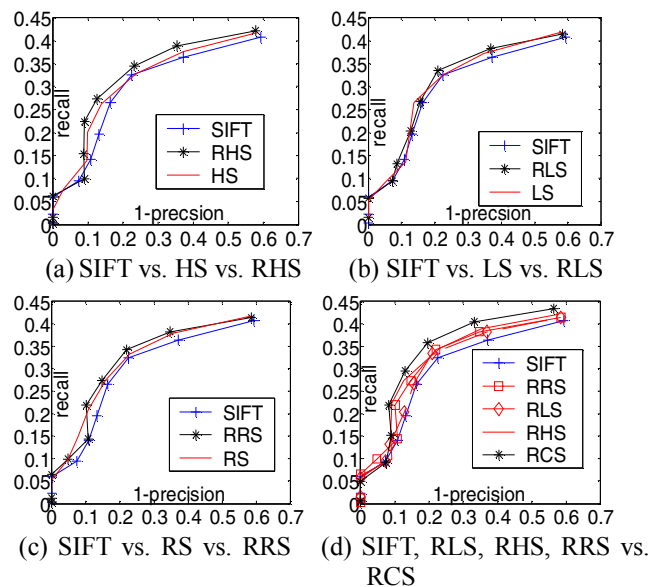


Fig. 2 Performance Comparison between SIFT, other related hybrid descriptors and our hybrid descriptor.

Secondly, we evaluate the performance when we use more color invariant features to characterize local color information. From the results shown in Fig. 2(d), we see that RCS outperforms $SIFT$, RHS, RLS and RRS. It implies that $hue, l_1l_2l_3$ and rgb invariants can supplement each other in characterizing color information. So combination of them

gives rise to a descriptor with more descriptive power than using only one of them.

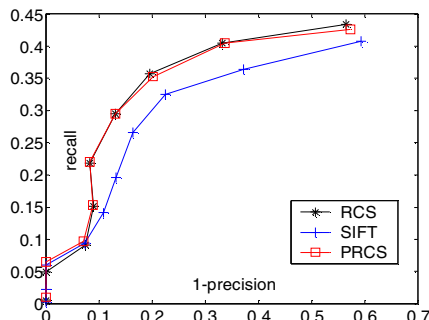


Fig. 3 Evaluation for *SIFT*, RCS, and PRCS (created by applying PCA to RCS)

Thirdly, we evaluate the performance of our compact hybrid descriptor PRCS. The experimental result is shown in Fig.3. We can observe that the new descriptor has almost the same performance compared with RCS, and is much better than SIFT. It also implies that the photometric color invariants hue , $l_1l_2l_3$ and rgb have a certain amount of redundant information. So PCA does not result in performance degradation. Table 2 gives a more intuitive experimental result for different descriptors at a specified 1-precision.

Table 2: The numbers of correct matches for the various descriptors when 1-precision is 0.2

Descriptor (dimension)	SIFT (128)	HS (200)	RHS (236)	LS (384)	RLS (371)
correct matches	79	85	105	82	89
Descriptor (dimension)	RS (384)	RRS (371)	RCS (722)	PRCS (188)	
correct matches	88	98	119	115	

The benefit of PRCS over RCS is the efficiency in matching speed. In our keypoints matching experiments, the time cost of using RCS is 12.048 seconds, while the time cost of using PRCS is 5.266 seconds, which includes the time of projecting the RCS descriptor into the PRCS descriptor. And the time for extracting descriptors is not included. The experiments about matching efficiency are performed in MATLAB 6.5 environment on a PC with 2.8GHz Pentium 4 processor and 256 MB memory.

4. CONCLUSION

In this paper, we present a new local descriptor combining three modified photometric invariant color descriptors with the famous *SIFT* descriptor. The extensive comparison experiments show the local hybrid descriptor outperforms the SIFT descriptor and other hybrid descriptors in which only one color invariant is involved. By applying PCA, our

proposed descriptor becomes more efficient in computation but keeps almost the same performance.

5. ACKNOWLEDGEMENT

We are grateful to Dr. Qin Lei and Dr. Qingfang Zheng for helpful advice, and Dr. Yuanning Li for offering his *SIFT* code. The work is supported by the research start-up fund of GUCAS, in part by National Hi-Tech Development Program (863 Program) of China under Grant: 2006AA01Z117 and National “242” project: 2006A09.

REFERENCES

- [1] Y. Dufournaud, C. Schmid and R. Horaud: “Matching Images with Different Resolutions”. In *CVPR*, Vol. 1, pp. 612-618, 2000.
- [2] K. Mikolajczyk and C. Schmid: “Scale & Affine Invariant Interest Point Detectors”. In *IJCV*, 60(1): pp.63-86, 2004.
- [3] D.G. Lowe: “Distinctive image features from scale invariant keypoints”. In *IJCV*, 60(2): pp.91-110, 2004.
- [4] Y. Ke and R. Sukthankar: “PCA-SIFT: A more Distinctive Representation for Local Image Descriptors”. In *CVPR*, Vol.2, pp.506-513, 2004.
- [5] L. Qin, W. Zeng, W. Gao and W.Q. Wang: “Local Invariant Descriptor For Image Matching”. In: *International Conference on Acoustics, Speech, and Signal Processing*, Vol.2, pp.1025- 1028, 2005.
- [6] K. Mikolajczyk and C. Schmid: “A performance Evaluation of Local Descriptors”. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(10): pp.1615-1630, 2005.
- [7] T. Gevers, S. Ghebreab and A.W.M. Smeulders: “Color Invariant Snakes”. In: *British Machine Vision Conference*, pp. 578-588, 1998.
- [8] V. Gouet and R. Deriche: “Differential Invariants for Color Image”. In: *International Conference on Pattern Recognition*, pp. 838-840, 1998.
- [9] T. Gevers and A.W.M. Smeulders: “Color Based Object Recognition”. In: *Pattern Recognition*, 32: pp. 453-464, 1999.
- [10] J. Geusebroek, R. Boomgaard and A.W.M. Smeulders: “Color Invariance”. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(12): pp.1338-1550, 2001.
- [11] T. Gevers and A.W.M. Smeulders: “Pictoseek: Combining Color and Shape Invariant Features for Image Retrieval”. In *IEEE Transactions on Image Processing*, 9(1): pp.102-119, 2000.
- [12] J. Weijer and C. Schmid: “Coloring Local Feature Extraction”. In *European Conference on Computer Vision*, Vol.2, pp.334 -348, 2006.
- [13] D.A. Forsyth, J. Ponce: *Computer Vision: A Modern Approach*. Electronic Industry Press, June, 2004.
- [14] R.O. Duda, P.E. Hart and D.G. Stork: *Pattern Classification*, 2nd ED. New York: John Wiley and Sons, 2001.
- [15] Interest point test sequences. <http://lear.inrialpes.fr/data>.