Harris Feature Vector Descriptor (HFVD)

X. G. Wang, F. C. Wu and Z. H. Wang
NLPR, Institute of Automation, Chinese Academy of Science, China
xgwang@nlpr.ia.ac.cn

Abstract

A new image feature called Harris feature vector is defined in this paper, which effectively describes the image gradient distribution. By computing the mean and the standard deviation of the Harris feature vector in key point neighborhood, a novel descriptor for key points matching is constructed, which is invariant to image rigid transformation and linear intensity change. Experimental evidence suggests that the novel descriptor has a good adaptability to slight view point changing, JPEG compression as well as nonlinear change of intensity.

1. Introduction

Key points matching is a key step for most of the problems in computer vision, such as object localization and recognition, image registration, three-dimensional model acquisition, video content understanding, etc. It has been a challenging problem because of different image types and geometric transformations. There are two major steps involved in the process of key points matching: the first and foremost is key points description and the other matching. A typical way of describing key points is to compute local descriptors. The descriptor can be roughly divided into two types, i.e. intensity-based descriptor and gradient-distribution-based one.

Intensity-based descriptor was first developed in [11], which is just a vector of image pixels. Zhang etc. [15] introduced a similar descriptor called Cross-correlation, and its high dimensionality results in a high computational complexity. Based on the work of Schmid and Mohr [14], Johnson and Hebert [7] introduced a more distinctive descriptor Spin Image, which is invariant to image rotation and has a low dimensionality. Among the gradient-distribution-based descriptors, SIFT [8] appeared as the most famous one. From then, a lot of similar work has been done, such as the Shape Context [2], GLOH [10] and SURF [1].

Some other kinds of techniques are also reported in the literature, such as Local Jet [3], steerable Filters [4], Moment Invariants [5], Complex Filters [12, 13] etc. Mikolajczy and Schmid [10] evaluated ten different popular descriptors on real images with different geometric, photometric transformations and scene types, and then concluded that the performance of descriptors is independent of feature detectors; SIFT Based descriptors perform best among the descriptors with high dimensionality.

In this paper, we introduce a new image feature called HFV (Harris feature vector) in order to effectively describe the gradient distribution. By computing the mean and the standard deviation of this feature, a novel descriptor called HFVD for key points matching is constructed , which is invariant to image rigid transformation as well as linear intensity change. The experimental results show that the novel descriptor has a good adaptability to slight view point changing, JPEG compression as well as nonlinear change of intensity.

The paper is organized as follows: Section 2 introduces the Harris feature vector and discusses its properties. Section 3 shows in detail how to construct HFVD for key points matching. The experimental results are reported in section 4 and section 5 concludes the paper.

2. Harris feature vector

In this paper, $\nabla f(\mathbf{x}) = (f_x(\mathbf{x}), f_y(\mathbf{x}))^\top$ denotes the gradient of the image point $\mathbf{x}$. $N(\nabla f(\mathbf{x}))$ and $N_\perp(\nabla f(\mathbf{x}))$ denote the normalized $\nabla f(\mathbf{x})$ and the unit orthogonal vector of $\nabla f(\mathbf{x})$ respectively, thus an orthogonal transformation can be formed as

$$T_{\mathbf{x}} = [N(\nabla f(\mathbf{x})), N_\perp(\nabla f(\mathbf{x}))]^\top \tag{1}$$

and

$$T_{\mathbf{x}}(\mathbf{y}) = T_{\mathbf{x}} \nabla f(\mathbf{y}) \triangleq \left(\hat{f}_x(\mathbf{y}), \hat{f}_y(\mathbf{y})\right)^\top \tag{2}$$

978-1-4244-2175-6/08/$25.00 ©2008 IEEE

Based on this, we define the positive part and the negative part of the transformation respectively as follows:

$$\begin{aligned} T_{\mathbf{x}}(\mathbf{y})_+ &= \frac{(|\hat{f}_x(\mathbf{y})|, |\hat{f}_y(\mathbf{y})|)^\top + (\hat{f}_x(\mathbf{y}), \hat{f}_y(\mathbf{y}))^\top}{2} \\ &\triangleq \left(\hat{f}_{x+}(\mathbf{y}), \hat{f}_{y+}(\mathbf{y})\right)^\top \end{aligned} \tag{3}$$

$$\begin{aligned} T_{\mathbf{x}}(\mathbf{y})_- &= \frac{(|\hat{f}_x(\mathbf{y})|, |\hat{f}_y(\mathbf{y})|)^\top - (\hat{f}_x(\mathbf{y}), \hat{f}_y(\mathbf{y}))^\top}{2} \\ &\triangleq \left(\hat{f}_{x-}(\mathbf{y}), \hat{f}_{y-}(\mathbf{y})\right)^\top \end{aligned} \tag{4}$$

and it is easy to see that $T_{\mathbf{x}}(\mathbf{y}) = T_{\mathbf{x}}(\mathbf{y})_+ - T_{\mathbf{x}}(\mathbf{y})_-$.

Then, we will define the Harris feature matrices, which are similar to the Harris second moment matrix [6]: Let $\Omega_\varepsilon(\mathbf{x}) = \{\mathbf{y} : \|\mathbf{y} - \mathbf{x}\| \leq \varepsilon\}$, a circle with center $\mathbf{x}$ and radius ε. The positive and negative Harris feature matrices are defined respectively as

$$\mu_+(\mathbf{x}) = \begin{bmatrix} \int_\Omega \hat{f}_{x+}(\mathbf{y})^2 d\mathbf{y} & \int_\Omega \hat{f}_{x+}(\mathbf{y})\hat{f}_{y+}(\mathbf{y}) d\mathbf{y} \\ \int_\Omega \hat{f}_{x+}(\mathbf{y})\hat{f}_{y+}(\mathbf{y}) d\mathbf{y} & \int_\Omega \hat{f}_{y+}(\mathbf{y})^2 d\mathbf{y} \end{bmatrix} \tag{5}$$

$$\mu_-(\mathbf{x}) = \begin{bmatrix} \int_\Omega \hat{f}_{x-}(\mathbf{y})^2 d\mathbf{y} & \int_\Omega \hat{f}_{x-}(\mathbf{y})\hat{f}_{y-}(\mathbf{y}) d\mathbf{y} \\ \int_\Omega \hat{f}_{x-}(\mathbf{y})\hat{f}_{y-}(\mathbf{y}) d\mathbf{y} & \int_\Omega \hat{f}_{y-}(\mathbf{y})^2 d\mathbf{y} \end{bmatrix} \tag{6}$$

and the vector

$$HFV(\mathbf{x}) = \left(det_+^{1/2}(\mathbf{x}), trace_+(\mathbf{x}), det_-^{1/2}(\mathbf{x}), trace_-(\mathbf{x})\right) \tag{7}$$

where

$$det_+^{1/2}(\mathbf{x}) = det\left(\mu_+(\mathbf{x})\right)^{1/2}, \tag{8}$$

$$trace_+(\mathbf{x}) = trace\left(\mu_+(\mathbf{x})\right), \tag{9}$$

$$det_-^{1/2}(\mathbf{x}) = det\left(\mu_-(\mathbf{x})\right)^{1/2}, \tag{10}$$

$$trace_-(\mathbf{x}) = trace\left(\mu_-(\mathbf{x})\right). \tag{11}$$

is called the Harris feature vector of point $\mathbf{x}$ with respect to Ω_ε. With the Harris feature vector above, we can distribute a 4-dimensional vector to every image point.

Figure 1 presents Harris feature energy images. (a) is an original image, (b) is the determinant of the positive Harris feature matrix,(c) is the determinant of the negative Harris feature matrix, (d) is the trace of the positive Harris feature matrix, and (e) is the trace of the negative Harris feature matrix. It can be seen that the determinants of both the positive and negative Harris feature matrices represent the feature points of image including corners and edge points with big curvature, and the trace of the two matrices represent the image edge information. It is obvious that the Harris feature vector can effectively distill the image feature.

It is not difficult to see that the Harris feature vector has the following properties concerning the image transformations:

1. For rigid transformation $g(\mathbf{x}') = f(\mathbf{x})$ $(\mathbf{x}' = R\mathbf{x} + \mathbf{t})$, the equation holds:

$$HFV(\mathbf{x}') = HFV(\mathbf{x})$$

2. For scale change $g(\mathbf{x}') = f(\mathbf{x})$ $(\mathbf{x}' = \sigma\mathbf{x})$, the equation holds:

$$HFV(\mathbf{x}') = (1/\sigma^2)HFV(\mathbf{x})$$

3. For linear change of intensity $g(\mathbf{x}') = \alpha f(\mathbf{x}) + \beta$, the equation holds:

$$HFV(\mathbf{x}') = \alpha^2 HFV(\mathbf{x})$$

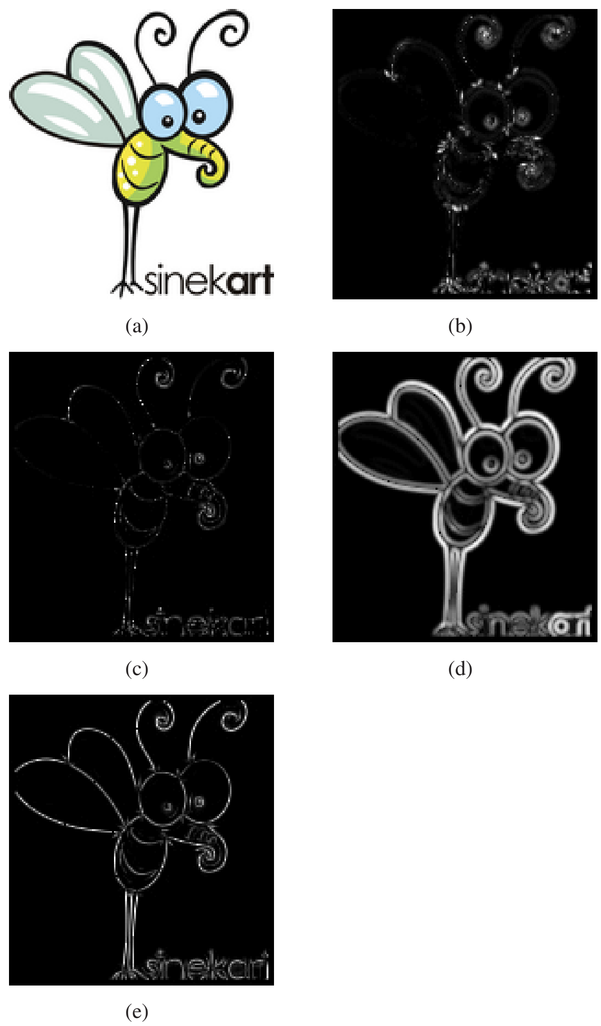

Figure 1. Harris feature energy images

3 Point descriptor

In this section, the Harris feature vector will be used to construct a new descriptor for point matching. Let $\Omega_r(\mathbf{x}) = \{\mathbf{y} : \|\mathbf{y} - \mathbf{x}\| \leq r\}$ be a circular neighborhood of key point $\mathbf{x}$ with radius r. In order to employ the statistical characteristics of the Harris feature vector in the neighborhood Ω_r, we divide Ω_r into certain sub-regions, such as $R_1, R_2, \ldots, R_M$. By computing the mean and the standard deviation of the Harris feature vector in the sub-region R_i, we obtain two 4-dimensional vectors:

$$m_i(\mathbf{x}) = \frac{1}{\#R_i} \sum_{\mathbf{y} \in R_i} HFV(\mathbf{y}), \tag{12}$$

$$sd_i(\mathbf{x}) = \sqrt{\frac{1}{\#R_i} \sum_{\mathbf{y} \in R_i} \big(HFV(\mathbf{y}) - m_i(\mathbf{x})\big)^2}. \tag{13}$$

Let

$$m(\mathbf{x}) = \big(m_1(\mathbf{x}), m_2(\mathbf{x}), \ldots, m_M(\mathbf{x})\big), \tag{14}$$

$$sd(\mathbf{x}) = \big(sd_1(\mathbf{x}), sd_2(\mathbf{x}), \ldots, sd_M(\mathbf{x})\big). \tag{15}$$

By normalizing $m(\mathbf{x})$ and $sd(\mathbf{x})$ to unit vector respectively and concatenating them into a single vector, we obtain an 8M-dimensional vector:

$$HFVD(\mathbf{x}) = \Big(\frac{m(\mathbf{x})}{\| m(\mathbf{x}) \|}, \frac{sd(\mathbf{x})}{\| sd(\mathbf{x}) \|}\Big) \in R^{8M} \tag{16}$$

This vector is called HFV descriptor of point $\mathbf{x}$, which is denoted as HFVD.

Different kinds of HFVDs can be obtained according to different partitions of Ω_r. In this paper, we set up $r = 16$ and consider only one simple partition as shown in figure 2.

Figure 2 shows that the circular neighborhood of a key point is divided into 4 isometric sectors based on the main orientation (Lowe [8]), and is simultaneously divided into 4 concentric rings of equal width, thus we get 13 sub-regions. With the neighborhood partition, each key point gets a descriptor, i.e. a 104-dimensional HFVD.

From the properties of Harris feature vector(HFV), we can see that the HFVD is invariant to image rigid transformations as well as linear change of intensity.

4 Experimental results

In this section, the performance of our descriptor on real images will be tested and compared with that of SIFT descriptor (We find that SIFT descriptors computed from a 32×32 array perform much better than those computed from a 16×16 sample array, so 32×32 sample array will be used in all the experiments of this section). The criterion of matching performance is the number of matches with the same precision [10]. In this paper, the match measure and the criterion are the Euclidean distance between descriptors and NNDR [10] respectively. In the first following experiment on image rotation, we use the epipolar geometry constrain (computed with RANSAC technique) to distinguish each match is true or false. In the other experiments, the correctness of each match is tested through the homography matrices offered by [10]. For the limited paper length, only the matching results of Harris point are reported here. Shown in our experiments with a large number of images, the matching results on the other kinds of key points, such as LOG point , are similar to that of Harris point.

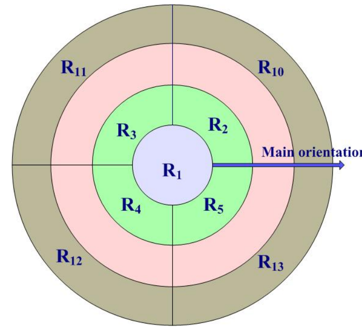

Figure 2. Neighborhood partition

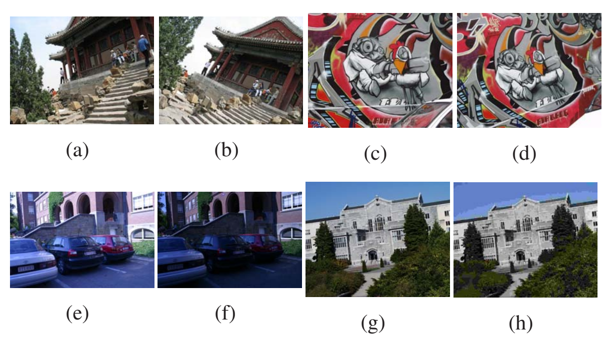

Figure 3. Images used in experiments.

Image rotation Figure 3 (a) and (b) is an image pair obtained by rotating the camera with a rotation angle of 50^o approximately; figure 4 (a) offers the matching result, from which we can see that our descriptor outperforms the SIFT descriptor on image rotation. With the same precision higher than 80%, the match numbers of our descriptor is almost as twice as that of SIFT's.

Image affine Figure 3 (c) and (d) is a pair of images with affine distortion from [10]; figure 4 (b) is the matching result, which indicates that although with lower dimension (104 vs. 128) our descriptor performs a little better.

Nonlinear intensity change Figure 3 (e) and (f) is a pair of images with nonlinear intensity change from [10]; figure 4 (c) is the matching result which shows that our descriptor does not perform as well as SIFT on image nonlinear intensity change. That is to say our descriptor is more sensitive to intensity change.

JPEG compression Figure 3 (g) and (h) is a pair of images with JPEG compression, also from [10]. It can be seen from figure 4 (d) that our descriptor outperforms SIFT apparently, which means that our descriptor has a very good adaptability to image JPEG compression.

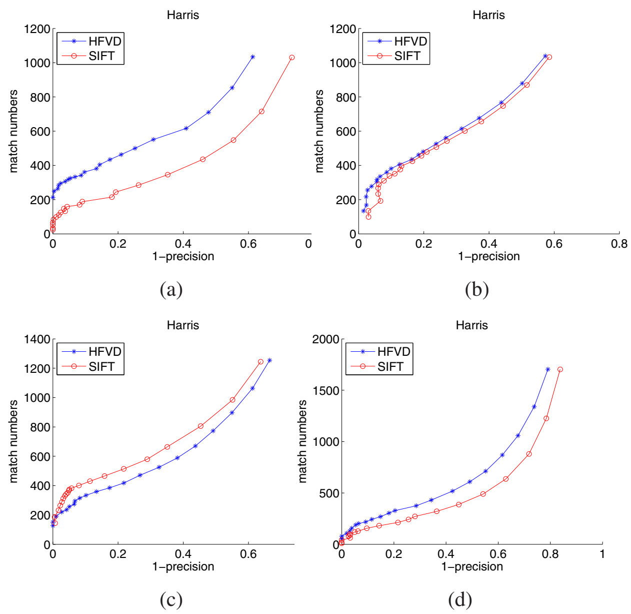

Figure 4. Experimental results.

5 Conclusion

We have described a novel descriptor that is constructed by a new image feature. The construction of the descriptor is simple since it only employs the mean and the standard deviation calculation. Since the very image feature is effective in capturing the gradient distribution, the descriptor constructed by it performs very well in key points matching.

The HFVD is powerful and needs to be explored in further detail. One practical question is how to make our descriptor invariant to scale and large view point changing, and the work in [8] and [9] will be good paradigms.

References

[1] H. Bay, T. Tuytelaars, and L. Gool. Surf: Speeded up robust features. 1:404–417, 2006. Proc. of ECCV.

[2] S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. *IEEE Trans. On Pattern Analysis and Machine Intelligence*, 24(4):509–522, 2002.

[3] L. Floract, B. Romeny, J. Koenderink, and M. Viergever. General intensity transformations and second order invariants. pages 338–345, 1991. In Proc. of the 7th conference on Image Analysis.

[4] W. Freeman and E. Adelson. The design and use of steerable filers. *IEEE Trans. On Pattern Analysis and Machine Intelligence*, 13(9):891–960, 1991.

[5] L. Gool, T. Moons, and D. Ungureanu. Affine/ photometric invariants for planar intensity patterns. pages 642–651, 1996. In Proc. of the 4th ECCV, Cambridge, UK.

[6] C. Harris and M. Stephens. A combined corner and edge detector, 1988. Proceeding 4th Alvey Vision Conference.

[7] A. Johnson and M. Hebert. Using spin images for efficient object recognition in cluttered 3d scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(5):433–449, 1999.

[8] D. Lowe. Distinctive image feature from scale invariant keypoint. *International Journal of Computer Vision*, 60(2):91–110, 2004.

[9] K. Mikolajczyk and C. Schmid. Scale and affine invariant interest point detectors. *International Journal of Computer Vision*, 60(1):63–86, 2004.

[10] K. Mikolajczyk and C. Schmid. A performance evaluation of local decriptors. *IEEE Trans On Pattern Analysis and Machine Intelligence*, 27(10):1615–1630, 2005.

[11] H. Moravec. Rover visual obstacle avoidance. pages 785–790, 1981. International Joint Conference on Artificial Intelligence.

[12] F. Schaffalitzky and A. Zisserman. Viewpoint invariant texture macthjng and wide baseline stereo. 2:636–643, 2001. In Proc. of ICCV.

[13] F. Schaffalitzky and A. Zisserman. Multi-view matching for unordered image sets. 1:414–431, 2002. In Proc. of ECCV.

[14] C. Schmid and R. Mohr. Local grayvalue invariants for image retrieval. *IEEE Trans. On Pattern Analysis and Machine Intelligence*, 19(5):530–534, 1997.

[15] Z. Zhang, R. Deriche, O. Faugeras, and O. Luong. A robust technique for matching two un-calibrated images through the recovery of unkown epipolar geometry. *Artificial Intelligence*, 78:87–119, 1995.