

MIFT: A Mirror Reflection Invariant Feature Descriptor

Xiaojie Guo, Xiaochun Cao, Jiawan Zhang, and Xuewei Li

School of Computer Science and Technology
Tianjin University, China
{xguo,xcao,jwzhang,lixuewei}@tju.edu.cn

Abstract. In this paper, we present a mirror reflection invariant descriptor which is inspired from SIFT. While preserving tolerance to scale, rotation and even affine transformation, the proposed descriptor, MIFT, is also invariant to mirror reflection. We analyze the structure of MIFT and show how MIFT outperforms SIFT in the context of mirror reflection while performs as well as SIFT when there is no mirror reflection. The performance evaluation is demonstrated on natural images such as reflection on the water, non-rigid symmetric objects viewed from different sides, and reflection in the mirror. Based on MIFT, applications to image search and symmetry axis detection for planar symmetric objects are also shown.

1 Introduction

Local image feature is a prerequisite in the areas of computer vision such as image retrieval [1], 3-D reconstruction, object recognition [2], camera calibration, robot navigation, gesture recognition, image search [3] and building panoramas [4]. To handle image variability caused by rotations, scale changes, occlusions, varying illuminations and even perspective distortions, numerous techniques such as Harris corner detector [5], SIFT [6], SURF [7], GLOH [8] and DAISY [9] are purposed. Harris corner detector can find feature points with invariance of rotation and illumination. Actually, what it detects are not just corners but also points that have great gradients in multiple directions at a fixed scale. In other words, scale changing is not handled. SIFT makes full use of the gradient property and utilizes a spatial weighting scheme to differentiate pixels in the corresponding collection, i.e. the histogram of each pixel in the collection is weighted by the gradient magnitude and a Gaussian weighted window. SURF adopts Fast-Hessian detector to quickly detect features within images which owes much to the integral images and uses the Harr wavelet response to capture the texture properties around interest points. Histogram of Oriented Gradients (HOG) [10] uses the distribution of gradients within a set of pixels nearby the point of being calculated. DAISY [9] is another fast solution of local image descriptor designed for dense wide-baseline matching purpose, and is intrinsically tolerant to rotations due to its circular interest region design.

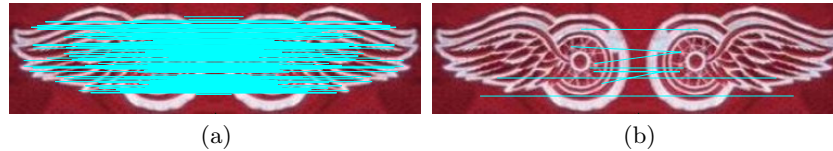


Fig. 1. Comparison of matching between MIFT and SIFT in the mirror reflection situation. The left image is the Red Wings logo from movie “Ferris Bueller’s Day Off”, the right one is the horizontally reflected version of it. (a) Matching result of MIFT. (b) Matching result of SIFT.

Although so many different schemes of feature extraction have been proposed, it is far from complete and robust enough to solve all of the problems encountered in the real life. While those above descriptors are successful in handling most of distortions and illumination variances, they fail in the situation of mirror reflection as shown in Fig. 1. Since the invariance of SIFT is remarkably robust, many works have been carried out based on it, such as the 3D SIFT descriptor [11] that is proposed to recognize actions and PCA-SIFT [12] which aims at creating a more robust and shorter descriptor to image deformations than SIFT. The purposed descriptor MIFT improves the invariance to mirror reflection and performs as well as SIFT when there is no mirror reflection. Particularly, we reorganize the structure of SIFT descriptor, and also adjust the matching strategy accordingly.

2 MIFT

MIFT improves SIFT to advance the invariance of features in mirror reflection situations. Note that the proposed solution is able to be used in other SIFT-like descriptors with minor changes. We choose SIFT as basis, because of its outstanding stability, robustness and distinctiveness.

2.1 Mirror Reflection

In the real world, mirror reflection generally appears in three different ways: horizontal reflection, vertical reflection, and combined reflection which is horizontally and vertically reflected. However, for rotation invariant descriptors, the horizontal reflection and the vertical reflection are equivalent by rotating the coordinate system. The same argument holds true for the original and the combined reflection. Therefore two cases, i.e. the original image and the horizontal reflected one, need to be handled. The relationship between the same region after specifying the dominant orientation in the original image and in the horizontally reflected one is that the row order of pixels and cells is identical and the column order inverse, i.e. the gradients of the same pixel in the interest region after orientation assignment in different reflection cases are approximate¹ bilateral symmetry.

¹ The noises and distortions might influence the gradients.

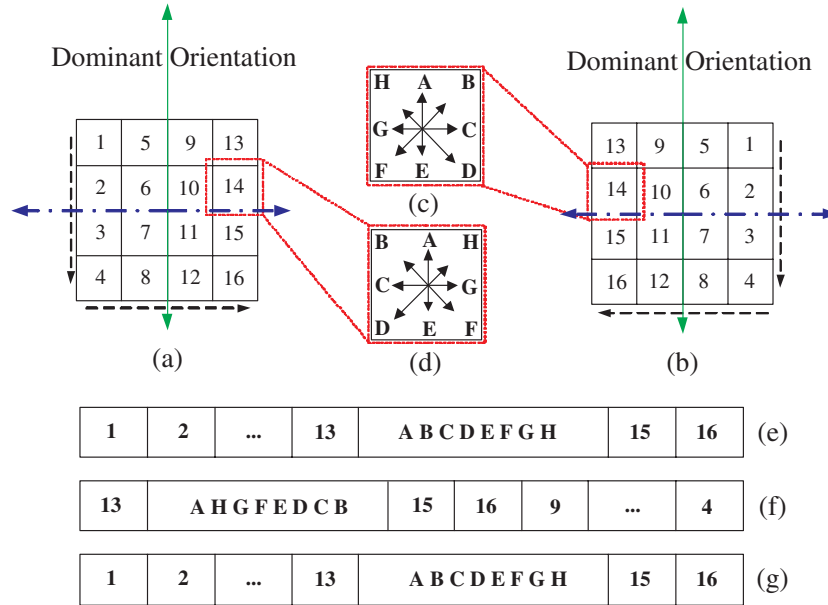


Fig. 2. Illustration of the descriptor organizations of MIFT and SIFT in the situations with and without mirror reflection. (a) A keypoint and its interest region in the original image. (b) (a) in the horizontally reflected image. (c) Distribution of eight orientations in the 14th cell of (b). (d) Distribution of eight orientations in the 14th cell of (a). (e) SIFT and MIFT descriptor for (a). (f) SIFT descriptor for (b). (g) MIFT descriptor for (b).

2.2 Descriptor Reorganization

A vector containing the magnitudes of all the orientation histogram entries in a region around the keypoint forms a descriptor, and a 4×4 array with 8 orientation bins in each is proved to be the successful format in [6]. As SIFT, a 128-D, i.e. $4 \times 4 \times 8$, vector is also chosen to support MIFT descriptor, although 128-D is not compulsory. Figure 2 (a) is a keypoint with its interest region in the original image, and Figure 2 (b) is (a) in the horizontally reflected image, both of which are after specifying dominant orientation. Note that there is no limitation about which direction the mirror reflection axis is alone because of the rotation invariance.

Order of 16 Cells. SIFT uses a fixed order to organize the 16 cells in the interest region after specifying the dominant orientation out of 36 candidates. Note that there might be multiple descriptors for the same combination of location and scale, because it is possible that multiple peaks of 36 orientations are close to the highest peak. As shown in Fig. 2 (a), SIFT might adopt the column-major-order encoding strategy. As a result, the 16 cells are ordered as Fig. 2 (e). However, the column order is reversed after mirror reflecting as shown in Fig. 2 (b). The original fixed encoding strategy used in SIFT would arrange the 16 cells as Fig. 2 (f). Although this encoding strategy is invariant to rotation and scale,

and even tolerant to affine transformation, it does not result in the same order in the situation of mirror reflection. We introduce an adaptive encoding technique that is able to preserve the order of the 16 cells in the mirror reflection case. Fortunately, there are only two directions to choose from, from left to right and vice versa, as the order of the rows are the same in the column-major-order encoding strategy. Intuitively, two magnitudes of the directly left- and right-pointing orientations (the blue dashed arrows) can be used to select which direction goes first. However, this method is sensitive to be relied on due to noises and distortions. The direction is instead decided based on the summations of the magnitudes of all left-pointing and right-pointing orientations,

$$m_r = \sum_{k=1}^{(N_{bin}-2)/2} L_{(n_d-k+N_{bin})\%N_{bin}}, \quad (1)$$

$$m_l = \sum_{k=1}^{(N_{bin}-2)/2} L_{(n_d+k+N_{bin})\%N_{bin}}, \quad (2)$$

where N_{bin} is the number of orientation bins that is 36 for MIFT, n_d is the dominant orientation index and L_i is the gradient magnitude in the $\frac{i \times 2\pi}{N_{bin}}$ direction. According to this measurement, we adaptively change the encoding strategy from the fixed order to the one indicated by the winner of m_l and m_r . Theoretically, this new encoding approach makes the sequence of cells unique in whichever case as shown in Fig. 2 (g). Nevertheless, m_l and m_r might be close to each other due to noises, lighting variations and other factors. To be more robust in such situations, an extra descriptor is created in opposite order to support the feature, when $\min\{m_l, m_r\} > \tau \max\{m_l, m_r\}$. In our experiments, τ is 70% for all the evaluations.

Order of 8 Orientations. Reorganization of cells described above is just one of two essential steps. The other equally significant procedure is to restructure the order of orientation bins in each cell. As shown in Fig. 2 (c) and (d), all gradients in each cell are divided into their nearest bins of eight directions. Figure 2 (c) and (d) present the same cell in the cases with and without mirror reflection. Main difference between them is order change of bins without serious influence on their strengths. Therefore, we encode them in anticlockwise order beginning with ‘A’ in the case of Fig. 2 (a), and in clockwise beginning with ‘A’ in the case of Fig. 2 (b) based on the comparison of m_l and m_r .

Finally, for the same keypoint, we obtain a unique descriptor in different mirror reflection cases by the method described above. For the features in the situations of Fig. 2 (a) and (b), the descriptors are identical which are shown in Fig. 2 (e) and (g).

2.3 Matching

Due to the change of the signature strategy described in Section 2.2, every descriptor from SIFT may map to one or two from MIFT, decided by the

similarity between m_l and m_r . Recall the original matching method of SIFT just compares the closest and the second closest descriptors. Therefore, it may miss pairs that should be on the list of good matches since MIFT might have multiple similar descriptors for one keypoint. To reduce the miss rate, we introduce an improved matching method. Algorithm 1, IMM, gives pseduo-code for the improved matching method.

Algorithm 1. Improved matching method (IMM)

Require: $k = 1$
Ensure: $\text{match}(i) = \text{IMM}(\text{des1}(i), \text{des2})$
 $[\text{vals}, \text{indices}] = \text{Sort}(\arccos(\text{des1}(i)^T \text{des2}), \text{ascend})$
repeat
 $k = k + 1$
 if $\text{vals}(1) < \text{distRatio} \times \text{vals}(k)$ **then**
 $\text{match}(i) = \text{indices}(1)$
 break
 else
 $\text{match}(i) = 0$
 end if
until The location and scale combinations of $\text{vals}(1)$ and $\text{vals}(k)$ are not identical

In the IMM, distRatio is set the same as SIFT, $\text{des1}(i)$ is the i^{th} descriptor in the set of the comparing image descriptors and des2 is the matrix containing all of the compared image descriptors. According to the characteristic of the matching strategy, the IMM aims at preventing potential true matches from ignoring due to similar matched descriptors for the same keypoint, we consider different combinations of location and scale as different keypoints, alternatively the descriptors at the same location and scale may be probably very close. Theoretically, the matching strategy needs to constrain the influence from close matched descriptors of the same keypoint. The improved matching performance is demonstrated in Section 3.

3 Results

MIFT aims at advancing state-of-the-art descriptors to be invariant to the mirror reflection. In order to show the benefit of MIFT, we tested the performance on natural images with and without mirror reflection against SIFT.

To quantify the performance of MIFT compared to SIFT in mirror reflection situations, we choose some typical scenes including symmetric image pairs, reflection on the water, non-rigid symmetric objects viewed from different sides and reflection in the mirror as shown in Fig. 3. The comparison results are shown in Fig. 4 (a). For the same four pairs of natural images with mirror reflection, SIFT fails to find reasonable numbers of matches, but MIFT finds (78, 70, 57, 36) matches, with very few mismatches (6, 2, 2, 0).

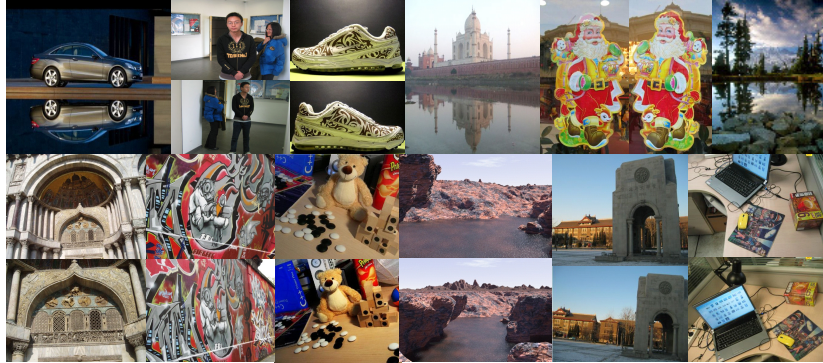


Fig. 3. Example images from image data set for the comparison between MIFT and SIFT. **Top two rows:** test images with mirror reflection. **Bottom two rows:** test images without mirror reflection.

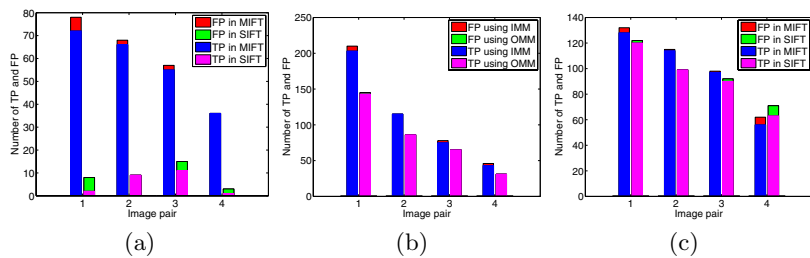


Fig. 4. Comparisons between true positive matches (TP) and false positive ones (FP). (a) Comparison between MIFT and SIFT for image pairs with mirror reflection. (b) Comparison between the improved matching method (IMM) and the original matching method (OMM). (c) Comparison between MIFT and SIFT for image pairs without mirror reflection.

As described in Section 2.3, the improved matching method aims at reducing the miss rate. As shown in Fig. 4 (b), we plot the differences between improved matching method and the original method referring to the method used in SIFT. For the same four pairs of images which are randomly selected from the data set with and without mirror reflection, the true positive matches increase a lot from (143, 86, 65, 31) to (203, 115, 75, 43). Simultaneously, the mismatches increase also for the test pair 1, 3, 4. However, the positive numbers are significantly bigger than the negative ones.

We also test and verify the performance of MIFT compared with SIFT on natural images without mirror reflection. Some of them are shown in Fig. 3. As shown in Fig. 4 (c), the numbers of the true positive and false positive matches of MIFT are similar to those of SIFT. However, for most cases of Fig. 4, MIFT performs slightly better than SIFT due to the characteristic of the IMM, i.e. the descriptors obtained at the same location and scale are considered as the same keypoint and are ignored.

4 Applications

4.1 Application to Image Search

Image feature is the basis of image understanding for computer, and it is the fundament of many multimedia applications such as image search from mass image data. No doubt directly comparing query images with the images in database is the simplest method. In addition, in [3], the authors present two ways to search images. 1) Hamming embedding (HE) and 2) weak geometric consistency constraint (WGC). These two approaches are based on the bag-of-feature method with k-means algorithm. However, all these methods can't solve the flip detection effectively but implement SIFT twice for every query image (SIFT₁ for the original query images, and SIFT₂ for the manually flipped ones).

Duplicate computation of SIFT for every query image is rather laborious and inflexible, we apply MIFT to address this tough problem. Note that we just implement the directly comparing method, however HE and WGC can be easily transplanted to MIFT from SIFT.

The image data are from INRIA Holidays dataset², and we randomly select 300 images from INRIA Holidays dataset as our database, and another 36 images as query set. The one third of the query images are without mirror reflection (NM), another one third are with horizontal mirror reflection (HM) and the remaining one third are with vertical mirror reflection (VM).

We separately use MIFT and SIFT to detect and describe features for every image from our database and query set. As shown in Table 1, MIFT successfully searches 35 images out of 36 in which 34 successfully registered images are first-ranking and 1 image second-ranking. There is only 1 failed to find the related images. This very image and its corresponding one in the database are shown in Fig. 5. For SIFT, as shown in Table 1, it only correctly finds 11 non-reflection images out of 36 for the first process which contains 12 images without mirror reflection, and for the second process it finds 23 first-ranking images with horizontal or vertical mirror reflection, and 1 third-ranking which is second-ranking using MIFT. The incorrect registration image pair is the same as MIFT, as shown in Fig. 5.

The reason why the left image pair in Fig. 5 is not first-ranking is that there are too many repetitive structures within the images that lead to close arc cosine values which are rejected by OMM and IMM. To handle this problem caused by repetitive structures, W. Zhang and J. Kosecka [13] propose an additional criterion which increases true positive rate while keeps false positive low. For the right image pair in Fig. 5, MIFT and SIFT can not obtain reasonable matches since the image pair are changed too much with respect to the scale and occlusion. Some search results are shown in Fig. 7, the images from the database with top 3 scores are shown for every query image.

² <http://lear.inrialpes.fr/~jegou/data.php>

Table 1. The results and their ranks of image search using MIFT and SIFT

#Image pairs	MIFT	SIFT ₁	SIFT ₂
12(NM)	TP, 11, 1 st ; FP, 1	TP, 11, 1 st ; FP, 1	FP, 12
12(HM)	TP, 12, 1 st	FP, 12	TP, 12, 1 st
12(VM)	TP, 11, 1 st , 1, 2 nd	FP, 12	TP, 11, 1 st , 1, 3 rd

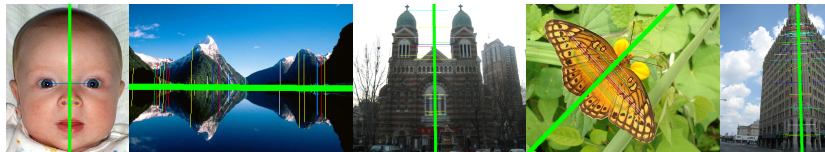
**Fig. 5.** **Left image pair:** the image pair with vertical reflection is queried at the second rank using MIFT, and at the third rank using SIFT. **Right image pair:** the image pair failed to be matched by both MIFT and SIFT.

4.2 Application to Detection of Symmetry Axis

In the real life, symmetric objects are ubiquitous, especially in urban area. Symmetry detection has been regarded as a useful and important technique in computer vision. Many symmetry detection schemes are proposed including human gait detection [14] and a focused version of Generalized Symmetry Transform [15]. As is known, the rules for planar symmetric objects are simple, symmetric points in the symmetric object lie on a single line perpendicular to the symmetry axis and their perpendicular distances to the symmetry axis are identical. However, the symmetric object under perspective projection is an exception of these rules. Although the ratios of lengths are not preserved under perspective projection, the following cross ratio constraint [16] is still valid,

$$\{\mathbf{v}_x, \mathbf{l}_i; \mathbf{c}_i, \mathbf{r}_i\} = \frac{(\mathbf{v}_x - \mathbf{l}_i)(\mathbf{c}_i - \mathbf{r}_i)}{(\mathbf{v}_x - \mathbf{r}_i)(\mathbf{l}_i - \mathbf{c}_i)} = 1 \quad (3)$$

where \mathbf{v}_x is the vanishing point calculated by the parallels linking matched pixels, \mathbf{l}_i and \mathbf{r}_i are the endpoints of the line and \mathbf{c}_i is the point on the symmetry axis.

**Fig. 6.** Example results of symmetry axis detection. The symmetry axes are marked with green straight line. Please zoom in the picture to see MIFT matches.

Query Images		MIFT			SIFT ₁ + SIFT ₂		
		1 st	2 nd	3 rd	1 st	2 nd	3 rd
H M							
V M							
N M							

Fig. 7. Example results of image search. SIFT results are obtained by implementing query procedure twice, while MIFT only once. **Top two rows:** The query images with horizontal mirror reflection (HM). **Middle two rows:** The query images with vertical mirror reflection (VM). **Bottom two rows:** The query images without mirror reflection (NM). The results are labeled as true positive (TP) or false positive (FP) below the searched images.

We use MIFT to detect the endpoints \mathbf{l}_i and \mathbf{r}_i . \mathbf{v}_x is then estimated from concurrent lines that link \mathbf{l}_i and \mathbf{r}_i . Finally we use RANSAC [17] to find the optimal symmetry axis. Some results are shown in Fig. 6. Light colorful lines link matched pixel pairs, outliers of matched symmetric points have been eliminated by RANSAC.

5 Conclusion

In this paper, we have proposed a mirror reflection invariant descriptor, MIFT, and have verified its improved performance over the state-of-the-art SIFT in mirror reflection situations. MIFT performs comparably in general cases when compared with SIFT. We elaborate the new strategy for organizing descriptor, and describe the matching method used to reduce recall loss. MIFT instead of SIFT applied to the image search can capably handle the mirror reflection

images rather than implementing SIFT twice for every query image. And the symmetry axis of a planar symmetric object can be easily detected using MIFT with RANSAC.

Acknowledgements

This work was supported by National Natural Science Foundation of China (No. 50735003), Tianjin University 985 research fund, and State Key Laboratory of Precision Measuring Technology and Instruments open fund.

References

1. Mikolajczyk, K., Schmid, C.: Indexing based on scale invariant interest points. In: ICCV, vol. 1, pp. 525–531 (2001)
2. Kleban, J., Xie, X., Ma, W.: Spatial pyramid mining for logo detection in natural scenes. In: ICME, pp. 1077–1080 (2008)
3. Jegou, H., Douze, M., Schmid, C.: Hamming embedding and weak geometric consistency for large scale image search. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part I. LNCS, vol. 5302, pp. 304–317. Springer, Heidelberg (2008)
4. Brown, M., Lowe, D.: Recognising panoramas. In: ICCV (2003)
5. Harris, C., Stephens, M.J.: A combined corner and edge detector. In: Alvey Vision Conference, vol. 20, pp. 147–152 (1988)
6. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. In: IJCV, vol. 60, pp. 91–110 (2004)
7. Bay, H., Tuytelaars, T., Gool, L.V.: Surf: Speeded up robust features. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3951, pp. 404–417. Springer, Heidelberg (2006)
8. Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors. PAMI 27, 1651–1630 (2004)
9. Tola, E., Lepetit, V., Fua, P.: A fast local descriptor for dense matching. In: CVPR, pp. 1–8 (2008)
10. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: CVPR, vol. 1, pp. 886–893 (2005)
11. Scovanner, P., Ali, S., Shah, M.: A 3-dimensional sift descriptor and its application to action recognition. In: ACM International conference on Multimedia, pp. 357–360 (2007)
12. Ke, Y., Suktnankar, R.: Pca-sift: A more distictive representation for local image descriptors. In: CVPR, vol. 2, pp. 506–513 (2004)
13. Zhang, W., Kosecka, J.: Image based localization in urban environments. In: 3DPVT, pp. 33–40 (2006)
14. Hayfron-Acquah, J.B., Nixon, M.S., Carter, J.N.: Automatic gait recognition by symmetry analysis. In: Pattern Recognition Letters, vol. 24, pp. 2175–2183 (2003)
15. Choi, I., Chien, S.I.: A generalized symmetry transfer with selective attention capability for specific corner angels. IEEE Signal Processing Letters 11, 255–257 (2004)
16. Hartley, R.I., Zisserman, A.: Multiple View Geometry in Computer Vision. Cambridge University Press, Cambridge (2004)
17. Fischler, M.A., Bolles, R.C.: Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. Commun. Assoc. Comp. Mach. 24, 381–395 (1981)