Combining Invariant and Corner-Like Features to Optimize Image Matching

Jimmy Addison Lee and Kin-Choong Yow

School of Computer Engineering, Nanyang Technological University,
50 Nanyang Avenue, Singapore 639798
{jimm0002,kcyow}@ntu.edu.sg

Abstract. Significance and usefulness of local invariant features and traditional corner-like features have been widely proven in the literature. In this paper, we novelly combine the two types of features to select salient keypoints with the invariant and corner-like properties, which are highly distinctive and improving match performance. We use moment-derived complex image patterns (e.g., corner, T-junction, sectional cut, and chess-cross) to find corner-like features. We further optimize the matching results by finding corner-like patterns in the invariant matched point correspondences; and rebuff point correspondences that have dissimilar pattern responses which are most likely false matches.

Keywords: Keypoint extraction, salient keypoints, distinctive keypoints, corner-like patterns.

1 Introduction

The use of salient features, also known as keypoints or interest points, to find correspondences across multiple images is a key step in many image processing and computer vision applications. Some of the most notable examples are panorama stitching [1,4,5], wide baseline matching [2,8,10], image retrieval [12,22], object recognition [3,7,13], and object class recognition [14,16,17]. These salient features are landmarks in an image which are often intuitively palpable to humans. They include corners of buildings, edges of objects, features (e.g., eyes) on human faces, etc. The traditional salient features such as edges and corners have been significantly useful and applied to many problems including tracking. We use complex image patterns (e.g., predefined shapes, contour junctions, etc.) in this paper to detect corner-like features.

In recent years, there has been an increased interest within the content-based image retrieval community in finding new types of salient features (e.g., SIFT features [13]) which provide properties robust to changes in scale and/or affine transformations. Such invariant features have proven useful in the context of image registration and object recognition. In this paper, we novelly combine these invariant features with the traditional ones (corner-like features). This combination can be used to select salient keypoints which comprise of the invariant and corner-like properties. These keypoints are highly distinctive points so that

T. Wada, F. Huang, and S. Lin (Eds.): PSIVT 2009, LNCS 5414, pp. 692–701, 2009.

they can be easily distinguished from other similarly extracted points in the same or another image, and therefore improving match performance. This will be discussed in Section 3.1.

In additional, after matching the invariant keypoints, we further improve the matching results by finding corner-like patterns in the matched invariant point correspondences; and rebuff point correspondences that have dissimilar pattern responses which are most likely false matches. This will be discussed in Section 3.2.

2 Related Work

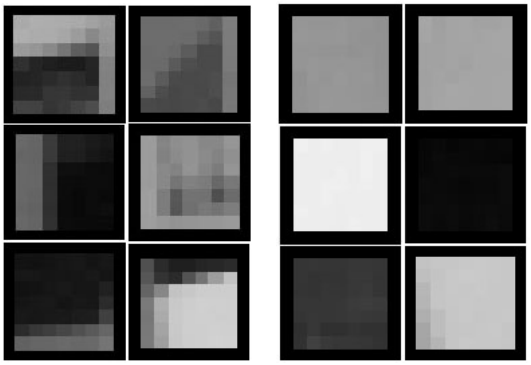
Many different keypoint detectors have been proposed with a wide range of definitions for what points in an image are interesting. Some detectors find points of high local symmetry, some find areas of highly varying texture, while others locate corner points. Corner points are interesting as they are formed from two or more edges and edges usually define the boundary between two different objects or parts of the same object. The earlier work on corner detectors can be traced back to the work of Moravec [18] used for stereo matching. It was then further improved by Harris and Stephens [9] to make it more repeatable under small image variations and near edges. While these detectors are called corner detectors, they are not selecting just corners, but rather any image location that has large gradients in all directions at a predetermined scale.

Complex image patterns, which also include corners and other predefined shapes, are informative as they are infrequent and provide rich description of images. In medical imaging, there is a tendency of replacing the traditional two-step object recognition (i.e. segmentation followed by shape identification) by methods directly extracting predefined objects from grey-level images (e.g., active contours [11,21]). Similarly, over the past years there has been considerable attention directed toward the detection of more complex contour features in raw-data images (e.g., [6,19]).

Recently, there has been impressive body of work on invariant local features which have been shown to be rather robust with respect to changes in scale and/or affine transformations (e.g., [2,13,15]). Lowe's Scale Invariant Feature Transform (SIFT) [13] is one of the well-known ones which looks promising for tracking applications. However the vast numbers of detected keypoints for matching at times can be rather time-consuming. There are also existence of false matches on occasion.

3 Combining Invariant and Corner-Like Features

In this paper, we demonstrate the combination of SIFT features (invarient features) and complex image patterns (corner-like features) for keypoint extraction and matching, which will be discussed in the following sub-sections (3.1 and 3.2). Complex image patterns will be discussed in sub-section 3.3.



(a) (b)

Fig. 1. Some invariant samples from SIFT. Some comprise of corner-like patterns as shown in (a), or no corner-like patterns as shown in (b).

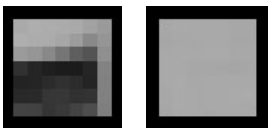
3.1 Keypoint Extraction

SIFT uses Difference of Gaussian (DOG) Extrema detector [13] to detect keypoints which are invariant to scale change. However due to the large numbers of extrema, there could be up to thousands of detected keypoints in an image. Each of these keypoints after sampled to a 15×15 square window, may comprise of some corner-like patterns (as shown in Figure 1(a)), or no corner-like patterns (as shown in Figure 1(b)).

Subsequently, each of the samples is computed against the complex image patterns (e.g., corner and T-junction, discuss in sub-section 3.3) to ascertain that the sampled patch comprises at least one of these corner-like patterns. If it does not, it will be eradicated. Figure 2 illustrates this, which the two samples are tested for any corner-like patterns. We can see from the response results that the sample in Figure 2(a) has some responses from the complex image patterns detection while the sample in Figure 2(b) does not. Sample in Figure 2(a) is more distinctive in this case, which we know is apparently better for matching.

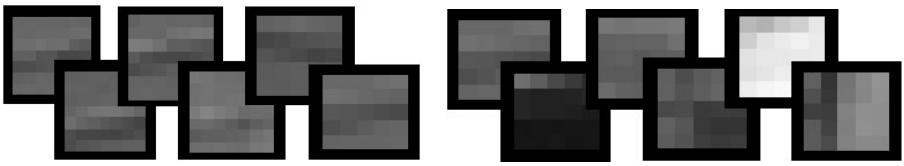
3.2 Matching

The matching is done through an Euclidean-distance based nearest neighbor approach. To increase robustness, matches are rejected for those keypoints for



(a) (b)

Fig. 2. Two invariant samples from SIFT. Complex image response for sample (a): Corner: 0.2218; T-junction: 0.5413, and Complex image response for sample (b): Corner: 0; T-junction: 0. Sample **(b)** does not comprise any corner-like patterns.



(a) (b)

Fig. 3. Example of sample group in (a) with small pattern response differences. Example of sample group in (b) with huge pattern response differences.

which the ratio of the nearest neighbor distance to the second nearest neighbor distance is greater than some threshold τ_r; or the other way round, a pair of keypoints is considered a match if the distance ratio between the nearest neighbor distance and a second nearest neighbor distance is below τ_r:

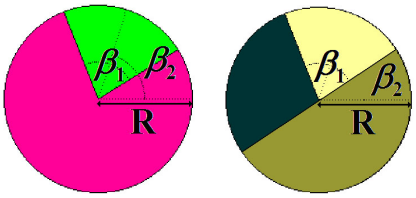
$$\frac{d^2(f, f_{1st})}{d^2(f, f_{2nd})} < \tau_r^2, \tag{1}$$

where $f \in \Re^n$ is the descriptor to be matched and f_{1st} and f_{2nd} are the nearest and the second nearest descriptors respectively, with d denoting the Euclidean distance between two descriptors. The threshold $\tau_r = 0.8$ suggested in [13] was found effective for general object recognition.

In additional, while SIFT provides invariant matched pairs, we scrutinize the difference in the responses of each pair of good distance match with complex image patterns. This can be done by sampling pixels within a 15 x 15 square window around the pair of point correspondences. If the difference in their pattern responses is huge (above a threshold of 0.5), it is very likely that this pair of point correspondences is a false match, and thus it will be eradicated. Figure 3 gives an example of two sample groups with similar and dissimilar patterns. The differences in the pattern responses of the group in Figure 3(a) (with similar pattern) are small, while the differences in the pattern responses of the group in Figure 3(b) (with dissimilar pattern) are huge.

3.3 Complex Image Patterns

Our work was inspired by moment-derived patterns [20] and we use them to find the pattern approximations of circular patches around keypoints. Corners and corner-like patterns (e.g., junctions) are predominantly significant as they generally preserve their geometry over a wide range of radii of circular patches. Thus, in this paper we demonstrated on four corner-like patterns which are commonly found in building images, i.e., proper corner, T-junction, sectional cut, and chess-cross. Two of them (corner and T-junction) will be discussed. The concept is identical for the rests. The model configuration of a corner over a circle of radius R is defined by two angles and two intensities (or colors) as illustrated in Figure 4(a). Similarly, the model configuration of a T-junction consists of two angles and three intensities (or colors) is illustrated in Figure 4(b).



(a) Corner (b) T-junction

Fig. 4. Model configurations of a corner (a) and a T-junction (b)

Corner Model. Given any circular image window of radius R, the parameters of its optimum corner approximation can be found from moment-based expression specified in [20]. The orientation angle β_2 in Figure 4(a) is extracted using

$$\beta_2 = \arctan 2(\pm m_{01}, \pm m_{10}) \tag{2}$$

while the angular width β_1 is computed from

$$\beta_1 = 2 \arcsin \sqrt{1 - \frac{16[(m_{20} - m_{02})^2 + 4m_{11}^2]}{9R^2(m_{10}^2 + m_{01}^2)}} \tag{3}$$

or

$$2 \arccos \frac{4}{3} \sqrt{\frac{(m_{20} - m_{02})^2 + 4m_{11}^2}{R^2(m_1 0^2 + m_{01}^2)}}. \tag{4}$$

The intensities A_1 (corner) and A_2 (background) are found using

$$A_1 = \frac{m_{00}}{\pi R^2} + \frac{3(2\pi - \beta_1)(m_{10} \cos \beta_2 + m_{01} \sin \beta_2)}{4\pi R^3 \sin 0.5\beta_1} \tag{5}$$

and

$$A_2 = \frac{m_{00}}{\pi R^2} - \frac{3\beta_1(m_{10} \cos \beta_2 + m_{01} \sin \beta_2)}{4\pi R^3 \sin 0.5\beta_1}. \tag{6}$$

T-junction Model. For T-junction in Figure 4(b), β_1 angular width and β_2 orientation angle can be computed using

$$\frac{\pi}{2} - \beta_2 - \frac{\beta_1}{2} = \frac{\arctan 2(\pm m_{02} \mp m_{20}, \pm 2m_{11})}{2} \tag{7}$$

and

$$m_{01} \cos \beta_2 - m_{10} \sin \beta_2 = \pm \frac{4}{3R} \sqrt{(m_{20} - m_{02})^2 + 4m_{11}^2} \tag{8}$$

respectively. The intensities are found as solutions of the following system of linear equations

$$\begin{cases} \frac{2m_{00}}{R^2} = A_1\pi + A_2\beta_1 + A_3(\pi - \beta_1) \\ \frac{3m_{10}}{R^3} = -2A_1c_2 + A_2(c_2 - c_{2-1}) + A_3(c_2 + c_{2-1} - 2s_2) \\ \frac{3m_{01}}{R^3} = -2A_1s_2 + A_2(s_2 - s_{2-1}) + A_3(s_2 + s_{2-1} + 2c_2) \end{cases} \tag{9}$$

where c_x and s_x indicate cos and sin functions of the corresponding arguments. Simple calculations can prove that results produced by Equations (2-4) and (7-8) are invariant to linear illumination changes and the angular width β_1 is invariant under any similarity transformation. Extensive experiments have also attested that the results are stable under both high and low frequency noise, image texturization and partial over and under saturation of image intensities.

The method is applicable to color images as well. In fact, it may be even more flexible since moments of color images are 3-dimensional vectors (R, G and B components) instead of scalars. Thus, the approximation equations for color patterns would be modified correspondingly. If scalar moments can be directly replaced by moment vectors, the gray-level solutions remain basically unchanged. For example, equation 3 is converted into

$$\beta_1 = 2 \arcsin \sqrt{1 - \frac{16\left(\|\overrightarrow{m_{20}} - \overrightarrow{m_{02}}\|^2 + 4\|m_{11}\|^2\right)}{9R^2\left(\|m_{10}\|^2 + \|m_{01}\|^2\right)}}. \tag{10}$$

If the direct replacement of scalars by vectors is not straightforwardly possible (e.g., equation 8), the scalar moments would be replaced by the largest components of vector-moments. Subsequently, equation 8 would be replaced by

$$\begin{aligned} & m_{01}(Z)\cos\beta_2 - m_{10}(Z)\sin\beta_2 \\ &= \pm\frac{4}{3R}\sqrt{\|\overrightarrow{m_{20}} - \overrightarrow{m_{02}}\|^2 + 4\|m_{11}\|^2} \end{aligned} \tag{11}$$

where Z could be R, G or B, depending for which color the value of $|m_{10}(Z)| + |m_{01}(Z)|$ is the largest. Colors of the approximations are calculated identically. We just apply equations 5, 6 and 9 separately to R, G and B colors, using the moments of the corresponding color.

4 Experimental Results

We have done experiments with images taken within the campus. As explained in Section 3.1, distinctive keypoints are found by selecting only those that comprise

(a) (b)

Fig. 5. In (a), 1326 invariant keypoints are detected from SIFT. Some of the keypoints without any corner-like patterns are circled in red. In (b), 1008 keypoints remain after selecting only those that comprise some corner-like patterns.

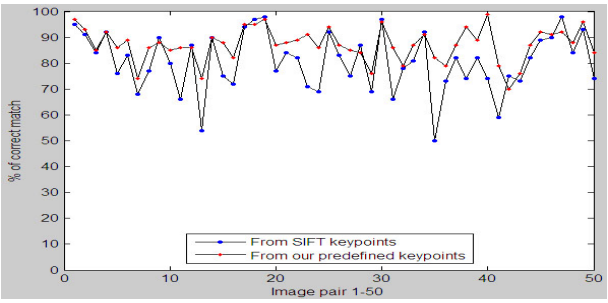


Fig. 6. Graph shows the percentage of correct match in 50 pairs of images with SIFT keypoints and predefined keypoints. Predefined keypoints improved matching results.

(a) 102 corresponding matches, with 11 false matches.

(b) 83 corresponding matches remain after eradicating those with dissimilar pattern responses. All false matches are being eradicated.

Fig. 7. Complex image patterns used to scrutinize the difference in responses of point correspondences in two images. Correspondences with dissimilar pattern responses are being eradicated in (b).

of the invariant and corner-like properties. This is illustrated in Figure 5. We can see that those keypoints that do not comprise any of the corner-like patterns are usually those that are rather homogeneous (circled in red in Figure 5(a)).

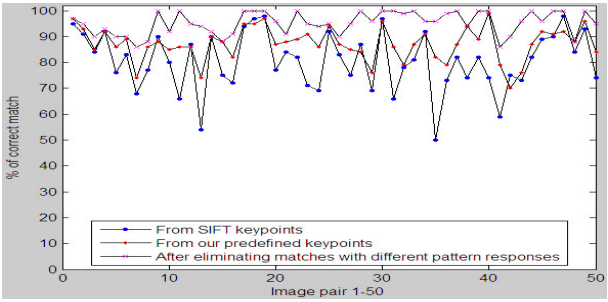


Fig. 8. Graph shows the percentage of correct match in 50 pairs of images. The matching results are further improved after eliminating matches/point correspondences that have different pattern responses.

These more distinctive keypoints can improve the match performance. For experiment, we took 51 images of similar campus scenes for matching. 1 image was taken as the main image while the rest of the 50 images were matched against it. The scenes were similar with some adjustments on the camera position at each shot, giving certain viewpoint and scale differences. Figure 6 displays a graph of the matching results from the 50 pairs of images. Most of the matchings were improved with the predefined keypoints.

Subsequently, the invariant keypoints were matched as described in Section 3.2. While SIFT gave invariant match, we scrutinized the difference in the pattern or corner-like responses of each pair of point correspondences. If the point correspondences were true match, the responses on the corner, T-junction, sectional cut, and chess-cross patterns would be similar. Figure 7 illustrates an example and we can see that some or most of the false matches (with dissimilar pattern responses) were eradicated as shown in Figure 7(b).

Figure 8 displays a graph of the matching results from the 50 pairs of images. As shown, the matching results have been further improved after eliminating matches/point correspondences with different pattern responses.

5 Conclusions

We have novelly combined the local invariant features (SIFT features in this paper) with the more traditional corner-like features for selecting salient keypoints. These keypoints were usually more distinctive and have shown to significantly improve on the match performance. The corner-like features were detected using moment-derived complex image patterns (i.e., proper corner, T-junction, sectional cut, and chess-cross).

While SIFT has provided invariant matched pairs of point correspondences, we further scrutinized the differences in the pattern or corner-like responses of each pair. Those pairs of point correspondences with dissimilar pattern responses which were most likely false matches have been eradicated. Through experimental results, we perceive the effectiveness of the approach which the matching results have been very much optimized.

References

1. Agarwala, A., Agrawala, M., Cohen, M., Salesin, D., Szeliski, R.: Photographing long scenes with multi-viewpoint panoramas. In: Proceedings of the 33rd Internationl Conference and Exhibition on Computer Graphics and Interactive Techniques (SIGGRAPH 2006), Boston, Massachusetts, USA, vol. 25, pp. 853–861 (2006)
2. Baumberg, A.: Reliable feature matching across widely separated views. In: IEEE Computer Society International Conference on Computer Vision and Pattern Recognition (CVPR 2000), Hilton Head Island, South Carolina, USA, vol. 1, pp. 774–781 (2000)
3. Belongie, S., Malik, J., Puzicha, J.: Shape matching and object recognition using shape contexts. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI 2002), 509–522 (2002)
4. Brown, M., Lowe, D.G.: Recognising panoramas. In: Proceedings of the 9th IEEE International Conference on Computer Vision (ICCV 2003), Nice, France, vol. 2, pp. 1218–1225 (2003)
5. Brown, M., Lowe, D.G.: Automatic panoramic image stitching using invariant features. International Journal of Computer Vision (IJCV 2007), 59–73 (2007)
6. Deriche, R., Giraudon, G.: A computation approach for corner and vertex detection. International Journal of Computer Vision (IJCV 1993), 101–124 (1993)
7. Frome, A., Huber, D., Kolluri, R., Bulow, T., Malik, J.: Recognizing objects in range data using regional point descriptors. In: Pajdla, T., Matas, J(G.) (eds.) ECCV 2004. LNCS, vol. 3023, pp. 224–237. Springer, Heidelberg (2004)
8. Goedeme, T., Tuytelaars, T., Van-Gool, L.: Fast Wide Baseline Matching for Visual Navigation. In: IEEE Computer Society International Conference on Computer Vision and Pattern Recognition (CVPR 2004), Washington, DC, vol. 1, pp. 24–29 (2004)
9. Harris, C., Stephens, M.: A combined corner and edge detector. In: 4th Alvey Vision Conference (AVC 1988), Manchester, UK, pp. 147–151 (1988)
10. Kannala, J., Brandt, S.S.: Quasi-dense wide baseline matching using match propagation. In: IEEE Computer Society International Conference on Computer Vision and Pattern Recognition (CVPR 2007), Minneapolis, Minnesota, USA, pp. 1–8 (2007)
11. Kass, M., Witkin, A., Terzopoulos, D.: Snakes: Active contour models. International Journal of Computer Vision (IJCV 1988), 321–331 (1988)
12. Katare, A., Mitra, S.K., Banerjee, A.: Content based image retrieval system for multi object images using combined features. In: Proceedings of the 17th International Conference on Computer Theory and Applications (ICCTA 2007), Alexandria, Egypt, pp. 595–599 (2007)
13. Lowe, D.G.: Distinctive Image Features from Scale-Invariant Keypoints. International Journal of Computer Vision (IJCV 2004) 60(2), 91–110 (2004)
14. Leordeanu, M., Hebert, M., Sukthankar, R.: Beyond local appearance: Category recognition from pairwise interactions of simple features. In: IEEE Computer Society International Conference on Computer Vision and Pattern Recognition (CVPR 2007), Minneapolis, Minnesota, USA (2007)
15. Mikolajczyk, K., Schmid, C.: An affine invariant interest point detector. In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (eds.) ECCV 2002. LNCS, vol. 2350, pp. 128–142. Springer, Heidelberg (2002)
16. Mikolajczyk, K., Leibe, B., Schiele, B.: Multiple object class detection with a generative model. In: IEEE Computer Society International Conference on Computer Vision and Pattern Recognition (CVPR 2006), New York, USA, pp. 26–36 (2006)

17. Mutch, J., Lowe, D.G.: Multiclass object recognition with sparse, localized features. In: IEEE Computer Society International Conference on Computer Vision and Pattern Recognition, New York, USA, vol. 1, pp. 11–18 (2006)
18. Moravec, H.: Rover visual obstacle avoidance. In: Proceedings of the 7th International Joint Conference on Artificial Intelligence (IJCAI 1981), Vancouver, Canada, pp. 785–790 (1981)
19. Ruzon, M.A., Tomasi, C.: Edge, junction, and corner detection using color distributions. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI 2001), 1281–1295 (2001)
20. Sluzek, A.: On moments-based local operators for detecting image patterns. IVC 23, 287–298 (2005)
21. Staib, L.H., Duncan, J.S.: Boundary fitting with parametrically deformable models. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI 1992), 1061–1075 (1992)
22. Wang, J., Zha, H., Cipolla, R.: Combining interest points and edges for content-based image retrieval. In: Proceedings of the 12th International Conference on Image Processing (ICIP 2005), Genova, Italy, vol. III, pp. 1256–1259 (2005)