

Improved SIFT Matching for Image Pairs with a Scale Difference

Y. Bastanlar, A. Temizel and Y. Yardımcı

Informatics Institute, Middle East Technical University, Ankara, 06531, Turkey

Published in IET Electronics, 4 March 2010, Volume 46, Issue 5, p. 346-348.

Published article can be reached at <http://dx.doi.org/10.1049/el.2010.2548>

Article History:

Submitted, 19 September 2009.

Accepted for publication, 5 February 2010.

Available online, 4 March 2010.

Abstract: When directly applied on images with different scales, SIFT matching performance decreases significantly. In this letter, this phenomenon is demonstrated and a simple method to increase the performance of SIFT matching is proposed. The proposed method includes preprocessing the images before matching and is compared to the previously proposed solutions which only eliminate the false matches.

Introduction: Scale Invariant Feature Transform (SIFT) [1] is considered as one of the best feature matching methods when robustness against rotation, scaling, illumination change and increased camera baseline is considered [2]. SIFT detects features in the so-called *scale space* comprising levels and octaves which are obtained by low-pass filtering and down-sampling the original image systematically. This enables the detection of features at different scales. In the case of matching points between images of different scales, we observed that a considerable number of false matches occur due

to matching finer scale features in the high-resolution image to the features in the corresponding image on a lower resolution.

To illustrate the problem we examine the two images in Fig. 1 which exhibit scale difference in addition to the change in camera viewpoint. Table 1 shows the number of extracted features in different octaves for these images. There is an approximate ratio of 2 between scales of correct correspondences. Corresponding octaves of correct matches are indicated in the table with same gray-level tones. SIFT extracts many features (~ 1600) at the first octave of the high-resolution image. It is quite likely that some of the candidates from this octave are incorrectly selected as the best match of features in the low-resolution image. For the given image pair, there are 23 false matches out of 84 and 17 of these false matches have a scale ratio less than 1.5 or greater than 3.0.

Previously proposed solutions: This phenomenon was observed and exploited to eliminate false matches in two independent studies [3,4]. Both of these methods use an eliminate-after-matching approach, hence they aim to reduce the false matches in the SIFT output. We refer them as *scale restriction* methods in the remainder of this letter. Yi *et al.* [3] form the histogram of scale differences and define a window around the peak of this histogram. The matches with scale differences outside this window are rejected. A limitation in their study is that only the image pairs with approximately the same scale are considered. Alhwarin *et al.* [4] divide the SIFT features according to the octaves they are extracted from. They detect the octave pair which yields

maximum number of matches and they assign the ratio between these octaves as the correct scale factor. All matches from other octave pairs are rejected. Since only the matches between octaves are analyzed, the scale ratio can be obtained only in the form of 2^k and ratios in between are not considered.

Proposed method: Our method also uses the dominant scale ratio between the images but it is not an eliminate-after-matching method. We preprocess the image pairs to adjust their scales and observe a significant improvement in SIFT matching.

Preprocessing is performed by low-pass filtering and down-sampling the high-resolution image. The down-sampling factor is extracted from the scale ratio histogram because the SIFT scale space ratio also reveals the scale ratio of the features in the images. To avoid aliasing, we need to low-pass filter the perspective image before down-sampling. We selected the cut-off frequency as $1/\sigma$ in the frequency domain and the standard deviation of the Gaussian filter becomes $\sigma = d/\pi$ where d is the down-sampling factor.

The advantage of the proposed method, when compared to the *scale restriction* approaches, is that not only false matches are eliminated but also number of correct matches is increased since the candidates from the incorrect octaves are eliminated in the first place. Last column of Table 2 shows the number of detected features where the high-resolution image is

preprocessed. For the given image pair, when the scale restriction results in a false/total match ratio of 6/65. With the proposed method, it is 5/81.

False match elimination using the geometric constraints is possible such as the one proposed by Lowe [1] which assumes an affine transformation between images. However, this should be considered as a last step of all matching methods since it eliminates 'geometrically' false matches regardless of the matching method. We exclude this step from our experiments.

Experimental comparison: We compare our method with both SIFT matching and scale restriction approaches. The latter is abbreviated as SR-SIFT and the steps are summarized as follows:

- 1) Apply SIFT matching on raw images and plot the histogram of scale ratios.
- 2) Extract the correct scale ratio (d) from the histogram as the mean of the most dominant Gaussian in the mixture.
- 3) Accept only the matches with a scale ratio between $0.6d$ and $1.4d$.

Please note that SR-SIFT explained here is the improved version of the method proposed in [3] since it employs scale ratio rather than scale difference. This maintains the effectiveness of the algorithm for images with larger scale differences to account for the increasing variance of the histogram. Also, the modified method is not limited to 2^k scale ratios as opposed to the method in [4].

Our method basically adds a preprocessing and a SIFT matching step after the 2nd step of SR-SIFT:

- 1) Low-pass filter the high-resolution image with a Gaussian filter with $\sigma = d/\pi$ and down-sample it by d both in horizontal and vertical directions.
- 2) Apply SIFT matching on the preprocessed images and plot the histogram of scale ratios. Now the scale ratio is close to one. A final elimination is applied by re-detecting d and employing the 3rd step of SR-SIFT.

Our motivation for this study was automatic feature matching between omnidirectional and perspective camera images, i.e. hybrid image pairs such as the one given in Fig. 2. Due to the different imaging geometry of omnidirectional cameras, appearance of objects varies between these images when compared to perspective-to-perspective matching. However, we observed that its effect is less significant and main causes of the degraded performance are wide baseline and scale difference. With the proposed approach, scale problem is eliminated and automatic point matching between hybrid camera images became possible for tolerable baseline lengths. For hybrid pairs, we preprocess the perspective image in the pair and employ the low-pass filtering parameter as $\sigma = 2d/\pi$. We obtained the best results with this parameterization but we also observed that slight variations in the selected parameters (σ , d) do not severely affect the results.

Results: To compare the three approaches (SIFT, SR-SIFT and proposed SIFT after preprocessing) we performed tests on a total of 30 image pairs viewing four different scenes (indoor and outdoor) which were taken with five different cameras including omnidirectional (both catadioptric and fisheye) and regular perspective cameras. Scale ratios between the images vary from 1.5

to 4.2. We adjusted the SIFT threshold (distance to the closest candidate feature / distance to the next closest feature) to obtain the same number of matches for all three approaches (changing from 50 to 100 from pair to pair) to be able to compare the number of correct matches.

In the original SIFT algorithm, it is possible for many points in the first image to be matched to the same point in the second image. This causes inconsistencies depending on which image is defined as "first". To eliminate this problem, we run the SIFT algorithm in both ways changing the order of images and declare a match only if it is found in both runs.

Table 2 shows the average FP rate ($\#$ false positives / $\#$ detected matches) and TP ($\#$ true-positives) for perspective-to-perspective matching and omnidirectional-to-perspective matching. FP rate is very low for both SR-SIFT and our method compared to directly applying SIFT. When the number of true-positives are considered, our method outperforms SIFT and SR-SIFT.

The proposed method is more successful for wide-baseline perspective image pairs as well as hybrid image pairs as can be observed from Table 2. In both cases, feature descriptors are not very close to their true matches in the corresponding image due to the different imaging geometry and increasing baseline. Thus, the ratio of distances to the first and second candidate matches (SIFT threshold) decreases resulting in decreased number of true matches and a greater number of features are distracted by false correspondences. For perspective pairs with short baseline, performances of

SIFT and SR-SIFT increase, however our approach is still able to increase the number of correct matches.

Conclusion:

The proposed algorithm has been experimented with images acquired with different perspective and hybrid camera pairs and has been shown to outperform SIFT and SR-SIFT. Different to the scale restriction approach which is able to eliminate a proportion of the false matches, the proposed algorithm increases the number of correct matches while eliminating the false matches. Improvement is especially significant for wide-baseline perspective image pairs and hybrid camera pairs.

References

- 1 LOWE, D.: 'Distinctive image features from scale invariant keypoints', International Journal of Computer Vision, 2004, 60, pp.91-110.
- 2 MIKOLAJCZYK, K., and SCHMID, C.: 'A performance evaluation of local descriptors', IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27(10), pp.1615-1630.
- 3 YI, Z., ZHIGUO, C., and YANG, X.: 'Multi-spectral remote image registration based on SIFT', Electronic Letters, 2008, 44(2), pp.107-108.
- 4 ALHWARIN, F., WANG, C., RISTIC-DURRANT, D., and GRASER, A.: 'Improved SIFT-features matching for object recognition', Visions of Computer Science - BCS International Academic Conference, 2008.

Authors' affiliations:

Y. Bastanlar, A. Temizel*, Y. Yardimci (Informatics Institute, Middle East Technical University, Ankara, 06531, Turkey)

E-mail address: atemizel@ii.metu.edu.tr

Tables

Table 1 Number of SIFT features detected in the images shown in Fig. 1. The boxes marked with the same gray level are the corresponding scales.

SIFT Octave	Approximate scale in SIFT scale space	Image in Fig.1a	Image in Fig.1b	Image in Fig.1a, low-pass filtered and down-sampled
-1	1	1599	1408	460
0	2	306	277	116
1	4	121	106	34
2	8	34	36	9
3	16	10	5	0
4	32	1	1	0

Table 2 Average values of FP rate (# false positives / # detected matches) and TP (# true-positives) for perspective and hybrid (omnidirectional with perspective) matching experiments.

	Perspective (Short-Baseline)		Perspective (Wide-Baseline)		Hybrid	
	FP rate	TP	FP rate	TP	FP rate	TP
SIFT	0.218	59.2	0.295	53.2	0.336	49.4
SR-SIFT	0.040	58.2	0.136	51.6	0.092	46.6
Proposed method	0.062	66.0	0.101	61.4	0.074	62.5

Figures

Fig. 1 Example image pair with an approximate scale ratio (700x760 pixels each)



(a)

(b)

Fig. 2 Example hybrid image pair

