Improving SIFT Matching by Interest Points Filtering

Rabih Younes
Virginia Tech
Blacksburg, VA
ryounes@vt.edu

Abstract

In this work, the goal is to minimize the number of outliers detected by interest point detectors. This approach presents many benefits like the fact that better homography matrices could be obtained and used for image stitching, RANSAC would take much less time to converge, better results would be obtained from the generalized Hough transform, etc. Four methods to reduce outliers were proposed and tested using the SIFT descriptor on a dataset of 1000 images containing image pairs subjected to different types of transformations and illumination changes. Three out of the four methods yielded good results and were combined in order to give the best result which was a 26.69% improvement in the true positive rate.

1. Introduction

In interest point matching applications such as image stitching, the best way to choose interest points in to do it manually, but most of the time a fast automated algorithm is needed and therefore an automated interest point detector is used. When detecting interest points using interest point detectors, a significant amount of outliers will be also detected. Depending on the textures, objects, and scenes present in the image, the amount of outliers can vary and it can lead to misleading results. For this reason, other techniques are used after the matching operation is performed in order to separate inliers from outliers such as RANSAC [3] and generalized Hough transform [4]. These techniques could take a significant amount of time and can slow the process by orders of magnitude, and they might not even converge if the outliers form a large portion of the detected interest points.

In this work, the detected interest points will be filtered in order to minimize the number of outliers using several methods which will be combined at the end to give the best results. Section 2 will discuss the related work. Section 3 will present the dataset used to test the approach. Section 4 will explain the developed approach. The obtained results are shown in section 5. And conclusions and proposed future work are presented in section 6.

2. Related work

In Mikolajczyk and Schmid work [1], the performance of the most popular local descriptors was evaluated with different interest point detectors and using images subjected to different types of changes and transformations. The conclusion was that SIFT [2] is the best point descriptor; it ranked first in all situations, except for the case when the images had illumination changes where steerable filters performed a little better that SIFT and presented the best results. Another important observation from [1] was that the choice of the interest point detector had no impact on the ranking of the performance of the point descriptors.

Since SIFT was proved to be the best interest point descriptor, some previous works tried improving SIFT performance by minimizing the feature vector in order to make it run faster, others tried to change some parts of the algorithm in order to make it a better fit for a certain application, and some studies tried improving the existing interest points detection algorithms [5][6]. This work will try to improve SIFT matching by removing outliers from the detected interest points or matches using many techniques that can take place before or after SIFT matching is performed.

3. Dataset

For this study, the same dataset used in [1] was used [7]. A sample of the dataset is shown in **Figure 1**. This dataset has 1000 images in total which contain image pairs representing an original image and another image which is the same as the original image but with some geometric or photometric transformations applied to it. These transformations are image rotation, scale changes, affine transformations and illumination changes. The transformations have been introduced by rotating the camera, varying the zoom and changing the viewpoint angle, while illumination is introduced by changing the brightness and the position of the light source. Planar scenes are used in order for homography to be used to verify the correct matches. Homography matrices for some of the image pairs are given in the dataset, and the rest of the homography matrices were computed using a tool which was developed. This tool shows the user a pair of images which does not have a homography matrix, and the user can click on both images specifying points that are the same in both of them (there should be at least 4 pairs of points), then the software computes the homography matrix based on these samples and stores it in the database. The software keeps showing image pairs to the user until homography matrices are obtained for all pairs.

Figure 1: Dataset sample [1].

4. Approach

In order to improve SIFT matching, 4 methods that are involved with interest points filtering were tried. At the end, the methods that gave good results were kept and combined to give even better results. Each method is performed on a pair of images in order to improve the SIFT matching between these two images, as shown in **Figure 2**. These methods will be explained in the following subsections. Note that the interest points detection (using Harris-Laplace interest point detection) code and the SIFT code of [8] were used.

4.1. Method 1: color spaces

In this first method, interest point detection is performed on two versions of each image of the image pair; one version is the grayscale image and the other version contains only the hue values of the HSV transformation of the image. After obtaining the interest points for both versions of the image, only interest points detected in both versions (within a 1 pixel distance) are kept and the rest are eliminated. After this process is performed on both images of the image pair, SIFT is used to describe the interest points' regions and then matching is performed based on the sum of squared distances score.

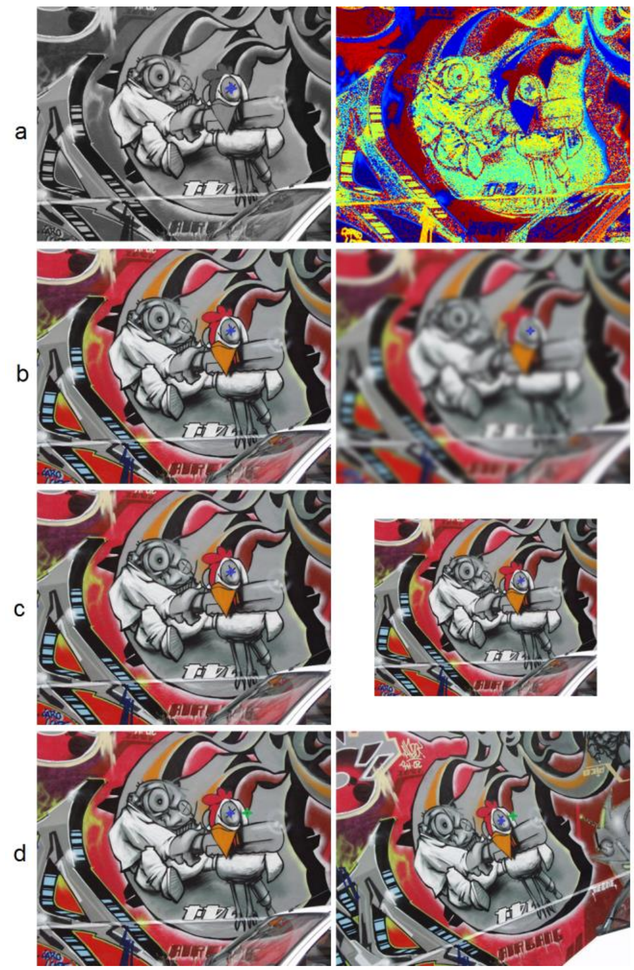

Figure 2: The blue cross in (a) shows an interest point that should be kept in this same image according to method 1. The blue cross in (b) shows an interest point that should be kept in this same image according to method 2. The blue cross in (c) shows an interest point that should be kept in this same image according to method 3. The blue cross in (d) shows 2 interest points that matched in the 2 images that should be kept since the examined points (green crosses) also matched according to method 4.

4.2. Method 2: blurring

In this method, interest point detection is performed on two versions of each image of the image pair; one version is the original image and the other version is a blurred version of the image. After obtaining the interest points for both versions of the image, the same process explained in section 4.1 is used to filter interest points and perform SIFT matching.

4.3. Method 3: resizing

For this method, interest point detection is performed on two versions of each image of the image pair; one version is the original image and the other version is a resized version of the image. After obtaining the interest points for both versions of the image, the same process explained in section 4.1 is used to filter interest points and perform SIFT matching.

4.4. Method 4: checking the match region

In this last method, instead of filtering the interest points in each image of the image pair before the SIFT matching is performed, filtering is done based on both images together after the SIFT matching takes place.

This method starts by detecting interest points in both images and performing SIFT matching based on these original matches. Then, for every match obtained, two points are examined which are near each interest point of the match. The region near each point of the points pairs is described using SIFT. If the difference between the sum of squared distances score of the two SIFT vectors is below a certain threshold, the match between the original interest points is kept, otherwise the match is discarded. The points that are chosen in the region of the original interest points are chosen using the following equations:

$$x_1' = x_1 + [m * sc_1 * \cos(or_1)] \quad \textit{(eq. 1)}$$
$$y_1' = y_1 + [m * sc_1 * \sin(or_1)] \quad \textit{(eq. 2)}$$
$$x_2' = x_2 + [m * sc_2 * \cos(or_2)] \quad \textit{(eq. 3)}$$
$$y_2' = y_2 + [m * sc_2 * \sin(or_2)] \quad \textit{(eq. 4)}$$

where (x_1,y_1) are the coordinates of the matched interest point in the first image, (x_2,y_2) are the coordinates of the matched interest point in the second image, (x_1',y_1') are the coordinates of the examined point in the first image, (x_2',y_2') are the coordinates of the examined point in the second image, sc_1 and or_1 are respectively the scale and orientation of the matched interest point region in the first image, sc_2 and or_2 are respectively the scale and orientation of the matched interest point region in the second image, and m is a scalar multiplier. Setting the value of m is discussed in section 5.

The choice of the examined interest points based on the equation above is due to the fact that if the matched interest points are truly the same then these examined points should be also the same if a flat surface is considered.

5. Results

Experiments were conducted on all 4 methods and parameters were tuned. 3 out of the 4 methods gave results that are better than the baseline and were combined later on in order to give the best results. The following sections present the evaluation criterion, the baseline, and the results for each method alone and for all 3 methods combined together.

5.1. Evaluation criterion and baseline

Many criteria can be used to evaluate the performance of SIFT matching such as the true positive rate, ROC curves, etc. Since there are a lot of parameters to tune in the used methods, it would be good to show how accurate the matching is while varying some parameters. For this reason, and after trying different representations, the evaluation criterion used to measure the performance of SIFT matching is the true positive rate versus the confidence region radius. The confidence region radius is the allowed range of the difference between pixels' positions in the transformed image based on the actual homography and the ones based on the computed homography.

To build the baseline curve, homography matrices for all image pairs were computed based on the used algorithm, then the transformed images obtained based on the computed homography matrices were compared to the ground truth output images which are obtained based on the given homography matrices.

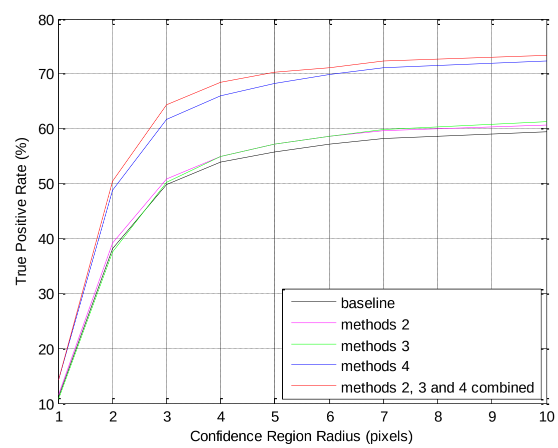

Figure 3: Results.

5.2. Method 1 results

Method 1 was tested and the results obtained were worse than using the grayscale version alone but better than using the hue values of the HSV space alone. This is due to the fact that there were different interest points detected in each of the two versions of each image; this is because the hue channel gives the image a different level of smoothness in its edges and creates some new sharper edges as seen in **Figure 2**a. Therefore when only the interest points in common were kept, a lot of these interest points were just random interest points which are outliers.

5.3. Method 2 results

For this method, different parameters needed to be tuned in order to find the best results. Some of these parameters were varied and it was found that using an average filter which has a size of 5x5 is the best filter to use for blurring in order to get the best results. The results of this method are shown in **Figure 3**.

Considering a confidence region radius between 1 and 10, the average improvement introduced by this method is: 2.48%.

5.4. Method 3 results

For method 3, one parameter needed to be tuned which is the resizing scale. This parameter was varied and it was found that using a scaling factor of 0.5 yields the best results. The results of this method are shown in **Figure 3**.

Considering a confidence region radius between 1 and 10, the average improvement introduced by this method is: 1.12%.

5.5. Method 4 results

For this last method, two parameters needed tuning which are the multiplier m and the threshold of the score. Since there are infinite combinations of these two parameters, only some combinations were tried and the best obtained results, shown in **Figure 3**, were for $m = 5$ and threshold = 4000.

Considering a confidence region radius between 1 and 10, the average improvement introduced by this method is: 23.84%.

5.6. Combined methods results

Since method 2, 3 and 4 yielded better results than the baseline, these 3 methods were the ones to be combined. Using the same parameters of each method mentioned earlier, the 3 methods were combined by first applying method 2 then method 3. Method 4 was performed after running SIFT and obtaining the candidate matches. The results of the combined methods are shown in **Figure 3**.

Considering a confidence region radius between 1 and 10, the average improvement introduced by these methods combined is: 26.69%. This result is better than any result obtained by using one of the 3 methods alone.

6. Conclusion and future work

Automated interest point detection saves a significant amount of time but its disadvantage is that a lot of incorrect interest points would be detected and consequently end up being matched incorrectly. In this study, it was proven that many techniques could be used in order to reduce the number of outliers. This leads to obtaining better homography matrices when matching points in several images for the same scene and trying to stich these images, it makes RANSAC converge much faster when used to obtain correct homography matrices, along with other benefits to many applications where interest point matching is needed.

This work could be further extended and results could be improved by integrating other methods to the approach or even optimizing the used methods. Method 1 could be optimized by trying different color channels which do not introduce strong edges. Method 2 could be optimized by trying some other custom filters with different sizes and distributions, since only the average and Gaussian filters were tried for some sizes of the filter matrix and with the default values of Matlab for the Gaussian distribution. Method 4 could be optimized much more since to values of the multiplier and the threshold are continuous and only a few combinations were tried in the wide range of available combinations.

References

[1] K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(10):1615-1630, 2005.

[2] D. Lowe. Object recognition from local scale-invariant features. *The Proceedings of the Seventh IEEE International Conference on Computer Vision*, vol. 2, pp. 1150-1157, 1999.

[3] M. Fischler and R. Bolles. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *Communications of the ACM*, 24(6):381–395, 1981.

[4] D. Ballard. Generalizing the Hough Transform to Detect Arbitrary Shapes. *Pattern Recognition,* 13(2):111-122, 1981.

[5] K. Liao, G. Liu and Y. Hui. An improvement to the SIFT descriptor for image representation and matching. *Pattern Recognition Letters,* 34(11):1211–1220, 2013.

[6] P. Du, Y. Zhou, Q. Xing, and X. Hu. Improved SIFT matching algorithm for 3D reconstruction from endoscopic images. *Proceedings of the 10th International Conference on Virtual Reality Continuum and Its Applications in Industry*, pp. 561-564, New York, 2011.

[7] Feature Detector Evaluation Sequences. Accessed in 2014. Retrieved from http://lear.inrialpes.fr/people/mikolajczyk/Database/index.html

[8] SIFT detector and descriptor. Accessed in 2014. Retrieved from http://www.vlfeat.org/overview/sift.html