Image Matching Using SIFT Algorithm And Pose - Illimination

Bhukya Shravan Nayak
M.Tech , Student,
VBIT College Ghatkesar Hyderabad India.

C.B.R. Lakshmi
M.Tech, Asst Prof,
VBIT College Ghatkesar Hyderabad India.

Abstract:

The challenges in local-feature-based image matching are variations of view and illumination. Many methods have been recently proposed to address these problems by using invariant feature detectors and distinctive descriptors. However, the matching performance is still unstable and inaccurate, particularly when large variation in view or illumination occurs.

In this paper, we propose a view and illumination invariant image-matching method. We iteratively estimate the relationship of the relative view and illumination of the images, transform the view of one image to the other, and normalize their illumination for accurate matching. Our method does not aim to increase the invariance of the detector but to improve the accuracy, stability, and reliability of the matching results. The performance of matching is significantly improved and is not affected by the changes of view and illumination in a valid range.

The proposed method would fail when the initial view and illumination method fails, which gives us a new sight to evaluate the traditional detectors. We propose two novel indicators for detector evaluation, namely, valid angle and valid illumination, which reflect the maximum allowable change in view and illumination, respectively. Extensive experimental results show that our method improves the traditional detector significantly, even in large variations, and the two indicators are much more distinctive.

Index Terms:

Feature detector evaluation, image matching, valid angle (VA), valid illumination (VI).

1. Introduction:

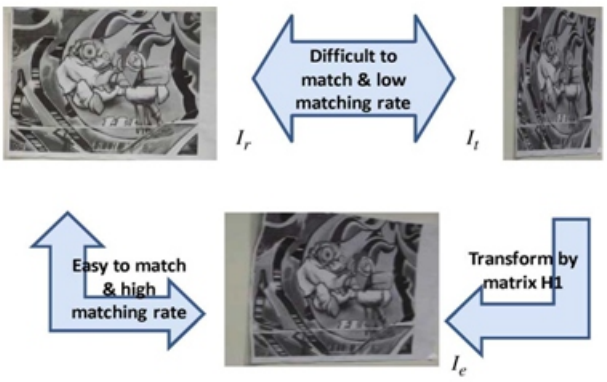

Fig. 1.Illustration of the proposed matching algorithm. I_r and I_t are the images to be matched. I_e is simulated from I_t by transformation T. I_r is difficult to match with I_t for the difference of view point and illumination, whereas I_e is easier to match with I_t since they are closer in the parameter space.to the changes of view and illumination.

The same interesting regions extracted from the matching images tend to be fewer and fewer when increasing the variation of view or illumination. For larger changes, there would be few invariant features that can be extracted from both images to be matched. This motivates us to think the essential difference of images with different view and illumination. Normally, a question need to be answered: whether an object in two images with different views and illumination looks like the same one, supposing there are two images with a large view change, as shown in Fig1. The two top images are the same object in different views.

They are so different in appearance that they can be considered as two different objects. We do not attempt to find invariant local feature detectors as in a previous work but focus on a better framework for image matching. Inspired by previous works [11], [19], [20] and the aforementioned perspective, we propose an iterative image-matching framework that iterates the estimation of pose and illumination to improve the matching performance. First, we transform the view and illumination of the image by estimating the pose and illumination correspondence between the matching pair by an initial detector, e.g., Harris [8], SIFT [11], SURF [12], and HLSIFD [14].

Then, we extract local features from the simulated image and match them with the features in another image. With this framework, the repeatability score (RS) and the number of correct matches (NCMs) could be stabilized under heavy variations in a valid range. Out of the valid range (larger view or illumination change), our method will fail to obtain correct matching result. We find that every feature detector under our framework has a considerable tolerance to the changes of view and illumination. When the initial estimation method, e.g., SIFT and SURF, fails, the proposed method also fails, which is a nature of the initial view and illumination estimation method.

2.Related work:

A. Image Matching With Local Features

The DDM framework is integrated in many systems. Brown and Lowe [1] create a system for fully automatic panorama stitching. SIFT is employed to detect local features from all images.

Then, they match the features and estimate the relationships, including location and rotation, for each connected component. Finally, multiband blending renders the panorama [2].Image stitching is easier than wide baseline matching since the main difference between the matching pair is the location and camera focus (scale).

Following the general evaluation, three criteria are often usedas feature evaluator.

1) NCMs are the number of total correct match pairs.

2) RS is the ratio between the NCM and the minimum of total number of features detected from the image pair RS NCM/TOTAL .

3) Matching precision (MP) is the radio between the NCM and the number of matches MP NCM/Matches NCM, RS, and MP are commonly used in the literature [23]–[26], [30], [31]. However, the meanings of these evaluators are not obvious.

The traditional evaluators cannot give intuitive comparison in choosing detectors according to the evaluation results. It is difficult to find which detector should be used because it not clear when the method would fail. To complement this blank, we propose two novel evaluators to evaluate some popular detectors in this paper.

3.View And Illumination Invariant Image Matching:

Proposed Method

Denote the reference image and test image to be matched as I_t and I_r. Suppose that the true pose transformation matrix from I_t to I_r is $\hat{H}$ and the illumination change function is $\hat{L}$. The relationship between I_t and I_r is

$$I_r(X) = \hat{T}(I_t) = \hat{L}(\hat{H}(I_t)) = \hat{L}(I_t(\hat{H}X)) \quad (1)$$

where is the true transformation between and , is thehomogeneous coordinates, and . If there existsapproximate estimations about illumination and transformation,the could be transformed to an estimated image , i.e.,

$$I(X) = T(I_t) = L(I_t(HX)) \quad (2)$$

where denotes the view point transformation and denotesthe illumination transformation. If is not a very rough estimationbetween and , the estimated image would be moresimilar to than itself. In other words, is closer to thanto . Thus, the matching between and will be easier, asshown in Fig. 1.

In this way, we propose the following iterative image-matchingprocess:

$$I_1 x = T_1(I_0) = L_1(I_0(H_1 x^T))\ (I_0 = I_t)$$

$$I_1 x = T_i(I_{i-1}) = L_i(I_{i-1}(H_i x^T))\ (i > 1) \quad (3)$$

Algorithm 1 *The proposed method*

Initial: $\{H_0, L_0\} = \{E, \vec{1}\}, T = T_0, \sigma_H, \sigma_L;$

Iterate

$i = i + 1;$

Estimate: $T_i : H_i, L_i;$

$T = T_i o T;$

$H = H_i * H.$

Transform I_{i-1} *to* I_i *by (3);*

Until $|H_i - E| < \sigma_H, |L_i - \vec{1}| < \sigma_L$ *or* $i > n$.

(E is the uint matrix, L_i *is a histogram transformation vector,*

σ_H *And* σ_L *are convergence thresholds.)*

Return *T, H*

The algorithm is summarized in Algorithm 1.

The final estimation of the $\hat{T}$ is

$$\hat{T} = ...oT_m oT_{m-1} oT_{m-2} o...oT_2 oT_1 \quad (4)$$

$$\approx T_n oT_{n-1} o...oT_2 oT_1 \quad (5)$$

Where " o" denotes function composition. Our experiments in Section IV-B show the convergence of the iteration with SIFT and the performance with respect to the number of iterations.

C. Estimate the Parameters and

General image-matching methods by local features focus on the first parameter since the concerned issue is the space correspondence between the two images. Illumination normalization could improve the performance of image matching because the images in the parameter space would be closer when the illuminations between them are similar.

One of the advantage of the proposed method is that it also estimates the illumination change, which makes matching much better when illumination has changed.

The purpose of general image-matching methods is to find the transformation matrix between the reference image and the test image. These methods are invariant to rotation, scale, and partially affine changes. The H can be easily estimated by the general methods without other information. First, we extract features from the matching images and obtain features descriptions (which method is used is not important).

Then, we match twofeatures when they are the nearest pair in the feature space. Here,norm is used to calculate the distance between to the features.

The RANSAC algorithm is employed to calculate transformation matrix . The general methods, i.e., HarAff, HesAff, SURF, SIFT, and HLSIFD, all can be used as the feature extraction method. We call them I-HarAff, I-HesAff, ISURF, ISIFT, and IHLSIFD ("I" indicates "Iterative"), respectively. Moreover, image matching is usually used in video sequences. We assume that the difference between two consecutive frames is not large, and the object or the camera smoothly moves. Thus, theth frame's transformation H_j can be approximated by theprevious results. Different detectors and descriptors , [1] have been developed to extract illumination invariant local features. The gradient direction histogram is normalized to form the descriptors. There is usually a tradeoff between the distinction and the invariance.

If we do not normalize the descriptors, they will be sensitive to illumination changes but more distinctive. Computing detectors and descriptors also cost much time. Conversely, thedetector will be more efficient if we do not require the detector to be invariant to illumination change. We want to keep both illumination invariant and descriptor distinctive in our method.

Thus, it is necessary to estimate the illumination change between the two images. Estimating the illumination is a challenging issue since the objects in the images are often accompanied by clutter background or noise. Benefitting from the estimation of the transformation matrix, we can warp the test image to another pose in which the object pose looks similar to that in the reference image.

Accordingly, approximate object segmentation would be obtained on the simulated image. To eliminate the occlusion, we only use the matched regions. The matchedregions are the region in the scale of the matched interesting points. First, we calculate the illumination histogram of the two images in the matched region. Second, we fix one image and calculate histogram translation function L from the other image to the fixed one. Suppose the histogram of the fixed image ish$_1$and the histogram of the other image is h$_2$. We calculate the cumulative functions of h$_1$ and h$_2$-h$_1$ and F$_2$. Finally, the translation function is

$$L = F_2^{-1} F_1. \quad (6)$$

Since the cumulative function of gray histogram is always monotonically increasing, inverse function always exists. We transform the histogram of the test image according to the histogram of the reference image to normalize the illumination between the pair.To sum up, we estimate transformation matrix between the matching pairs by feature detector, estimate the illumination relationship, and change one of the images according to the color histogram of the other to map the pose and illumination of the object in one image to the other.

D. Relationship Between the Iterative Algorithm and ASIFT

The proposed iterative method is similar to ASIFT [9], [2]. In ASIFT, the features are not invariant to affine change, but they cover the whole affine space, as shown in the middle blockin Fig. 4. Every simulation of the reference image is one pose of the image in the affine space. Therefore, parts of the simulations of the reference image and the test image should have similar poses in the affine space theoretically.

The simulations of the reference image and the test image are independently constructed. No mutual information is used in the simulations. Simulating in a high density in the affine space, many supposed image poses are constructed, and then, they are matched in a general way.

The number of matches increases with the number of the simulations. ASIFT indeed increases the invariability of the image-matching method. However, it does not care what the transformation matrix between the reference and test images is, by trying many possible transformations and combining the matches. Thus, ASIFT can be regarded as a sampling method

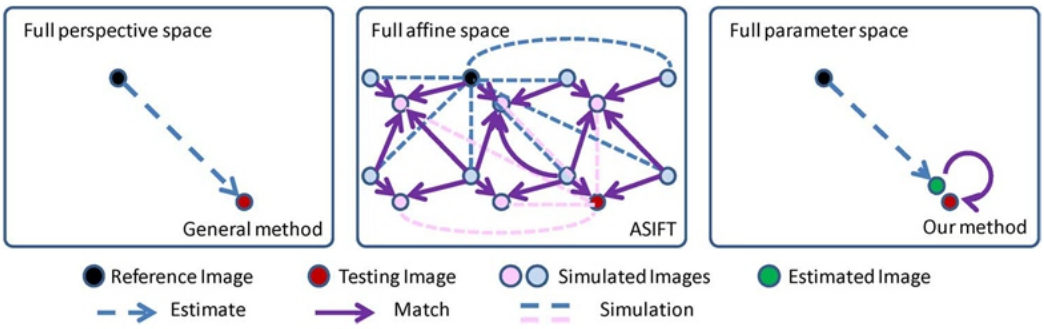

Fig. 2.Relationship among the general framework, ASIFT, and the proposed method. (Left block) The general framework, (middle block) ASIFT, and (right block) ours. The general DDM framework directly estimates the transformation between two images. It is simple but coarse. ASIFT simulates many poses of the two images to cover the affine space, whereas our method estimates the transformed pose first and then accurately matches in the projective space.

TABLE I:
Comparison of ASIFT and Our Method

*	ASIFT	Proposed
Simulation to reference image	√	×
Simulation to test image	√	√
Number of simulations	many	few
Number of features	$10^4 \approx 10^5$	10^3
Pose simulation	√	√
Illumination simulation	×	√
NCM	high	high
RS	Very low	high
Affine invariancy	Full	Partial
Computational cost	high	low
Real time	×	√

around the original points in parameter space , whose properties are shown in the left column of Table I.

Essentially, our method also constructs "simulation." We simulate the image not only in the pose but also in illumination, as shown in the right part of Fig.2. In addition, we transform one simulation per iteration, and in most tasks, two iterations are enough.

We will give an experiment to illustrate this, Benefiting from few simulations, the computational cost of our method is very low, compared with ASIFT, which simulates much more images than our method. A coarse-to-fine scheme can reduce the computational time of ASIFT to three times of the SIFT, whereas our method only costs two times. One drawback of the proposed method is that it does not increase the invariability of the original method. When the initial method fails in matching images, the proposed method also fails.

One promising method to overcome this shortage is to combine the proposed method with the ASIFT, which improves both the invariability and theaccuracy. Furthermore, the histogram matching may amplify noise that seems to affect the performance. A few more key points would be extracted after the histogram matching, but they would not affect the performance too much. We will show this in Section IV-C.

Experimental results show that the performance of the proposed framework reaches a comparable level, compared with ASIFT with much fewer features totally detected.Therefore, the RS of our method is much higher than that of ASIFT. The computational cost of our method is much

4.Experimental Results:

A. Database:

In the first experiment, we want to show the performance of the proposed method. We capture two images with changes both in illumination and view. This experiment is not used for comparison, but it only shows the effectiveness of the proposed method. To evaluate the performance of the proposed image-matching framework, we do experiments on the database provided by Mikolajczyk.1 This database contains eight groups of images with challenging transformations. We compare the proposed method with http://www.robots.ox.ac.uk/vgg/research/affine/ ASIFT and the usual DDM framework with the state-of-the-art detectors: HarAff, HesAff, SURF, SIFT, and HLSIFD.

In addition, two evaluations on the detectors through our strategy are proposed. One of them tests the adaptive capacity on the view change, and the other tests the capacity on the illumination change. To finish the two evaluations, we build two databases. One of them contains 88 frames with view changes from 0 to 87 . The other one contains 55 frames with light exposure changes from 40 to 14 (0.1 EV). The two databases contain continuous transformation frames. Thus, we can evaluate the view invariant ability of the detectors at a 1 interval and the illumination change invariant ability at a step of 0.1 EV. Such databases seldom appear in the open literature, and they will be currently available on the Internet [3].

B. Convergence:

As we mentioned in Section III-B, the number of iteration is an important parameter. A question that should be answered is whether more iterations bring better performance. Experiments show that, under the proposed framework, our method converges very fast. Fig. 6 shows an experiment on matching two images. The reference image is captured from a frontal view, and the test image is captured from a view angle of 60. Here, SIFT is used as the base detector. The RS andNCMof our method and theDDM.

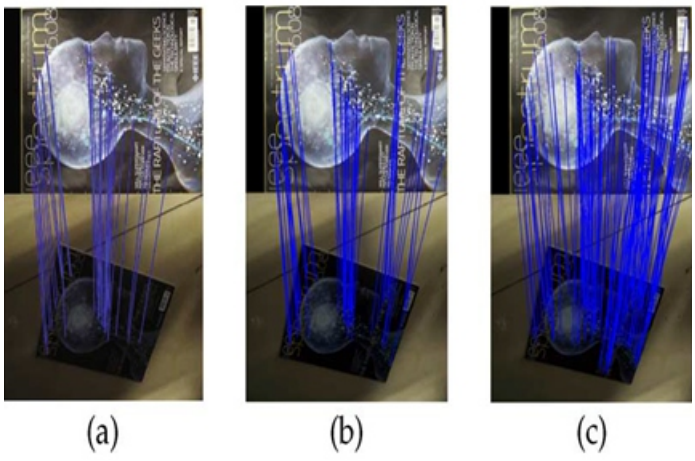

Fig. 3. Matching results of SIFT and ISIFT. (a) Matching result of SIFT. (b)Matching result of ISIFT with only pose simulation (c) Result of ISIFT with both pose and illumination (and) simulation.

TABLE II
Performance of SIFT, ISIFT with Only Pose Estimation, And ISIFT With Both Pose and Illumination Estimation.

	SIFT	ISIFT(H)	ISIFT(H& L)
Total detected	436	388	2021
Total matched	50	64	169
NCM	39	57	153
RS (%)	8.95	14.7	7.50
MP (%)	78.0	89	91

framework with SIFT are drawn for comparison, as shown in Fig. 3(c) The results show that more iterations do not necessarily increase the performance significantly, whereas it increases the computation time linearly. When , the performance significantly increases. TheNCM increases more than 300 matches from only 12 to 365, and the RS increases from 12.1% to 37.1%. However, as further increases the performance little, the NCM only moves around 360, and the RS moves around 37%. Thus, two iterations are enough in general situations, and we use in the following experiments. Moreover, all the features in this experiment and the following experiments are described by a SIFT [10] descriptor, except SURF, which is described by a SURF descriptor [1].

C. Performance:

In this experiment, a brief view of the performance of the proposed method is given.We use SIFT as the base detector in this experiment (ISIFT). Two images with both view and illumination changes are matched here. We first match the two images by SIFT, and then, we only simulate the pose of the left image in our strategy. Finally, we simulate both pose and illumination. The matching results are shown in Fig.3 and Table II. View and illumination changes both degrade the performance of the general method. SIFT could achieve 8.95% RS with 39 correct matches. ISIFT, with the pose estimation only, could achieve14.7% RS with 57 correct matches.

When we estimate the pose and illumination changes, the number of total detected features rapidly increases, and the NCM increase to 153. Because histogram matching amplifies noise in simulation, many fake features are detected, and the RS is reduced to 7.57%. This experiment is only a brief view of our strategy, and more experiments will be presented in the following. We estimate the global illumination change between the matching pair to increase the NCMs. The illumination change is usually continuous in the image. Thus, revising the illumination of part of the image could benefit to other regions.

Our algorithm does not increase the invariance of the original detector, but it increases the accuracy, stability, and reliability of the matching results. When SIFT fails, our method also fails. However, when SIFT works, but not robust, the proposed method will play an important role. More matches could not increase the invariance, but it can increase the accuracy of alignment when the matching by SIFT is inaccurate.

In other words, the advantage of the proposed method is that the performance does not degrade with the increase in the pose change or transition tilt, which is addressed in [1] and [2] in the valid range. Additionally, the local key point location will be more accurate than that of the original detected point. To corroborate this pointof view, we show an extra experiment in the following. The first row in Fig. 8 is the matching results of SIFT, and the second row is the results of ISIFT. Both the matches and the alignment residual error are shown. From this experiment, we can find that our algorithm can obtain less error than SIFT, and the NCM affects the accuracy of matching very much.

D. Comparison:

We compare ISIFT and IHLSIFD with the state-of-the-art methods on scale, affine, and illumination changes. We choose the database provided by Mikolajczyk and compare them with HarAff, HesAff, SURF, SIFT, HLSIFD, and ASIFT. Four pairs of images with scale, view, and illumination change are tested. the images on top are the reference image,and those at the bottom are the test image. Table III is a comparison of this experiment in terms of NCM, RS, and MP. Our method estimates the pose and illumination of the matching pairs and simulates the reference image. Therefore, the simulated image is closer to the original image, which contains most information of the original image, shortening the distance of the matching pairs in the parameter space.

First, the NCM of the IHLSIFD and ISIFT is much higher than that of the traditional methods. They obtain 726 and 584 matches, respectively, whereas HLSIFD obtains 48 matches, and SIFT obtains 46 matches in the Graf (affine change situation; second row in Fig. 9). We increase about 14 and 11 times of matches. Moreover, the total number of features that we extracted is 1797 and 1605, whereas HLSIFD and SIFT obtain 2419 and 2837 features, respectively. Thus, the RS of IIHLSIFD and ISIFT increases to 40.4% and 36.4%, whereas that of HLSIFD and SIFT is only 1.98% and 1.62%. This implies that the efficacy of IIM framework is much better than the traditional DDM framework.

We increase about 19 times and 21 times RS in this view-change experiment. With the significant increasing performance, we can make the matching more stable and reliable. Similarly, more correspondences are found in other experiments, particularlyunder affine and illumination change situations. Our method does not significantly increase NCM under only scale change comparing to SIFT, SURF, and HLSIFD since they are theoretically scale invariant. The RS and MP also significantly increase.

However, in extreme situations when SIFT fails in the first matching, our algorithm also fails. The proposed method can increase the stability, reliability, and accuracy of the original detector, but it cannot increase the invariance. A solution is integrating the proposed method into ASIFT as the second layer.Some matching results of ISIFT in video frames. Zoom in for better view. Frames 1, 100, 500, 805, 806, 807, 811, and 824 are shown, and the blue lines are calculated by the transformation matrix.

5.Conclusion:

In this paper, we have proposed a novel image-matching algorithm based on an iterative framework and two new indicators for local feature detector, namely, the VA and the VI. The proposed framework iteratively estimates the relative pose and illumination relationship between the matching pair and simulates one of them to the other to degrade the challenge of matching images in the valid region (VA and VI). Our algorithm can significantly increase the number of matching pairs, RS, and matching accuracy when the transformation is not beyond the valid region.

The proposed method would fail when the initial estimation fails, which is relative to the ability of the detector. We have proposed two indicators, i.e., the VA and the VI, according to this phenomenon to evaluate the detectors, which reflect the maximal available change in view and illumination, respectively. Extensive experimental results show that our method improves the traditional detectors, even in large variations, andthe new indicators are distinctive.

References:

1.J. Shi and C. Tomasi, "Good features to track," in Proc. Comput. Vis. Pattern Recognit., Jun. 1994, pp. 593–600.

2.Y. Li, Y. Wang, W. Huang, and Z. Zhang, "Automatic image stitching using sift," in Proc. Audio, Lang. Image Process., Jul. 2008, pp. 568–571.

3. R. Szeliski, "Image alignment and stitching: a tutorial," Found. Trends Comput. Graph Vis. vol. 2, no. 1, pp. 1–104, 2006 [Online]. Available:http://dx.doi.org/10.1561/0600000009.

4.M. Brown and D. Lowe, "Unsupervised 3D object recognition and reconstruction in unordered datasets," in Proc. Int. Conf. 3-D Digit. Imag.Model., Jun. 2005, pp. 56–63.

5.A. Davison, W. Mayol, and D. Murray, "Real-time localization and mapping with wearable active vision," in Proc. Int. Symp. Mixed AugmentedReality, 2003, pp. 18–27.

8.C. Harris and M. Stephens, "A combined corner and edge detection," in Proc. 4th Alvey Vis. Conf., 1988, pp. 147–151.

9.S. M. Smith and J. M. Brady, "Susan—A new approach to low level image processing," Int. J. Comput. Vis., vol. 23, no. 1, pp. 45–78, May1997.

10.F. Mokhtarian and R. Suomela, "Robust image corner detection through curvature scale space," IEEE Trans. Pattern Anal. Mach.Intell., vol. 20, no. 12, pp. 1376–1381, Dec. 1998.