A Novel Algorithm for View and Illumination Invariant Image Matching

Yinan Yu, *Student Member, IEEE*, Kaiqi Huang, *Senior Member, IEEE*, Wei Chen, *Student Member, IEEE*, and Tieniu Tan, *Fellow, IEEE*

***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).**

I. Introduction

IMAGE matching is a fundamental issue in computer vision. It has been widely used in tracking [1], image stitching [2], [3], 3-D reconstruction [4], simultaneous localization and mapping (SLAM) systems [5], camera calibration [6], object classification, recognition [7], and so on. Image matching aim to find the correspondence between two images of the same scene or objects in different pose, illumination, and environment. In this paper, we focus on local feature-based image matching. The challenges of this work reside in stable and invariant feature extraction from varying situations and robust matching.

Manuscript received April 28, 2010; revised November 13, 2010, February 16, 2011, and May 14, 2011; accepted May 23, 2011. Date of publication June 27, 2011; date of current version December 16, 2011. This work was supported in part by the National Natural Science Foundation of China under Grant 60736018 and Grant 60723005 and in part by the National Hi-Tech Research and Development Program of China under Grant 2009AA01Z318. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Kenneth K. M. Lam.

The authors are with the National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China (e-mail: tnt@nlpr.ia.ac.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2011.2160271

In image matching, key region or point of interest is often used as the local feature due to its stable performance in detection and description. A region feature is usually derived from a circle or ellipse with certain location and radius and is effective and efficient, compared with other types of features such as edges and contours. Therefore, region features are extensively used in real applications. Generally speaking, the framework of a region feature based image matching consists of three steps.

Detecting stable regions. Interesting points are extracted from images, and the region of interest is the associated circular (or elliptical) region around the interesting point. Generally, researchers use corner (Harris [8], SUSAN [9], CSS [10], etc.) or center of silent region (SIFT [11], SURF [12], DoH [13], HLSIFD [14], etc.) as the interesting point since they are stable and easy to locate and describe. The radius of the region is determined by *a priori* setting (Harris corner) or the region scale (scale invariant features). The total number of features detected is the minimum number of the features extracted from the matched images.

Describing regions. Color, structure, and texture are widely used to describe images in the recent literature. Descriptors with edge orientation information (SIFT and HOG) are also very popular since they are more robust to scale, blur, and rotation.

Matching features. Local features from two images are first matched when they are the nearest pair. A handful of distances can be used in practice, such as L_1 distance, L_2 distance, histogram intersection distance [15], and earth mover's distance [16]. If the nearest distance is higher than k times ($k \in (0,1]$ empirically) of the second nearest distance, the nearest matching pair will be removed. These are the very initial matching results. Then, the priori hypothesis of the object transform filters the un-uniform transformed matches. In this paper, we simply use planar objects to show the effectiveness of the proposed method. For the multitransform problem, the proposed method could be also integrated. Random sample consensus (RANSAC) [17], [18] is used to select the uniform or multiple transformations set from all the matches.

The three parts of the detect–describe–match (DDM) framework determine the performance of image matching. The first step is the basis of this framework. Unstable and variant features increase the difficulties of the next two steps. Researchers mostly focus on the first step for invariant feature extraction and have proposed many excellent detectors [8], [11], [12], [14]. However, an important experience of a pervious work is that all the aforementioned feature detectors are not strictly invariant

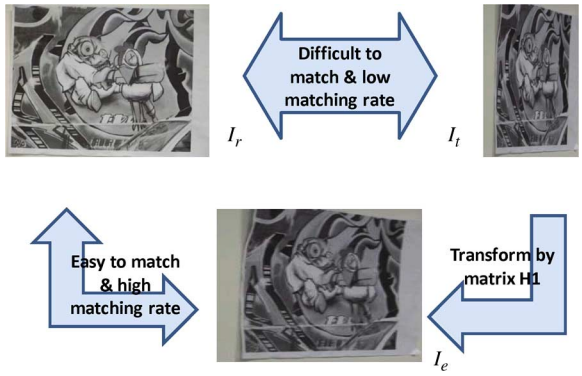


Fig. 1. Illustration of the proposed matching algorithm. I_r and I_t are the images to be matched. I_e is simulated from I_t by transformation T. I_r is difficult to match with I_t for the difference of view point and illumination, whereas I_e is easier to match with I_t since they are closer in the 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 Fig. 1. 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. Thus, two new evaluators, which are termed "valid angle" (VA) and "valid illumination" (VI), are proposed to show this nature ability of local features.

The reminder of this paper is organized as follows: Section II gives a brief introduction to a related work. Then, a new image-matching strategy is presented in Section III. Experiments and feature detector evaluation with our strategy is introduced in Section IV. Section V concludes this paper.

II. Related Work

A. Image Matching With Local Features

The DDM framework is integrated in many systems. Brown and Lowe [21] 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 [21]. Image stitching is easier than wide baseline matching since the main difference between the matching pair is the location and camera focus (scale).

To cope with the change of view in image matching, a feasible solution is simulating the original image to every possible view, extracting features, and matching, respectively. Recently, Morel and Yu [19], [20] have proposed a promising image-matching framework called affine-SIFT (ASIFT). They simulate the original image to discrete poses. The simulations are controlled by two variables: ϕ (horizontal angle) and θ (vertical angle). Choosing a group of values for the two variables, they construct simulations of the image to cover the whole affine space. Finally, SIFT is used to extract features from these simulations. It turns out that the SIFT is not fully affine invariant, whereas ASIFT is fully affine invariant, which is credited to the framework. ASIFT can find the correspondence between the matching pair, even if they are much different in view. The improvement by the novel ASIFT image-matching framework gives us a new viewpoint to image matching.

Real-time applications are constantly proposed as many fast algorithms are developed. Recently, Ta *et al.* [22] have proposed an efficient algorithm called SURFTrac for continuous image matching. SURFTrac extracts interest points in each video frame incrementally. Then, motions of the key points are predicted between consecutive frames. Finally, the key points are updated in the next frame according to the predicted motions. SURFTrac is very efficient, and an interesting application is that this algorithm has been embedded in a Nokia N95 mobile phone, with six to seven frames per second, which proves the efficiency of the algorithm. The SURFTrac algorithm amends the key points pose via the matching results of the preceding frame, which gives pose information about the object in the present frame. The key points extracted from the present frame are restricted by the location and pose of the key points of the preceding frame.

B. Detector Evaluation

Many local invariant feature detectors have been proposed in recent years. Baker [23] develops a feature detector and evaluates it with others. Schmid *et al.* [24] evaluate the interest point detectors using two comparison criteria: repeatability rate and information content. Later, Xiao *et al.* [25] analyze and compare several feature detectors. They focus on the properties and faults of gradient-based feature detectors. Mikolajczyk *et al.* [26] make systematic comparison among several existing affine invariant feature detectors, including Harris-Affine (HarAff), Hessian-Affine (HesAff), MSER [27], EBR [28], and salient region [29]. In their tests, MSER obtains the highest score in view change and Hessian affine follows. However, no detector

outperforms all other detectors for all types of scenes and transformations. To evaluate the feature detectors in 3-D tasks, Moreels and Perona [30] explore the performance of some popular detectors and descriptors. More than 100 3-D objects are used to test the detectors in 3-D situation. They find that no detector could perform well when the viewpoint changes more than 20 °–35 °. Recently, detectors used in the SLAM system are of great interest to researchers. Gil *et al.* [31] compare the interest point detectors and local descriptors for visual SLAM system. Recall-precision curve and RS are compared. However, there is no comparison about the efficiency of the detectors.

Following the general evaluation, three criteria are often used as 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.

III. View and Illumination Invariant Image Matching

A. General Definition of Image Matching

Two images of the same object or scene are shown as two points in parameter space $\mathcal{P}$ of the object (scene). Let I be the original appearance of an object, $I_v = \mathbf{L}(\mathbf{H}(I))$ be the real appearance of the object shown from an image, where $\mathbf{L}$ indicates the illumination, and $\mathbf{H}$ is the object transformation factor from a normal pose. Here, we define the parameter space of a given image $\mathbf{I}$ as $\mathcal{P}_I = \{\mathbf{H_I}, \mathbf{L_I}\}$ (simply written as $\mathcal{P} = \{\mathbf{H}, \mathbf{L}\}$ in the following). Translation $\mathbf{L}$ and $\mathbf{H}$ is a point in the parameter space; thus, the observed image is shown as a point in the parameter space, which is expanded by object I. Therefore, the purpose of image matching is to find transformation T between the two points in the parameter space ($\{\mathbf{L_r}, \mathbf{H_r}\} \xrightarrow{T} \{\mathbf{L_t}, \mathbf{H_t}\}$ or, in other words, $T|I_r = T(I_t), I_r, I_t \in \mathcal{P}_{\mathcal{I}}$). The purpose is to find the coordinate differences between the two points. The norm of this space is difficult to define since illumination factor $\mathbf{L}$ and transformation $\mathbf{H}$ are totally independent and cannot be combined together. In this paper, we simply use images with planar objects; therefore, $\mathbf{H}$ is the homography transform matrix, and $\mathbf{L}$ is the histogram matching function that transform the histogram of one image to a specific one.

B. Proposed Method

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

$$I_r(\mathbf{X}) = \hat{T}(I_t) = \hat{\mathbf{L}}\left(\hat{\mathbf{H}}(I_t)\right) = \hat{\mathbf{L}}\left(I_t(\hat{\mathbf{H}}\mathbf{X})\right) \tag{1}$$

where $\hat{T}$ is the true transformation between I_t and I_r, $\mathbf{x}$ is the homogeneous coordinates, and $\mathbf{x} = (x, y, 1)$. If there exists approximate estimations about illumination and transformation, the I_t could be transformed to an estimated image I, i.e.,

$$I(\mathbf{X}) = T(I_t) = \mathbf{L}\left(I_t(\mathbf{H}\mathbf{X})\right) \tag{2}$$

where $\mathbf{H}$ denotes the view point transformation and $\mathbf{L}$ denotes the illumination transformation. If T is not a very rough estimation between I_r and I_t, the estimated image I would be more similar to I_r than I_t itself. In other words, I_r is closer to I than to I_t. Thus, the matching between I and I_r will be easier, as shown in Fig. 1.

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

$$\begin{aligned} I_1\mathbf{x} &= T_1(I_0) = \mathbf{L_1}\left(\mathbf{I_0}(\mathbf{H_1}\mathbf{x^T})\right) \qquad (I_0 = I_t) \\ I_i\mathbf{x} &= T_i(I_{i-1}) = \mathbf{L_i}\left(\mathbf{I_{i-1}}(\mathbf{H_i}\mathbf{x^T})\right) \qquad (i > 1). \end{aligned} \tag{3}$$

Algorithm 1 The proposed method

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

Iterate

$\mathbf{i} = \mathbf{i} + 1$;

Estimate T_i: $\mathbf{H_i}$, $\mathbf{L_i}$;

$T = T_i \circ T$;

$\mathbf{H} = \mathbf{H_i} * \mathbf{H}$.

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

Until $|\mathbf{H_i} - \mathbf{E}| < \sigma_H$, $|\mathbf{L_i} - \overrightarrow{1}| < \sigma_L$ or $i > n$. ($\mathbf{E}$ is the uint matrix, $\mathbf{L_i}$ is a histogram transformation vector, σ_H and σ_L are convergence thresholds.)

Return T, $\mathbf{H}$

The algorithm is summerized in Algorithm 1. The final estimation of the $\hat{T}$ is

$$\hat{T} = \cdots \circ T_m \circ T_{m-1} \circ \cdots \circ T_2 \circ T_1 \tag{4}$$

$$\approx T_n \circ T_{n-1} \circ \cdots \circ T_2 \circ T_1 \tag{5}$$

where " $\circ$" 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 n.

C. Estimate the Parameters $\mathbf{H}$ and $\mathbf{L}$

General image-matching methods by local features focus on the first parameter $\mathbf{H}$ since the concerned issue is the space correspondence between the two images. Illumination normaliza-

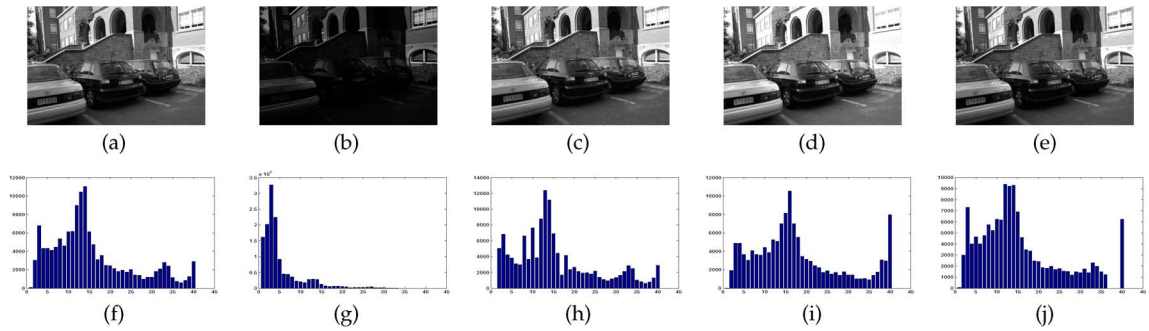


Fig. 2. Illustration of the histogram transformation. (a) The original image. (b) Darker image. (c) Transformed image from (b) according to the histogram of (a). (d) Brighter image. (e) Transformed image from (d) according to the histogram of (a). (f)–(j) The corresponding histograms of (a)–(e).

tion 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 $\mathbf{H}$ 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 two features when they are the nearest pair in the feature space. Here, L_2 norm is used to calculate the distance between to the features. The RANSAC algorithm is employed to calculate transformation matrix $\mathbf{H}$. 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, the ith frame's transformation $\mathbf{H_i}$ can be approximated by the previous results.

Different detectors and descriptors [11], [12] 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, the detector 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 matched regions 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 $\mathbf{L}$ from the other image to the fixed one. Suppose the histogram of the fixed image is h_1 and the histogram of the other image is h_2. We calculate the cumulative functions of h_1 and h_2 — F_1 and F_2. Finally, the translation function is

$$\mathbf{L} = \mathbf{F_2^{-1}F_1}. \tag{6}$$

Since the comulative function of gray histogram is always monotonically increasing, inverse function F^{-1} 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, as shown in Fig. 2, and the whole procedure is illustrated in Fig. 3.

To sum up, we estimate transformation matrix $\mathbf{H}$ 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 [19], [20]. In ASIFT, the features are not invariant to affine change, but they cover the whole affine space, as shown in the middle block in 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

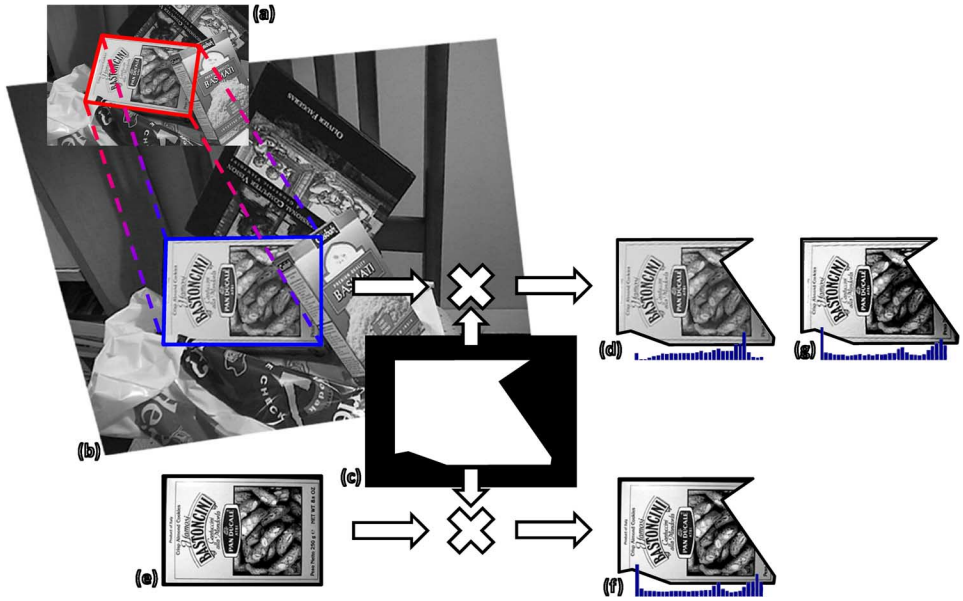


Fig. 3. Procedure of illumination estimation. (a) The test image. Warp (a) by the estimated transformation matrix to generate (b). (c) Mask with the matched regions labeled as 1, and the unmatched regions labeled as 0. (d) The inner product of (b) and (c). (e) The reference image. (f) The inner product of (c) and (e). (g) Illumination simulated image from (d) according to the histogram of (f).

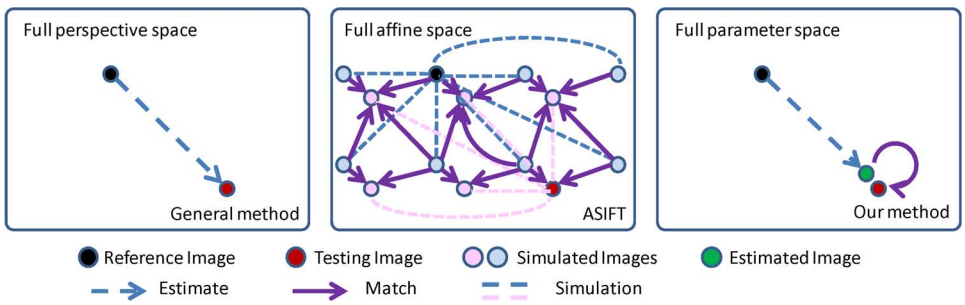


Fig. 4. Relationship among the general framework, ASIFT, and the proposed method. (Left block) The general framework, (middle block) ASIFT, and (right block) ours. The general DDM framework directly estimates the transformation between two images. It is simple but coarse. ASIFT simulates many poses of 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	Ours
Simulation to ref image	$\surd$	$\times$
Simulation to test image	$\surd$	$\surd$
Number of simulations	many	few
Number of features	$10^4 \sim 10^5$	10^3
Pose simulation	$\surd$	$\surd$
Illumination simulation	$\times$	$\surd$
NCM	high	high
RS	very low	high
Affine invariancy	Full	Partial
Compuational cost	high	low
Real-time	$\times$	$\surd$

around the original points in parameter space $\mathcal{P}$, 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. 4. In addition, we transform one simulation per iteration, and in most tasks, two iterations are enough. We will give an experiment to illustrate this in Section IV-B. 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 the accuracy. 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, as shown in Table III. Therefore, the RS of our method is much higher than that of ASIFT. The computational cost of our method is much

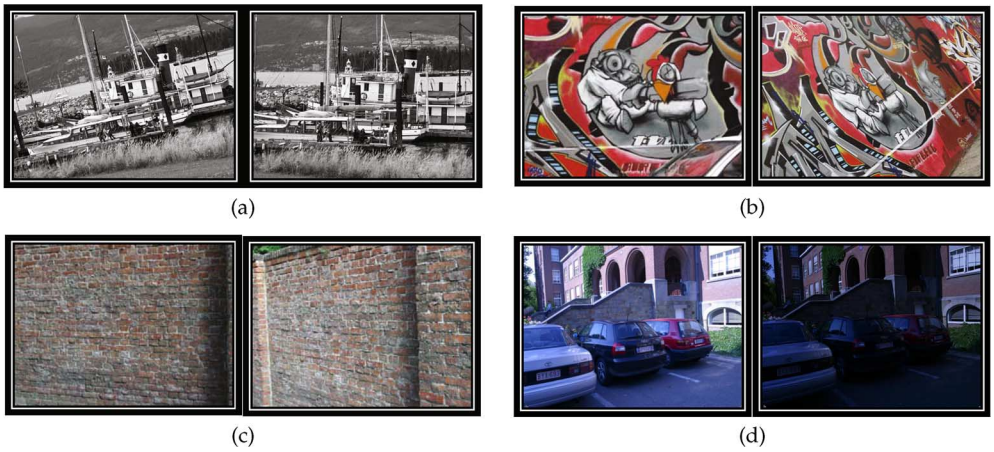


Fig. 5. Four groups of images that we used for comparison [33]. Each group contains one or two transformations with six images, and only parts of them are shown here. (a) Boats (scale +rotation). (b) Graf (view). (c) Wall (view). (d) Leuven (illumination).

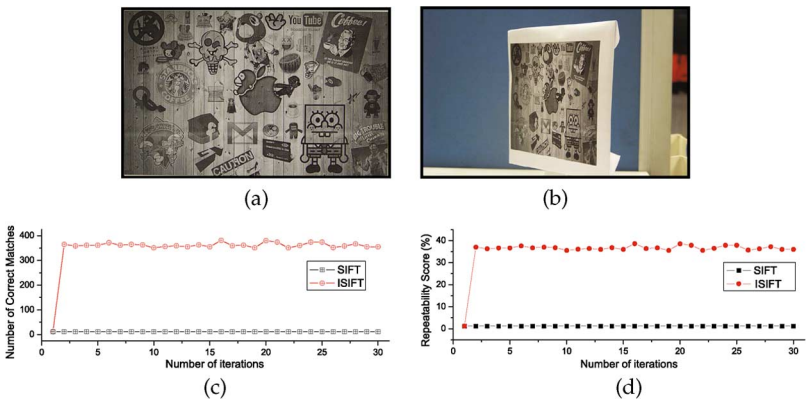


Fig. 6. Experiments of convergence. (a) The reference image. (b) The test image. (c)–(d) The NCM and RS of ISIFT compared with SIFT.

lower than that of ASIFT because much fewer features are required. Above all, there are some common properties between iterative SIFT (ISIFT) and ASIFT. Instead of directly matching the original images, both methods find good simulations of the original pairs. ASIFT samples the imaginary images in the whole affine space, whereas our method directly estimates in the whole parameter space. We should point out here that these comparisons and the experiments shown in the following section are all under the situation that the original method, i.e., SIFT, still works. When it fails, the proposed method also fails, whereas the ASIFT can still obtain a valid result.

IV. Experimental Results

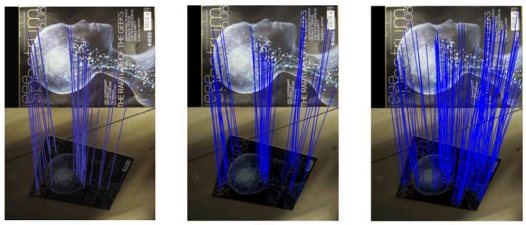
A. Database

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. Parts of them are shown in Fig. 5. We compare the proposed method with 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 [32].

[1] http://www.robots.ox.ac.uk/vgg/research/affine/

B. Convergence

As we mentioned in Section III-B, the number of iteration n 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 °, as shown in Fig. 6(a) and (b), respectively. Here, SIFT is used as the base detector. The RS and NCM of our method and the DDM



(a) (b) (c)

Fig. 7. Matching results of SIFT and ISIFT. (a) Matching result of SIFT. (b) Matching result of ISIFT with only pose simulation ($\mathbf{H}$). (c) Result of ISIFT with both pose and illumination ($\mathbf{H}$ and L) simulation.

TABLE II
PERFORMANCE OF SIFT, ISIFT WITH ONLY POSE ESTIMATION, AND ISIFT WITH BOTH POSE AND ILLUMINATION ESTIMATION

	SIFT	ISIFT($\mathbf{H}$)	ISIFT($\mathbf{H}$ & L)
Total detected	436	388	2021
Total mathes	50	64	169
NCM	39	57	153
RS(%)	8.95	14.7	7.57
MP(%)	78.0	89.1	90.5

framework with SIFT are drawn for comparison, as shown in Fig. 6(c) and (d). The results show that more iterations do not necessarily increase the performance significantly, whereas it increases the computation time linearly. When $n = 2$, the performance significantly increases. The NCM increases more than 300 matches from only 12 to 365, and the RS increases from 12.1% to 37.1%. However, as n 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 $n = 2$ in the following experiments. Moreover, all the features in this experiment and the following experiments are described by a SIFT [11] descriptor, except SURF, which is described by a SURF descriptor [12].

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. 7 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 achieve 14.7% RS with 57 correct matches. When we estimate the pose and illumination changes, the number of total detected features rapidly increase, 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 [19] and [20] in the valid range. Additionally, the local key point location will be more accurate than that of the original detected point. To corroborate this point of 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, as shown in Fig. 9. 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, particularly under 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

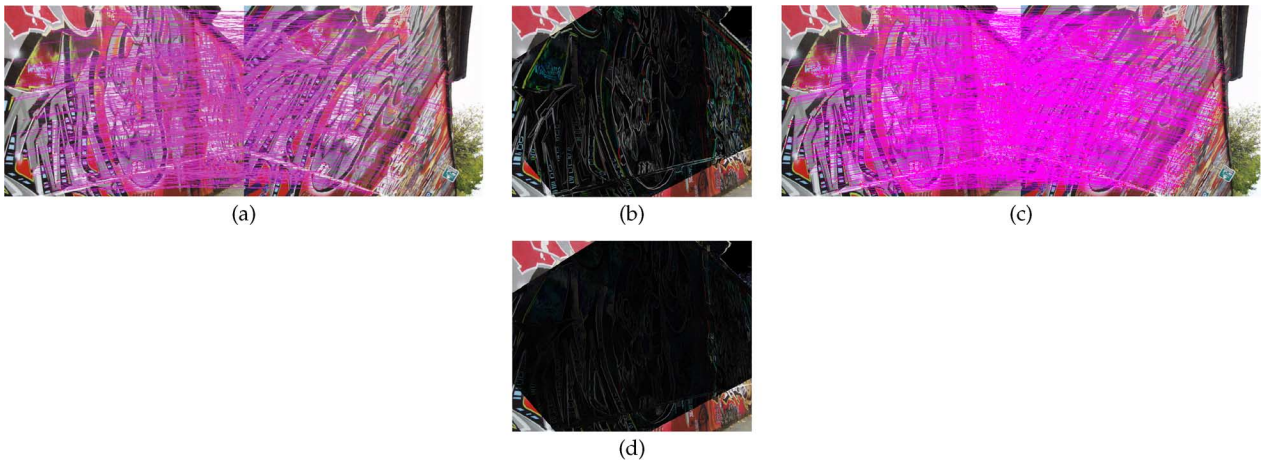


Fig. 8. Matching error of the SIFT and the proposed method. (a) The matches of SIFT. (b) The residual error of SIFT. (c) The matches of ISIFT. (d) The residual error of ISIFT.

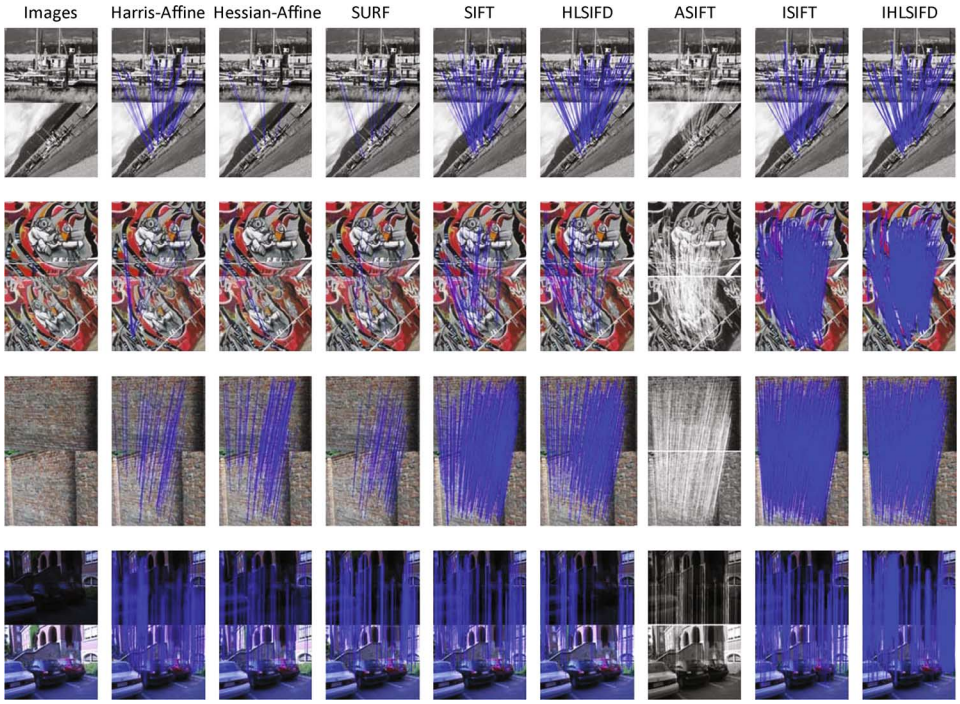


Fig. 9. Matching results of four groups of images. (Test images from top to bottom) Boat, Graf, Wall, and Leuven. The results of the correct matches are drawn in blue or white lines.

to refine the original matching results. We will show an experiment in Section IV-E.

ASIFT also obtains 105, 465, 556, and 157 matches from Boat, Graf, Wall, and Leuven matching images (61, 46, 409, and 259 matches are found by SIFT, respectively). However, these matches are calculated from 29 985, 45 151, 64 908, and 22 562 extracted features. Indeed, ASIFT increases the NCM, but they need to extract much more features from the images, which cost much time in computation. More detail results are summarized in Table III.

In this paper, we try to link our method with the general optimization theory. Essentially, the target of image matching is finding the correspondence. We want to find the transformation function between the matching pair, which can minimize the matching error. Thus, we optimize the view difference and then optimize the illumination. With the two-step optimization, our method can find more accurate transformation function. Different from ASIFT, the proposed method does not increase the invariance of the original detector, but it increases the stability and reliability.

E. Feature Evaluation

In real tasks, given an invariant region detector, we must answer this question: What is the valid condition of this detector? The valid condition includes the tolerance of view and illumination changes. A special property of the iteration framework is that the performance can stay at a stable level with transformations in the valid range. However, when the transformation goes beyond the valid range, our method would fail. According to this property, the proposed method could be used as an evaluator to

TABLE III
COMPARISON OF THE ALGORITHMS ON VIEW CHANGE PAIRS

	Methods	HarAff	HesAff	SURF	SIFT	**HLSIFD**	ASIFT(HR)	**ISIFT**	**IHLSIFD**
Boat/Scale	Total	1457	833	722	7986	**3061**	29985	**615**	**622**
	Matches	190	70	43	125	**73**	/	**94**	**89**
	NCM	37	5	8	61	**63**	105	**79**	**87**
	RS(%)	2.54	1.08	1.25	0.764	**2.06**	0.35	**12.8**	**14.0**
	MP(%)	19.5	7.14	18.6	48.8	**86.3**	/	**84**	**97.8**
Graf/Affine	Total	2739	1325	793	2837	**2419**	45151	**1605**	**1797**
	Matches	330	137	34	210	**103**	/	**586**	**836**
	NCM	10	4	9	46	**48**	465	**584**	**726**
	RS(%)	0.365	0.302	1.14	1.62	**1.98**	1.03	**36.4**	**40.4**
	MP(%)	3.03	2.92	26.5	21.9	**46.6**	/	**99.7**	**86.8**
Wall/Affine	Total	2479	3592	1730	7094	**4735**	64908	**5358**	**2441**
	Matches	206	317	78	452	**247**	/	**834**	**681**
	NCM	71	120	40	409	**241**	556	**833**	**676**
	RS(%)	2.86	3.34	2.31	5.77	**5.09**	0.857	**15.5**	**27.7**
	MP(%)	34.5	37.9	51.3	90.5	**97.6**	/	**99.9**	**99.3**
Leuven/Illumination	Total	3216	1153	647	999	**705**	22562	**1159**	**1250**
	Matches	1160	430	172	289	**213**	/	**379**	**621**
	NCM	536	192	161	259	**199**	157	**344**	**601**
	RS(%)	16.7	16.7	24.9	25.9	**28.2**	6.96	**29.7**	**48.1**
	MP(%)	46.2	44.7	93.6	89.6	**93.4**	/	**90.8**	**96.8**

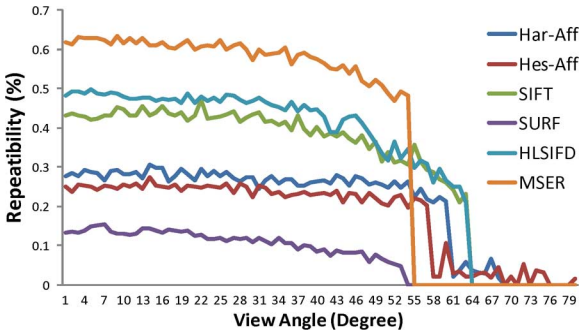


Fig. 10. Five algorithms are used as the basic detector in our strategy. The performances of the HarAff, HesAff, SURF, SIFT, and HLSIFD are shown. The VA point of each method is labeled on the RS line. The VA of HarAff, HesAff, SURF, SIFT, and HLSIFD are 61 °, 58 °, 54 °, 65 °, and 64 °, respectively.

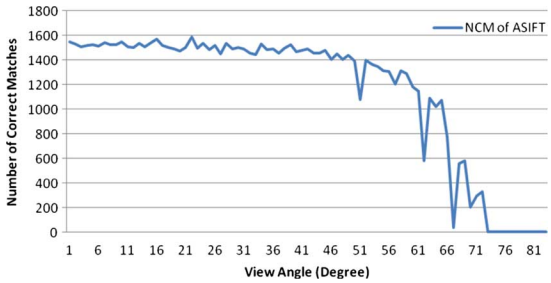


Fig. 11. Numbers of correct matches of ASIFT in different view angles.

the feature detectors. Our method could give numerical evaluation to the detectors in view and illumination changes. When the change degree of viewpoint is larger than a threshold, which we call the VA, or the illumination change degree is beyond a threshold, which we call VI, the method will fail. The VA and VI rely on the tolerance of the basic feature detector.

The VA is similar to the "transition tilt," particularly to the maximum "transition tilt." There are two differences between them. First, the transition tilt is proposed to measure the change of the matching images in pose, whereas the VA is proposed to describe the ability of the original detector. Second, the

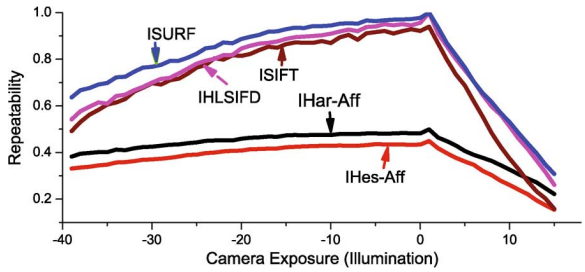


Fig. 12. Five kinds of algorithms are used as the basic detector in illumination change test. All the detectors pass this test. ISURF, IHLSIFD, and ISIFT win the first three, whereas IHar-Aff and IHes-Aff are the last two.

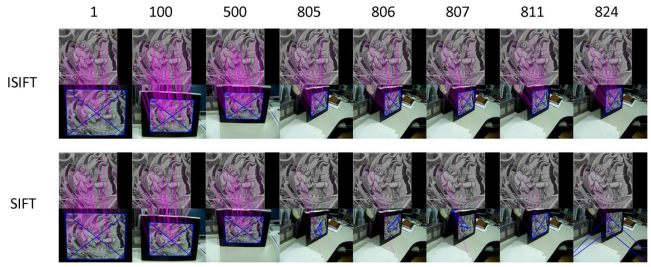


Fig. 13. Some matching results of ISIFT in video frames. Zoom in for better view. Frames 1, 100, 500, 805, 806, 807, 811, and 824 are shown, and the blue lines are calculated by the transformation matrix.

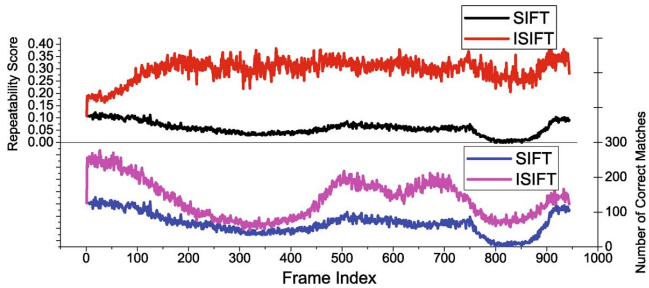


Fig. 14. Matching results in video frames. (Top) RSs of ISIFT and SIFT and (bottom) the NCMs.

transition tilt is found by experimental results and concluded by a human, whereas the VA can be calculated by our framework

Fig. 15. Our real-time image matching system. (a)–(f) SIFT is used as the base detector, and (g)–(l) SURF is used as the base detector. Total number of features detected, NCM, and computation time are presented. (a) ISIFT: 181 features, 31 matches, 70.0 ms. (b) ISIFT: 148 features, 85 matches, 63.2 ms. (c) ISIFT: 150 features, 96 matches, 62.1 ms. (d) ISIFT: 152 features, 53 matches, 62.6 ms. (e) ISIFT: 219 features, 41 matches, 77.1 ms. (f) ISIFT: 100 features, 20 matches, 52.2 ms. (g) ISURF: 305 features, 30 matches, 29.8 ms. (h) ISURF: 474 features, 54 matches, 37.3 ms. (i) ISURF: 302 features, 45 matches, 29.5 ms. (j) ISURF: 529 features, 176 matches, 38.6 ms. (k) ISURF: 326 features, 52 matches, 29.5 ms. (l) ISURF: 187 features, 39 matches, 23.4 ms.

itself. The VA is the maximum angle that the detector still works or the matching algorithm gives right result. The definition of "matching algorithm giving right results" is not an easy work for SIFT or ASIFT. However, it can be automatically obtained by our framework. If transform matrix $\mathbf{H}$ does not converge with some matching pairs, the matching algorithm fails in the task and the same to VI. We combine the proposed framework with HarAff, HesAff, SURF, SIFT, HLSIFD, and MSER as the basic detectors in real-scene frames with sequential view changes from 0 ° to 87 °n. Here, the reference image is set in front view (0 °). The RS of the proposed framework with the five state-of-the-art methods (IHar-Aff, IHes-Aff, ISURF, ISIFT, and IMSER) and HLSIFD (IHLSIFD) are shown in Fig. 10. The RS curves are stable from 0 ° to a large view range, but the performance considerably degrades when they outrange the VA. The VA is a very important indicator for feature detectors because the detectors are only effective in the VA range while invalid out of the VA. Experimental results in Fig. 10 show that SIFT with 65 ° is the best one in view change.

Why does SIFT win in view change? Image areas contain local structure information. The first-order (I_x, I_y, I_xI_y) and second-order (I_{xx}, I_{yy}, I_{xy}) gradients can represent this structure simply. They construct two matrix called squared difference matrices and Hessian matrix as

$$M = \begin{bmatrix} I_x^2 & I_xI_y \\ I_xI_y & I_y^2 \end{bmatrix} \tag{7}$$

$$H = \begin{bmatrix} I_{xx} & I_{xy} \\ I_{xy} & I_{yy} \end{bmatrix}. \tag{8}$$

The eigenvalues of the matrices represent the edgelike of the area. Thus, if the suppression of the method to the edgelike is very strong, this area will be no more salient. Yu *et al.* [14] find that the HLSIFD has the strongest suppression to the edges, and the DoG gives the lowest suppression. Thus, more edgelike features could be obtained by SIFT, whereas the HLSIFD ignores the features that turn from salient to edgelike areas.

ASIFT is not a detector but a framework as our method. However, we also show the advantage of ASIFT here, as shown in Fig. 11. The NCM of ASIFT under our framework is computed, as in previous evaluation. The ASIFT has a wider range than the detectors previously mentioned.

To test the effectiveness with illumination, we test the five detectors in consecutive frames with illumination change. All the detectors succeed in this test, as shown in Fig. 12. We find in this experiment that the performance of ISURF, ISIFT, and IHLSIFD is much better than that of IHar-Aff and IHes-Aff. This illustrates that HarAff and HesAff are not very distinctive in feature extraction. Many similar features around a region are extracted. Thus, the nearest neighborhood matching method would not be effective to such features.

F. Real-Time Image Matching

An important application of image matching is object detection and pose estimation in video frame. Suppose that the camera smoothly moves and the reference image can be matched with the first frame, the estimation of the transformation matrix from the reference image to certain frame in video can be initialized from the matching of the previous frame. In addition, we match the first frame with the reference image directly by local-feature-based image-matching method. We directly use SIFT here. The RS and NCM of our method and SIFT are shown in Fig. 14, and parts of the matching results of ISIFT and SIFT are shown in Fig. 13. The RS of our method (ISIFT is used here) stays around 30%, and NCM is always higher than 100 pairs in this experiment. The RS of SIFT is running around 7%. Only a small part of features are useful

for the correspondence calculation. The NCM of SIFT is about 70 matches, which is lower than that of the proposed method. The mean of the RS and NCM of the ISFIT and SIFT is, respectively 29.6%, 137, 5.7%, and 66. Our method accurately calculates matches all through the video frames, even in large view changes such as frames 750 to 900. To sum up, ISIFT is very accurate and stable in real applications.

We develop a real-time image-matching system to show the efficiency. The proposed method could cope with a wide range of view and illumination changes with stable matches, as shown in Fig. 15. We compare the real performance of SURF and SIFT by using them as our basic detector. ISURF is faster than ISIFT; however, it is not as stable as ISIFT. The system is implemented on a computer with two dual-core 2.8-GHz central processing unit, and the processed image size is 640×480. The matching could be finished in 80 ms, with parallel coding in a algorithmic level.

V. 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, and the 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 Augmented Reality*, 2003, pp. 18–27.

[6] B. Telle, M. J. Aldon, and N. Ramdani, "Camera calibration and 3d reconstruction using interval analysis," in *Proc. Int. Conf. Image Anal. Process.*, 2003, pp. 374–379.

[7] D. Lisin, M. Mattar, M. Blaschko, E. Learned-Miller, and M. Benfield, "Combining local and global image features for object class recognition," in *Proc. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2005, p. 47.

[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, May 1997.

[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.

[11] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Nov. 2004.

[12] H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool, "Speeded-up robust features (SURF)," *Comput. Vis. Image Understand.*, vol. 110, no. 3, pp. 346–359, Jun. 2008.

[13] T. Lindeberg, *Scale-Space Theory in Computer Vision*. Norwell, MA: Kluwer, 1994.

[14] Y. Yu, K. Huang, and T. Tan, "A Harris-like scale invariant feature detector," in *Proc. Asian Conf. Comput. Vis.*, 2009, pp. 586–595.

[15] A. Barla, F. Odone, and A. Verri, "Histogram intersection kernel for image classification," in *Proc. Int. Conf. Image Process.*, 2003, vol. 3, p. III–513–16 [Online]. Available: http://dx.doi.org/10.1109/ICIP.2003.1247294, vol.2

[16] Y. Rubner, C. Tomasi, and L. J. Guibas, *A Metric for Distributions With Applications to Image Databases*. Washington, DC: IEEE Comput. Soc., 1998, p. 59.

[17] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, Jun. 1981.

[18] J. Rabin, J. Delon, Y. Gousseau, and L. Moisan, "MAC-RANSAC: A robust algorithm for the recognition of multiple objects," in *Proc. 3D'PVT*, 2010, pp. 1–8.

[19] J. M. Morel and G. Yu, "Asift: A new framework for fully affine invariant image comparison," *SIAM J. Imag. Sci.*, vol. 2, no. 2, pp. 438–469, Apr. 2009.

[20] G. Yu and J. Morel, "A fully affine invariant image comparison method," in *Proc. IEEE ICASSP*, 2009, pp. 1597–1600.

[21] M. Brown and D. G. Lowe, "Automatic panoramic image stitching using invariant features," *Int. J. Comput. Vis.*, vol. 74, no. 1, pp. 59–73, Aug. 2007.

[22] D. Ta, W. Chen, N. Gelfand, and K. Pulli, "Efficient tracking and continuous object recognition using local feature descriptors," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2009, pp. 2937–2944.

[23] S. Baker, "Design and evaluation of feature detectors," Ph.D. dissertation, Columbia Univ., New York, 1998.

[24] C. Schmid, R. Mohr, and C. Bauckhage, "Evaluation of interest point detectors," *Int. J. Comput. Vis.* vol. 37, no. 2, pp. 151–172, Jun. 2000 [Online]. Available: http://perception.inrialpes.fr/Publications/2000/SMB00

[25] Z. Xiao, M. Yu, C. Guo, and H. Tang, "Analysis and comparison on image feature detectors," in *Proc. Int. Symp. Electromagn. Compat.*, May 2002, pp. 651–656.

[26] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. V. Gool, "A comparison of affine region detectors," *Int. J. Comput. Vis.*, vol. 65, no. 1/2, pp. 43–72, Nov. 2005.

[27] J. Matas, O. Chum, U. Martin, and T. Pajdla, "Robust wide baseline stereo from maximally stable extremal regions," in *Proc. Brit. Mach. Vis. Conf.*, 2002, vol. 1, pp. 384–393.

[28] T. Tuytelaars and L. Van Gool, "Content-based image retrieval based on local affinely invariant regions," in *Proc. Int. Conf. Vis. Inf. Syst.*, 1999, pp. 493–500.

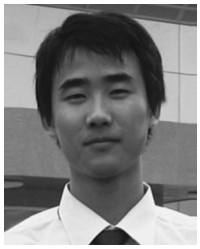
[29] T. Kadir, A. Zisserman, and M. Brady, "An affine invariant salient region detector," in *Proc. Eur. Conf. Comput. Vis.*, 2004, pp. 228–241.

[30] P. Moreels and P. Perona, "Evaluation of features detectors and descriptors based on 3d objects," in *Proc. Int. Conf. Comput. Vis.*, Oct. 2005, vol. 1, pp. 800–807.

[31] A. Gil, O. Mozos, M. Ballesta, and O. Reinoso, "A comparative evaluation of interest point detectors and local descriptors for visual SLAM," *Mach. Vis. Appl.*, vol. 21, no. 6, pp. 905–920, Oct. 2010.

[32] Y. Yu, A Database of Sequential Images for Image Matching Apr. 2010 [Online]. Available: http://www.cbsr.ia.ac.cn/users/ynyu/IIM.html

[33] K. Mikolajczyk, "Detection of Local Features Invariant to Affines Transformations" Ph.D. dissertation, Inst. Nat. Polytech. Grenoble, Grenoble, France, Jul. 2002 [Online]. Available: http://perception.inrialpes.fr/Publications/2002/Mik02



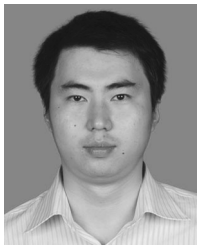
Yinan Yu (S'09) received the B.S. degrees in telecommunication engineering from Beijing University of Posts and Telecommunications, Beijing, China, in 2007. He is currently working toward the Ph.D. degree in computer vision and pattern recognition in the National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing.

His current interests include image/video analysis, visual object recognition, perception, and visual cognition.

Kaiqi Huang (SM'07) received the B.Sc. and M.Sc. degrees from Nanjing University of Science Technology, Nanjing, China, and the Ph.D. degree from Southeast University, Nanjing.

He has worked with National Laboratory of Pattern Recognition (NLPR), Institute of Automation, Chinese Academy of Science, Beijing, China, and he has been an Associate Professor with NLPR since 2005.

Dr. Huang was the Deputy General Secretary of the IEEE Beijing Section from 2006 to 2008.



Wei Chen (S'10) received the B.E. degree from the University of Electronic Science and Technology of China, Chengdu, China, in 2007 and the M.E. degree from the Institute of Automation, Chinese Academy of Science, Beijing, China, in 2010. He is currently working toward the Ph.D. degree at the University at Buffalo, The State University of New York, Buffalo.

Since spring 2011, he has been with the Department of Computer Science and Engineering, UB. He is currently working with Dr. Y. Fu and Dr. J. Corso on his Ph.D. degree. His research interests include applied machine and computer vision.

Tieniu Tan (F'04) received the B.Sc. degree in electronic engineering from Xi'an Jiaotong University, Xi'an, China, in 1984 and the M.Sc. and Ph.D. degrees in electronic engineering from Imperial College of Science, Technology and Medicine, London, U.K., in 1986 and 1989, respectively.

In October 1989, he joined the Computational Vision Group, Department of Computer Science, University of Reading, Berkshire, U.K., where he worked as a Research Fellow, a Senior Research Fellow, and a Lecturer. In January 1998, he returned to China to join the National Laboratory of Pattern Recognition (NLPR), Institute of Automation, Chinese Academy of Sciences (CAS), Beijing, China, where he is currently a Professor and the Director of NLPR and was a former Director General of the Institute from 2000 to 2007. He is also a Deputy Secretary General of CAS. He has published more than 300 research papers in refereed journals and conference proceedings in the areas of image processing, computer vision, and pattern recognition. His current research interests include biometrics, image and video understanding, information hiding, and information forensics. He has given invited talks at many universities and international conferences.

Dr. Tan is a Fellow the International Association of Pattern Recognition (IAPR) and a member of the IEEE Computer Society. He currently serves as the executive vice president of the Chinese Society of Image and Graphics. He is or has served as an Associate Editor or a member of the editorial board of many leading international journals, including the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, the IEEE TRANSACTIONS ON AUTOMATION SCIENCE AND ENGINEERING, the IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, PATTERN RECOGNITION, Pattern Recognition Letters, and *Image and Vision Computing Journal*. He is the editor-in-chief of the *International Journal of Automation and Computing and Acta Automatica Sinica*. He has served as the chair or a program committee member for many major national and international conferences. He is the vice chair of the IAPR and a founding chair of the IAPR/IEEE International Conference on Biometrics and the IEEE International Workshop on Visual Surveillance.