A New Shape Adaptation Scheme to Affine Invariant Detector

Congxin Liu, Jie Yang, Yue Zhou and Deying Feng
Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University
Shanghai 200240, China
[e-mail: phillx@sjtu.edu.cn]
*Corresponding author: Congxin Liu

Received July 26, 2010; revised August 30, 2010; revised October 8, 2010; accepted October 28, 2010; published December 23, 2010

Abstract

In this paper, we propose a new affine shape adaptation scheme for the affine invariant feature detector, in which the convergence stability is still an opening problem. This paper examines the relation between the integration scale matrix of next iteration and the current second moment matrix and finds that the convergence stability of the method can be improved by adjusting the relation between the two matrices instead of keeping them always proportional as proposed by previous methods. By estimating and updating the shape of the integration kernel and differentiation kernel in each iteration based on the anisotropy of the current second moment matrix, we propose a coarse-to-fine affine shape adaptation scheme which is able to adjust the pace of convergence and enable the process to converge smoothly. The feature matching experiments demonstrate that the proposed approach obtains an improvement in convergence ratio and repeatability compared with the current schemes with relatively fixed integration kernel.

Keywords: Local image feature, affine invariant feature, scale invariant feature, adaptive kernel shape

This research is supported by projects of international cooperation between Ministries of Science and Technology, No: 2009DFA12870 and aviation science fund, China, No: 2008CZ57. We would like to thank the anonymous reviewers for their valuable comments.

DOI: 10.3837/tiis.2010.12.016

1. Introduction

Local image features have shown excellent performance in many applications such as image matching and registration, image retrieval and classification, object and texture recognition, 3D reconstruction and stereo matching. The fundamental difficulty of using local image features resides in extracting features invariant to differing views. There are many impressive works in the literature trying to address the problem, which are shown in Section 2. Harris interest point [1], which is a very popular local feature, is robust to geometric and photometric deformations, but is not invariant to scaling [2]. Mikolajczyk, K. and Schmid [3] extended Harris to be invariant to scale and affine transformation based on previous works [4][5]. This detector is called Harris-Affine interest point in this paper. However, the convergence stability of this algorithm needs to be further investigated [3].

Lindeberg and Garding [4] introduced the second moment matrix defined in affine Gaussian scale space as the metric matrix (determined by an integration Gaussian kernel and a differentiation Gaussian kernel, which are determined by an integration scale matrix and a differentiation scale matrix respectively. See equation (3) in Section 3.1.2 for more details) of local image pattern to find blob-like affine features. During the iteration process of finding affine invariant regions, to reduce the search space, Lindeberg et al. assumed that the integration scale matrix and differentiation scale matrix are coupled and derived fixed point property based on the above assumption, which constitutes the theory basis of later iteration algorithms [3][4][5]. According to the fixed-point property, these iteration algorithms all required that the expected integration and differentiation scale matrices (next iteration) are proportional to the current second moment matrix (current iteration) at the given point. However, such linear relation between them is probable to cause overshoot of anisotropy measured by the second moment matrix at the next iteration, and may hence reduce the stability of convergence.

This paper regards the expected integration and differentiation scale matrices as the function of the current second moment matrix and its anisotropy. Based on the anisotropy measured in each iteration, we dynamically adjust the relation between the two scale matrices and the second moment matrix. To sum up, our approach has the following properties:

(1) The relation between the expected integration and differentiation scale matrices and the current second moment matrix is always non-linear.

(2) The relation between the expected integration and differentiation scale matrices and the current second moment matrix is dynamically updated.

(3) With the increasing number of iteration, the shapes of the expected integration and differentiation scale matrices will gradually approach to the shape determined by the current second moment matrix.

The remainder of this paper is organized as follows: In Section 2, we review the previous works in scale- and affine-invariant feature detection. In Section 3, the details of the proposed method are presented. In Section 4, we provide the detailed experiment results on feature matching experiments. Section 5 concludes the paper.

2. Related Work

Many scale- and affine- invariant features detection algorithms have been explored in the literature [6].

2.1 Scale Invariant Feature

The automatic scale selection for various keypoint detectors have been thoroughly explored by Lindeberg [7]. For the detection of blob-like feature points, Lindeberg proposed to search for the local absolute extremum of $\sigma_D^2 \operatorname{Trace}\left(\mathrm{H}\left(\mathbf{x};\sigma_D\right)\right)$ (also called LOG, where σ_D is the differentiation scale) and $\sigma_D^4 \det \mathrm{H}\left(\mathbf{x};\sigma_D\right)$ in the 3D scale-space.

$$\mathrm{H}\left(\mathbf{x};\sigma_D\right)=\begin{pmatrix} L_{xx}\left(x,y;\sigma_D\right) & L_{xy}\left(x,y;\sigma_D\right) \\ L_{xy}\left(x,y;\sigma_D\right) & L_{yy}\left(x,y;\sigma_D\right) \end{pmatrix} \tag{1}$$

LOG can be well approximated by DOG (difference of Gaussian) [8]. By convolving the image with the difference of Gaussian kernels at a variety of scales and selecting local maxima in 3D scale-space, DOG achieves better computation efficiency compared with LOG. To further accelerate the detection, Bay et al. [9] provided an efficient implementation of $\sigma_D^4 \det \mathrm{H}\left(\mathbf{x};\sigma_D\right)$ by applying the integral image for the computation of image derivatives.

Mikolajczyk and Schmid [3] proposed to use the multi-scale Harris detector (or structure tensor) to determine spatial interest points and the characteristic scales at these points were determined by searching local peaks of $\sigma_D^2 \operatorname{Trace}\left(\mathrm{H}\left(\mathbf{x};\sigma_D\right)\right)$ over scales, leading to the Harris-Laplace detector. This detector provides complementary features points to blob-like detectors. To optimize keypoint detection to achieve stable local descriptors, Gyuri Dorko and Schmid [10] constructed a scale space based on the stability of region descriptor. The characteristic scale was selected at the stability extremum of the local descriptor. Wei-Ting Lee et al. [11] followed this idea and extract feature points directly on the histogram-based representations. Förstner et al. [12][13] proposed a new scale space framework based on the structure tensor and the general spiral feature model to detect scale-invariant features (SFOP) which have shown good repeatability under scale and rotation transformation.

2.2 Affine Invariant Feature

Many methods such as MSER [14], EBR [15], IBR [15], affine saliency [16], Harris-Affine [3], and Hessian-Affine [3][17] are motivated by the fact that only scale invariance cannot arrive at a reliable image matching under significant view changes.

MSER (Maximally Stable Extremal Regions), EBR and IBR are all region and boundary based detector. MESER extracts a nest of regions called Extremal Regions, where each pixel intensity value in a region is less (or greater) than a certain threshold, and all intensities around the boundary are greater (or less) than the same threshold. An Extremal Region is a Maximally Stable Extremal Region when it remains stable over a range of thresholds. IBR [15] first detects local maxima of the image intensity over multiple scales and then a region boundary is determined by seeking for intensity singularity of the intensity along rays radially emanating around these points. The regions extracted by MSER and IBR are very similar. EBR starts from a corner, then finds another two anchor points along the two edges by the relatively invariant geometrical constraint [15], and finally obtain parallelogram regions. Kadir et al. [16] extracts circular or elliptic regions in the image as maxima of the entropy scale-space of region. On one hand the detector can extract many features, but on the other hand it is known to have higher computational complexity and lower repeatability than most other detectors [17]. Harris-Affine and Hessian-Affine first detect feature points with multi-scale second moment matrix and Hessian matrix. Subsequently, affine covariant regions around these points are obtained by affine shape adaptation with the second moment matrix.

A detailed performance comparison between them has been presented in [17], where it is shown that Hessian-Affine and Harris-Affine not only provides more features compared with other methods, but also gives good repeatability and matching scores. Therefore, we focus on them and try to enhance their performance.

Scale space based affine invariant feature Harris-Affine and Hessian-Affine experienced the following developing process. Lindeberg and Garding [4] developed a method for finding blob-like affine features using an affine adaptation scheme based on the second moment matrix, deriving the affine invariance property of the second moment matrix and the shape adapted fixed-point property. Later, Baumberg [5] found the scheme can be used to determine the affine deformation of an isotropic structure. He first extracted multi-scale Harris points and then adapted the shapes of their neighborhoods to the local image structures using the iteration scheme proposed by Lindeberg. During the adaptation, the location and scale of the feature point keep invariant, thereby introducing some localization and shape bias. F. Schaffalitzky and A. Zisserman [18] employed the Harris-Laplace [3] detector to extract feature point and computed an affine invariant neighborhood using the method proposed by Baumberg. Likewise, this approach is also likely to result in localization and shape bias when matching image pairs under significant viewpoint. Mikolajczyk and Schmid [3] improved the previous methods and implemented a full scale adaptation approach of Lindeberg, that is the scale, location and affine shape are updated in each iterative step, thereby gaining better localization effect. Our method tries to further enhance the performance of the detector by dynamically changing the shapes of integration and differentiation kernels in iteration process.

3. The Proposed Affine Invariant Detector

3.1 Background

3.1.1 Scale Normalized Second Moment Matrix

Lindeberg and Garding [4] proposed to use the second moment matrix to estimate the anisotropic shape of a local image structure. Scale normalized second moment matrix, also known as the local shape matrix, is defined as follows:

$$\mu(\mathbf{x};\sigma_D,\sigma_I)=\sigma_D^2 G(\mathbf{x};\sigma_I)\otimes(\nabla L(\mathbf{x};\sigma_D)\nabla L(\mathbf{x};\sigma_D)^T) \tag{2}$$

where σ_D is the differentiation scale, σ_I is the integration scale, $\nabla L(\mathbf{x};\sigma_D)$ is the gradient vector computed at $\mathbf{x}$ with the Gaussian derivative at scale σ_D. This matrix can also be used to extract multi-scale Harris points [3][5].

3.1.2 Affine Second Moment Matrix

Many textured patches show distinct anisotropy, varying scales along different directions. To obtain a more accurate description of textured patches, Lindeberg et al. [4] used the second moment matrices defined in affine Gaussian scale space to describe the anisotropic shape of the local image structure. The normalized second moment matrices are defined as follows:

$$\mu(\mathbf{x};\Sigma_I,\Sigma_D)=\det(\Sigma_D)G(\mathbf{x};\Sigma_I)\otimes(\nabla L(\mathbf{x};\Sigma_D)\nabla L(\mathbf{x};\Sigma_D)^T) \tag{3}$$

where:

$$G(\mathbf{x};\Sigma_I)=\frac{1}{2\pi\sqrt{\det(\Sigma_I)}}\exp(-\frac{\mathbf{x}^T\Sigma_I^{-1}\mathbf{x}}{2}) \tag{4}$$

$$\nabla L(\mathbf{x};\Sigma_D) = \nabla G(\mathbf{x};\Sigma_D) \otimes I(\mathbf{x}) \tag{5}$$

In (4) and (5), Σ_I^{-1} and Σ_D^{-1}, which are 2×2 covariance matrices, refer to integration scale matrix and differentiation scale matrix. These two matrices can determine the shapes of the integration and differentiation Gaussian kernels.

Suppose two image patches are related by an affine transformation $\mathbf{x}_R = \mathbf{A}\mathbf{x}_L$. Then the affine second moment matrices μ_L and μ_R computed at $\mathbf{x}_L$ and $\mathbf{x}_R$ are shown as follows:

$$\mu_L = \mu(\mathbf{x}_L;\Sigma_{I,L},\Sigma_{D,L}), \quad \mu_R = \mu(\mathbf{x}_R;\Sigma_{I,R},\Sigma_{I,R})$$

Lindeberg et al. [4] had verified the following properties:

$$\mu_L = \mathbf{A}^T\mu_R\mathbf{A}, \quad \Sigma_{I,R} = \mathbf{A}\Sigma_{I,L}\mathbf{A}^T, \quad \Sigma_{D,R} = \mathbf{A}\Sigma_{D,L}\mathbf{A}^T \tag{6}$$

Based on the above equation(6), it's easy to verify the property of affine invariance of μ.

3.1.3 Invariance Property of Fixed-point

In a realistic situation, the affine transformation $\mathbf{A}$ is always unknown. Therefore, to find corresponding μ_L and μ_R, we must compute all possible combinations of kernel parameters on brute force, which is unpractical. But if the second moment μ_L is calculated in such a way:

$$\mu_L = t\Sigma_{I,L}^{-1}, \quad \mu_L = s\Sigma_{D,L}^{-1} \quad (t,s \in \Re^+) \tag{7}$$

Lindeberg et al. [4] claimed the above fixed -point property is persevered under affine transformation. In other words, the following equations also hold if $\mathbf{x}_R = \mathbf{A}\mathbf{x}_L$:

$$\mu_R = t\Sigma_{I,R}^{-1}, \quad \mu_R = s\Sigma_{D,R}^{-1} \quad (t,s \in \Re^+) \tag{8}$$

During the process of detection, suppose estimated Σ_I^{-1} and Σ_D^{-1} verify (7) or(8), we assume that there is an unknown affine transformation $\mathbf{A}$ between the two patches.

Based on the fixed-point property, the previous iterative schemes in [3][4] and [5] all let Σ_I^{-1} and Σ_D^{-1} be proportional to the second moment matrix in each iteration.

In addition, all the operations in affine shape adaptation are performed in the transformed image domain [3][5]. The points in neighborhood around feature points are normalized by transformation $\mathbf{x}_{norm} = \mu^{1/2}\mathbf{x}$, and then uniform Gaussian kernels can be applied to calculation. If computed $(\mu, \Sigma_I^{-1}, \Sigma_D^{-1})$ verifies the condition (7) or (8), μ will be isotropic in the transformed image domain.

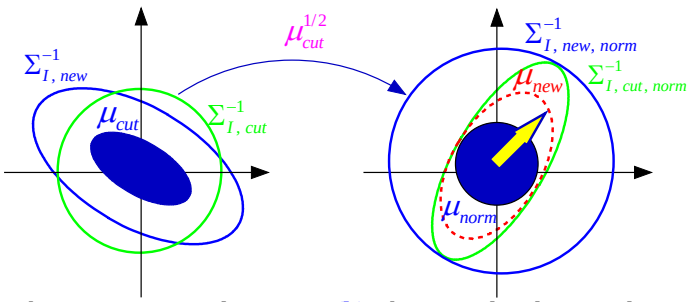
3.2 Adaptive Kernel-shape based Affine Invariant Detector

In the following, we discuss adaptive kernel-shape based affine invariant detector. Our approach does not restrict the type of interest point. Without loss of generality, we choose multi-scale Harris as detectors. For each feature point, the second moment matrix is calculated and the corresponding integration and differentiation scales are selected automatically.

3.2.1 Motivation

At first, in order to make following description more clear, we assume the shape of a matrix $\mathbf{A}$ is the elliptical region defined by $\mathbf{x}^T\mathbf{A}\mathbf{x} = 1$ and its aspect is defined as the direction of the eigenvector corresponding to the minimum eigenvalue of $\mathbf{A}$. **Fig. 1** shows the shape

relationships of the kernels between the two image domains in the previous methods, where the relatively fixed integration kernel is adopted. From this figure, we can observe if the expected integration scale matrix $\Sigma_{I,new}^{-1}$ for the next iteration is proportional to the current second moment matrix μ_{cut}, then there is always an apparent shape difference between the expected integration kernel (shown in blue ellipse) and the current integration kernel (shown in green ellipse) in the normalized image domain. Such shape difference makes the obtained second moment matrix μ_{new} in the next iteration always anisotropic with a great probability.



(a) The current image domain **(b)** The normalized image domain

In **(a)**, the blue filled elliptic region is determined by the current second moment matrix; the green circle represents the shape of current integration kernel; the blue ellipse refers to the expected integration kernel; **(b)** illustrates shapes of the matrices in the normalized image domain, the color labels correspond to **(a)**. As we can see that under the smoothing of the Gaussian kernel labeled with green color, the local image structure is transformed into an isotropic structure region shown in the blue filled circle.

Fig. 1. The shape relations between the kernels when the relatively fixed integration kernel is adopted

The above conclusions can be proved as follows:

Proof:

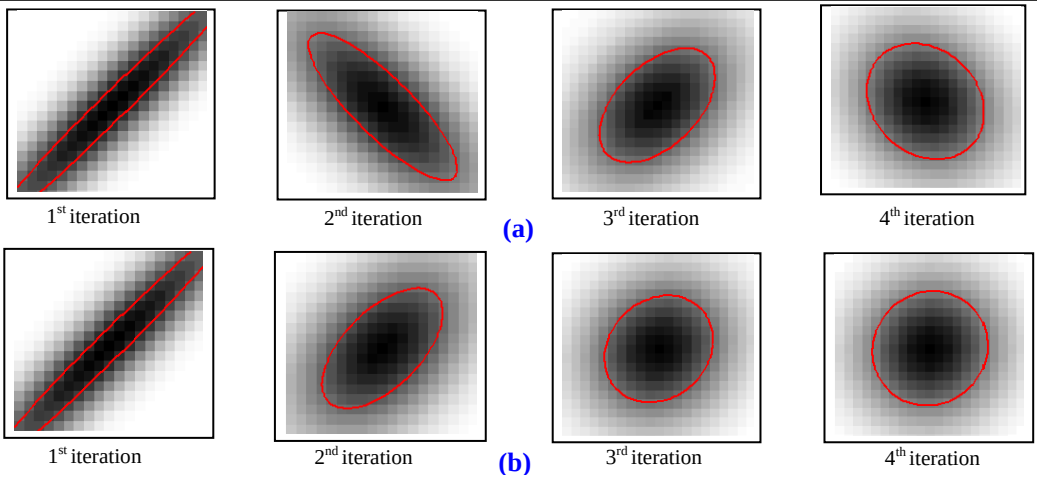
Suppose $\mathbf{x}_{norm} = \mu_{cut}^{1/2}\mathbf{x}$, based on (6), we can get:

$$\mu_{cut} = \left(\mu_{cut}^{1/2}\right)^T \mu_{norm} \left(\mu_{cut}^{1/2}\right) \Rightarrow \mu_{norm} = \left(\mu_{cut}^{-1/2}\right)^T \mu_{cut} \left(\mu_{cut}^{-1/2}\right) = \mathbf{I} \tag{9}$$

$$\Sigma_{I,cut,norm} = s\left(\mu_{cut}^{1/2}\right)\mathbf{I}\left(\mu_{cut}^{1/2}\right)^T = \mu_{cut} \Rightarrow \Sigma_{I,cut,norm}^{-1} = s\mu_{cut}^{-1}, \quad \left(s \in \Re^+\right) \tag{10}$$

$$\Sigma_{I,new,norm} = \left(\mu_{cut}^{1/2}\right)\Sigma_{I,new}\left(\mu_{cut}^{1/2}\right)^T \Rightarrow \Sigma_{I,new,norm}^{-1} = t\mathbf{I}, \quad \left(t \in \Re^+\right) \tag{11}$$

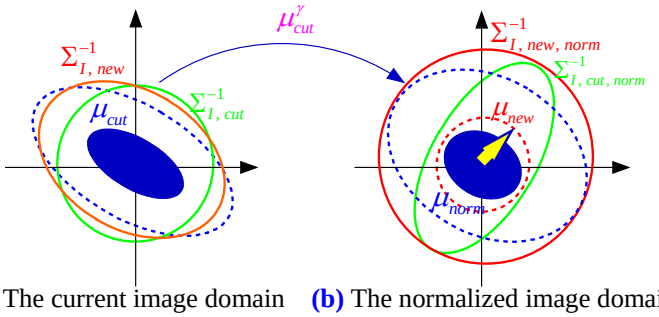
The subscript *cut* represents the current matrices; *norm* refers to the forms of the current matrices in the normalized image domain and *new* denotes the expected matrices in the next iteration. From equation (9), as can be observed, μ_{norm} is always isotropic when using the Gaussian kernel with $\Sigma_{I,cut,norm}$. Based on this conclusion, it's reasonable to predict that the aspect of μ_{new}, which is calculated using the Gaussian kernel with $\Sigma_{I,new,norm}$, will be in accordance with the aspect of μ_{cut}^{-1}. This can be confirmed by **Fig. 2-(a)**, where the affine iteration procedures of an ideal blob structure are illustrated with Hessian matrix as measure matrix (affine Hessian matrix is also affine covariant which is proved in Appendix A). In **Fig. 2-(a)**, the aspects between the ellipses in consecutive iterations are always perpendicular. We call this phenomenon overshoot.



(a) The affine iteration process of an ideal blob structure with the relatively fixed integration kernel.
(b) The affine iteration process of an ideal blob structure with the adaptive integration kernel.
The red ellipses represent the anisotropy of the local image structure measured by Hessian matrix.

Fig. 2. The iteration process of an ideal blob structure with two different methods

Intuitively, the greater the value of $\lambda_{\max}(\mu_{cut})/\lambda_{\min}(\mu_{cut})$ is, the larger anisotropy of μ_{new} may be. This phenomenon can also be observed in **Fig. 2-(a)**. Therefore, when the anisotropy of μ_{cut} is larger, the iteration process may fall into a state of oscillation, thereby increasing the possibility of divergence.



(a) The current image domain **(b)** The normalized image domain

In **(a)**, the blue filled elliptic region is determined by the current second moment matrix; the green circle represents the shape of the current integration kernel; the blue dotted ellipse refers to the expected integration kernel in previous methods; the red ellipse denotes the candidate integration kernel of the proposed method; **(b)** illustrates shapes of the matrices in the normalized image domain ($\gamma < 0.5$), the color labels correspond to **(a)**. The local structure is still an anisotropic structure region under the smoothing of the green Gaussian kernel, as shown in the blue filled circle. However, at this time, if the local structure is smoothed by the red integration kernel, it is probable for it to be close to an isotropic structure shown in the dotted red circle.

Fig. 3. The shape relations between the kernels when the adaptive integration kernel is used

Based on the above analysis and comparison, we propose a method which is able to estimate the shape of expected integration kernel and adjust the convergence pace. As shown in **Fig. 3-(a)**, if $\Sigma_{I,new}^{-1}$ can share the shape similar to the red ellipse, then in the normalized image

domain the probability $p(\lambda_{\max}(\mu_{new})/\lambda_{\min}(\mu_{new}) \to 1)$ is greater than the one in previous methods. In this case, μ_{norm} is no longer isotropic and its aspect has become consistent with that of μ_{cur}. This is illustrated in blue filled elliptic region in **Fig. 3-(b)**. Under the smoothing of the Gaussian kernel with $\Sigma_{I,new,norm}$, the shape of μ_{new} is probable to approach to a circle, as shown in red dotted circle in **Fig. 3-(b)**.

The second row of **Fig. 2** shows the iteration procedures when using the proposed method. From the figure, we can observe that the overshoot phenomenon has been effectively suppressed and the convergence process becomes smoother. Compared to previous methods, the anisotropy of the blob structure is getting smaller in each step when our method is adopted. In the following, we will discuss more details of our method.

3.2.2 Approximate Estimation of the Shape of Integration Kernel

First, we use the eigenvalue ratio ξ [3] to measure the anisotropy of μ:

$$\xi = \frac{\lambda_{\max}(\mu)}{\lambda_{\min}(\mu)} \qquad \xi \in [1, \infty] \tag{12}$$

When ξ is equal to 1, the local image structure is a perfect isotropic structure.

As can be observed from **Fig. 1-(b)**, if there is linear relationship between the expected integration kernel and the shape of μ_{cur} with a large anisotropy, then μ_{new} probably has a large anisotropy. To such integration kernel, it's reasonable to assign it a low confidence level. Furthermore, we also expect that the estimated integration kernel can continue to reduce the anisotropy of the local image structure. Based on these criteria, the shapes of $\Sigma_{I,new}^{-1} = t\mu_{cut}^{\eta}, \eta \in [0.5,1]$ (shown in blue ellipses in **Fig. 4**) can meet our requirements very well.

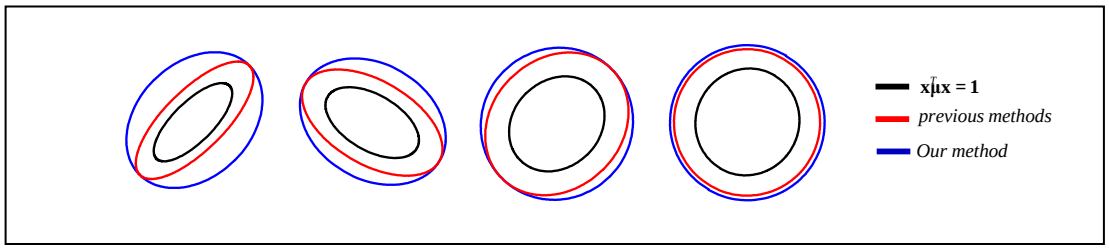


Fig. 4. The comparison of the integration kernels in iteration process between our method and the previous methods

Same as [3], our iterative shape adaptation scheme also works in the normalized image domain. Assume that the estimated integration kernel with $\Sigma_{I,new}^{-1} = t\mu_{cut}^{\eta}$ can be transformed into an isotropic kernel using a linear transformation μ_{cur}^{γ} (γ is called normalization exponent here). Since μ_{cur} is a positive definite symmetric matrix, then the following equation holds:

$$\eta = 2\gamma \qquad \gamma \in [0.25, 0.5] \tag{13}$$

According to equation (13), we propose two functions to estimate the normalization exponent γ directly.

$$\gamma = F_1(\xi) = \begin{cases} 0.5 & \xi \in [1,1.5] \\ \dfrac{(\vartheta - 0.5)(\xi - 1.5)}{\tau - 1.5} + 0.5 & \xi \in [1.5, \tau] \\ \vartheta & others \end{cases} \tag{14}$$

$$\gamma = F_2(\xi) = \begin{cases} \dfrac{(\vartheta - 0.5)}{(\tau - 1)^2}(\xi - 1)^2 + 0.5 & \xi \in [1, \tau] \\ \vartheta & others \end{cases} \tag{15}$$

Their figures are shown in **Fig. 5**. For the function F_1, we assume that γ is piecewise linear with ξ while the relationship between them becomes nonlinear in F_2. In these two figures, when ξ approaches to 1, there is a sufficient margin in which γ approaches to 0.5. One reason for the above assumptions is that such forms enable the extracted local image regions to better meet the fixed point condition (see Section 3.2.5).

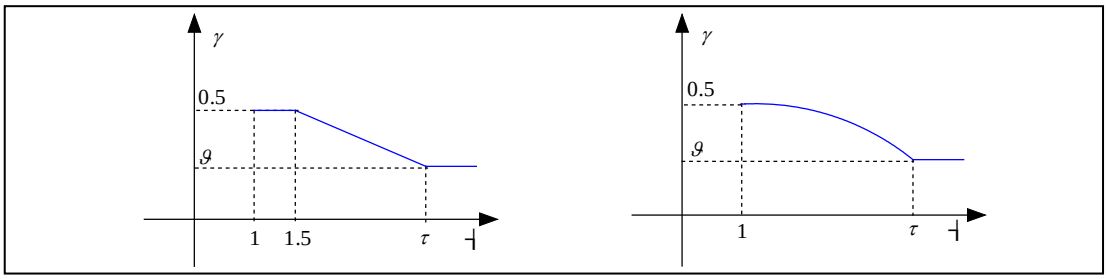


Fig. 5. The diagrams of the two estimation functions

Since γ is less than 0.5, the shape of μ_{norm} is transformed into an anisotropic structure illustrated in **Fig. 3-(b)** with the blue elliptical region. The diagram of the shape evolution of μ_{norm} in the normalized image domains can be observed in **Fig. 6**.

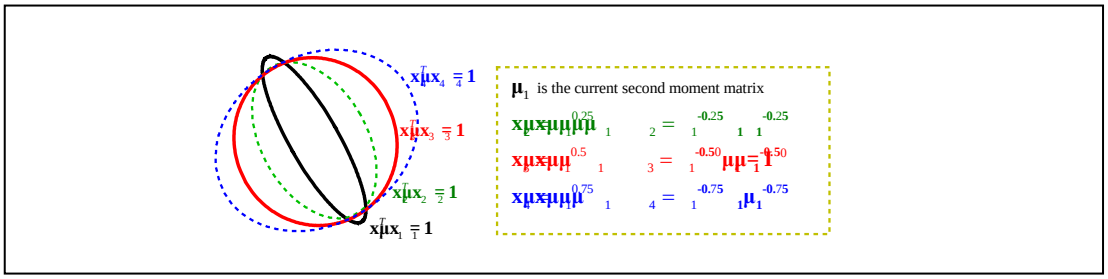


Fig. 6. The shape evolution of the second moment matrix in the normalized image domain

3.2.3 Modified Estimation of the Shape of Integration Kernel

In our approach, γ is estimated in each iteration. However, due to the noise effect, γ may have small variations. To make γ immune against to noise, we combine the current γ and the last γ to achieve a modified estimation of γ:

$$\gamma = (\gamma_{cur},\ \gamma_{last})\mathbf{w},\ \|\mathbf{w}\|_1 = 1 \tag{16}$$

3.2.4 Shape Adaptation Matrix

The proposed scheme is carried out in the transformed image domain. Often it requires several iterations before convergence. The second moment matrix μ is computed in each iteration. Based on its anisotropy, γ can be assigned a reasonable value by the above proposed method. Combining them, we can obtain a staged transformation matrix μ^{γ}. The final shape adaptation matrix $\mathbf{M}$, which is applied to the transformation of the neighborhoods of feature points in original image, can be achieved by concatenating these staged transformation matrices. The matrix $\mathbf{M}$ obtained in k^{th} step of affine iteration is shown as follows:

$$\mathbf{M}^{(k)} = \prod_{i=1,\,k} \mu^{\gamma_i}\mathbf{M}^{(0)}, \quad \gamma_i \in [0.25, 0.5] \tag{17}$$

If the affine shape adaptation converges, then:

$$\mathbf{M} \approx \mathbf{M}^{(k)} \tag{18}$$

3.2.5 Integration Scale Matrix

Although each staged transformation matrix is positive definite symmetric, the finally obtained transformation matrix $\mathbf{M}$ is not a symmetric matrix. The expected second moment matrix and integration scale matrix for the feature point assume the following forms:

$$\mu = \mathbf{M}^T\mathbf{M} \tag{19}$$

$$\Sigma_I^{-1} = t\mathbf{M}^T\mathbf{M}, \qquad t \in \Re^+ \tag{20}$$

In order to fulfill the fixed-point property (7), the obtained $\mu = \mathbf{M}^T\mathbf{M}$ must be relatively affine invariant.

Proposition 1: If the normalization exponent γ in the last staged transformation matrix is equal to 0.5, then $\left(\mu,\ \Sigma_I^{-1}\right)$ is a fixed-point. Here, the obtained feature point is assumed to undergo all the staged affine deformations $\mu^{\gamma_i}, i \in (1, k)$, k is the iterative number for convergence.

Proof:

Insert $\mathbf{M}^{(k)} = \prod_{i=1,\,k} \mu^{\gamma_i}\mathbf{M}^{(0)}$, $\left(\mathbf{M}^{(0)} = \mathbf{I}\right)$ into (19), it follows that:

$$\mu = \mathbf{M}^T\mathbf{M} = \left(\prod_{i=1,\,k} \mu^{\gamma_i}\mathbf{M}^{(0)}\right)^T \left(\prod_{i=1,\,k} \mu^{\gamma_i}\mathbf{M}^{(0)}\right) = \mathbf{M}^{(0)}\left(\prod_{i=k-1,\,1} \left(\mu^{\gamma_i}\right)^T\right)\left(\mu^{\gamma_k}\right)^T \mu^{\gamma_k}\left(\prod_{i=1,\,k-1} \mu^{\gamma_i}\right)\mathbf{M}^{(0)}$$

Enforcing that $\mathbf{B} = \left(\prod_{i=1,\,k-1} \mu^{\gamma_i}\right)\mathbf{M}^{(0)}$, we have

$$\mu = \mathbf{B}^T\left(\mu^{\gamma_k}\right)^T \mu^{\gamma_k}\mathbf{B} \tag{21}$$

If $\gamma_k = 0.5$, then

$$\mu = \mathbf{B}^T\mu_k\mathbf{B} \ \Rightarrow\ \mu_k = \mathbf{B}^{-T}\mu\mathbf{B}^{-1} \tag{22}$$

Inserting (22) into (20), we get

$$\Sigma_I^{-1} = t\mathbf{M}^T\mathbf{M} = t\mu = t\mathbf{B}^T\mu_k\mathbf{B} \ \Rightarrow\ \Sigma_I = \mathbf{B}^{-1}\left(t\mu_k\right)^{-1}\mathbf{B}^{-T} = \mathbf{B}^{-1}\Sigma_k\mathbf{B}^{-T} \tag{23}$$

Combining (22), (23), (19) and (20), we can obtain that ($\mu = \mathbf{M}^T\mathbf{M}$, $\Sigma_I^{-1} = t\mathbf{M}^T\mathbf{M}$) is a

fixed-point.

In practice, if $\lim_{i \to k} \gamma_i = 0.5$, then the fixed-point can be obtained with sufficient accuracy. The problem has been considered in the proposed algorithm by letting γ_k be close to 0.5 in the later iterations (see **Fig. 5**).

Proposition 2: If the local regions around points $\mathbf{x}_L$ and $\mathbf{x}_R$ are linked by an affine transformation $\mathbf{A}$ and are normalized by shape adaptation matrices $\mathbf{x}_{L,norm} = \mathbf{M}_L \mathbf{x}_L$ and $\mathbf{x}_{R,norm} = \mathbf{M}_R \mathbf{x}_R$, then the relation between these two normalizations is up to an orthogonal transformation $\mathbf{R}$ and $\mathbf{A} = \mathbf{M}_R^{-1} \mathbf{R} \mathbf{M}_L$.

Proof:

Based on equation (6), we can obtain:

$$\mu_L = \mathbf{M}_L^T \mu_{L,norm} \mathbf{M}_L \Rightarrow \mu_{L,norm} = \mathbf{M}_L^{-T} \mu_L \mathbf{M}_L^{-1} = \mathbf{M}_L^{-T} \mathbf{M}_L^T \mathbf{M}_L \mathbf{M}_L^{-1} = \mathbf{I}$$

In the same way, we can get: $\mu_{R,norm} = \mathbf{I}$.

If $\mathbf{x}_{R,norm} = \mathbf{B}\mathbf{x}_{L,norm}$, then $\mu_{L,norm} = \mathbf{B}^T \mu_{R,norm} \mathbf{B} \Rightarrow \mathbf{B}^T \mathbf{B} = \mathbf{I}$.

Therefore, $\mathbf{B}$ is an orthogonal transformation $\mathbf{R}$.

Furthermore, $\mathbf{x}_{R,norm} = \mathbf{R}\mathbf{x}_{L,norm} \Rightarrow \mathbf{M}_R \mathbf{x}_R = \mathbf{R}\mathbf{M}_L \mathbf{x}_L \Rightarrow \mathbf{A} = \mathbf{M}_R^{-1} \mathbf{R} \mathbf{M}_L$.

Thus, the proposition 2 has been proved.

The affine normalization process can be observed in **Fig. 7**. The ambiguity of rotation can be removed by other geometric constraints, such as the dominant direction [8].

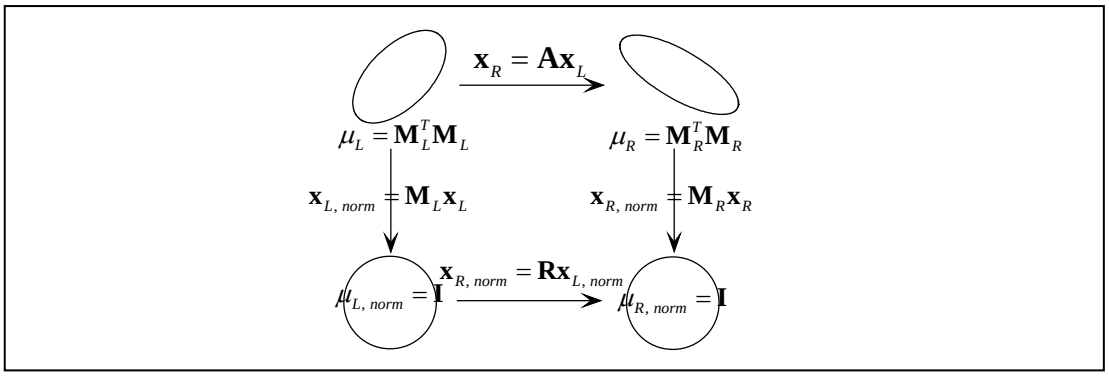


Fig. 7. Diagram illustrates the affine normalization based on the second moment matrices. Image coordinates are transformed with the shape adaptation matrix **M**. The transformed images are related by an orthogonal transformation

3.2.6 Convergence Criteria

We stop iterating when the matrix μ_{new} is sufficiently close to a pure rotation and favors the fixed-point condition. Referring to [3], we propose the following convergence criteria:

$$\frac{\lambda_{\max}(\mu_{new})}{\lambda_{\min}(\mu_{new})} < 1.05 \quad \text{and} \quad \left\| \gamma_k - 0.5 \right\|_2 < 0.1 \tag{24}$$

where k is the number of iteration.

4. Evaluation for Image Matching

4.1 Data Set

Fig. 8. Data set. Viewpoint change for structured scene **(a)**, and for textured scene **(b)**; scale and rotation for structured scene **(c)**; blur for textured scene **(d)**; **(e)** light change; **(f)** JPEG compression. In the experimental comparisons, the left most image of each set is used as the reference image

In this section, we evaluate adaptive kernel-shape based affine invariant detector and compare it with the relatively fixed kernel detectors on a standard dataset, which comes from [19], a popular dataset for the evaluation of local feature properties. The dataset consists of eight image sequences including textured-images and structured-images. There are different deformations among these images, like viewpoint change, scale and rotation change, light change, blur, and JPEG compression. The test images of the standard dataset used in the experiments are shown in **Fig. 8**. These image pairs are either planar scene or captured from fixed-position camera during acquisition. Thus, the relation between them can be modeled by a 2D homography matrix. In this way, the image set has also provided homographies for some pair of images.

4.2 Evaluation Metric

We adopt the metrics in [17] to evaluate the performance of affine detectors. More specifically, the following three metrics are adopted:

(1) Correct match metric : $1-\frac{R_{\mu_a} \bigcap \mathbf{A}^T R_{\mu_b} \mathbf{A}}{(R_{\mu_a} \bigcup \mathbf{A}^T R_{\mu_b} \mathbf{A})} < \varepsilon$

where R_μ represents the elliptic region defined by $\mathbf{x}^T \mu \mathbf{x} = 1$; $\mathbf{A}$ is a locally affine transformation of the homography between the two images; $R_{\mu_a} \bigcap \mathbf{A}^T R_{\mu_b} \mathbf{A}$ and $R_{\mu_a} \bigcup \mathbf{A}^T R_{\mu_b} \mathbf{A}$ refer to the area intersection and area union respectively; ε represents the overlap error.

(2) Repeatability:

$$\text{Repeatability} = \frac{\#\text{of correspondences}}{\min\left(\#\text{of correspondences } (\text{image1}), \#\text{of correspondences } (\text{image 2})\right)}$$

(3) Convergence ratio:

$$\text{convergence ratio} = \frac{\#\text{of feature points of convergence}}{\#\text{of total feature points}}$$

4.3 Experimental Results

We compare the proposed method to previous method in two schemes: multi-scale Harris points and using the second moment matrix as affine shape adaptation; multi-scale Hessian points and employing the Hessian matrix as affine shape adaptation. Both of the two schemes employ the same matrices in feature point extraction and affine shape adaptation, so the shape adaptation matrices definitely contain sufficient information of the local image structures. The proposed approach is based on the prediction of the local image structure, so it naturally fits in such framework. We have corrected some errors and bugs of the code of Laptev (*http://www.nada.kth.se/~laptev/code.html*) and employ it for the next experiments. 1200 candidate feature points are first extracted at three scales: 2, 6 and 10. Then these points proceed to scale selection, location modification and affine shape adaptation, and finally converge to fixed-points. To reduce the complexity of the algorithm, we choose $\Sigma_I = 2\Sigma_D$ across iterations. We aim at comparing the performance of relatively fixed kernel and adaptive kernel. Simplifying relevant steps is also helpful to observe the performance variation between them. The test programs are running on Intel(R) Xeon(R) E5520 2.26GHz.

4.4.1 Scheme Selection

There are three unknowns required to be determined in our scheme: the estimation function, two thresholds ϑ and τ, and weighting vector $\mathbf{w}$. In general, if ϑ and τ are larger, the estimation functions will reduce the ability to adjust iteration process. In contrast, if ϑ and τ are smaller, this is likely to cause premature convergence of the iterative algorithm. As a result, the obtained image regions cannot satisfy the fixed-point property (7). Hence it is important for our scheme to select proper parameters. We determine these unknowns by referring to [3] and doing experiments. For a given estimation function, we fix a threshold and just vary the other one. The repeatability and matching score on the standard image set [19] are used as evaluation metric. In these experiments, we have found that the performance of our method is not very sensitive to these parameters. We propose to use the linear function with $\vartheta = 0.25$, $\tau \in (6,13)$ and the quadratic function with $\vartheta = 0.25$, $\tau \in (6,9)$. As for $\mathbf{w}$, it is proposed that $w_1 > w_2$. In the next experiment, the quadratic function with $\vartheta = 0.25$, $\tau = 6$, and $\mathbf{w} = (0.9, 0.1)^T$ is selected as the candidate scheme.

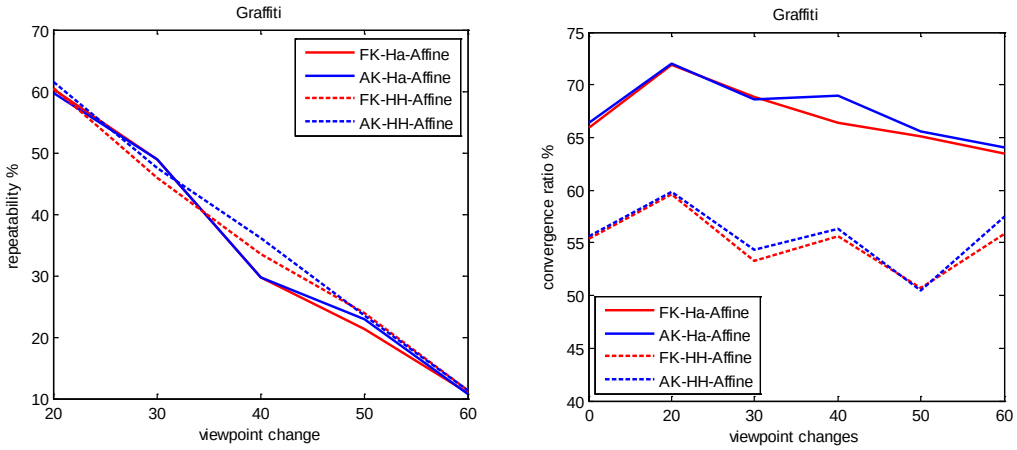
4.4.2 Experimental Results

Fig. 9, **Fig. 10**, **Fig. 11**, **Fig. 12**, **Fig. 13**, and **Fig. 14** show the experiment results of our method on the different image pairs on the standard dataset. In **Fig. 9** and **Fig. 10**, we show the comparative results of two types of methods: the methods based on relatively fixed kernel shape (FK-Ha-Affine and FK-HH-Affine) and the methods based on adaptive kernel shape (AK-HH-Affine and AK-HH-Affine), under affine transformation. The symbol 'Ha-Affine' represents Harris points with the second moment matrix for affine adaptation and 'HH-Affine' refers to Hessian points with the Hessian matrix for affine adaptation. **Fig. 11** shows the result under scale change and rotation. Moreover, **Fig. 12** shows the performance comparison of methods under significant amount of image blur in the structured scene. Finally, the comparative results under illumination change and JPEG compression are shown in **Fig. 13** and **Fig. 14**. In all of these **Fig.**, plot (a) shows repeatability and plot (b) the corresponding convergence ratio.

As can be seen from these plots, in most cases, AK- based methods achieve better convergence ratio and repeatability compared with FK-based methods. This demonstrates the effectiveness of the proposed method on improving the convergence stability of Harris-Affine although the performance improvement is not very significant.

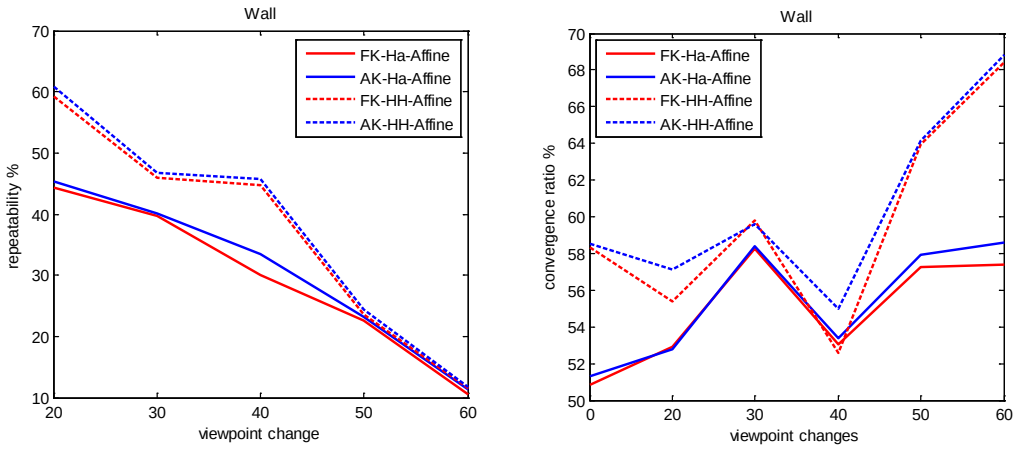
Moreover, in **Fig. 15**, we also show the experiment results of descriptor matching score which are used to evaluate the distinctiveness of local image regions obtained by various kinds of detectors [17]. Our methods and Harris-Affine extract same local image regions. Their difference is only residing in affine shape adaptation. Thus, the scores of descriptor matching and repeatability should be similar between them. The plots in **Fig. 15** confirm this inference, and the experiment results are in accordance with conclusions in [10][17].

Finally, in terms of the number of iteration, our methods achieve comparable iteration times compared to the FK-based methods, as can be shown in **Fig. 16**. Each data mark in the plots represents the total iteration times from all the feature points in one image sequence (6 images). In addition, from the view of computation complexity, our approach mainly increases a linear discrimination process to each iteration, which assumes a very low computation cost and can be fully neglected compared to the Gaussian smoothing of scale-space in each iterative process.



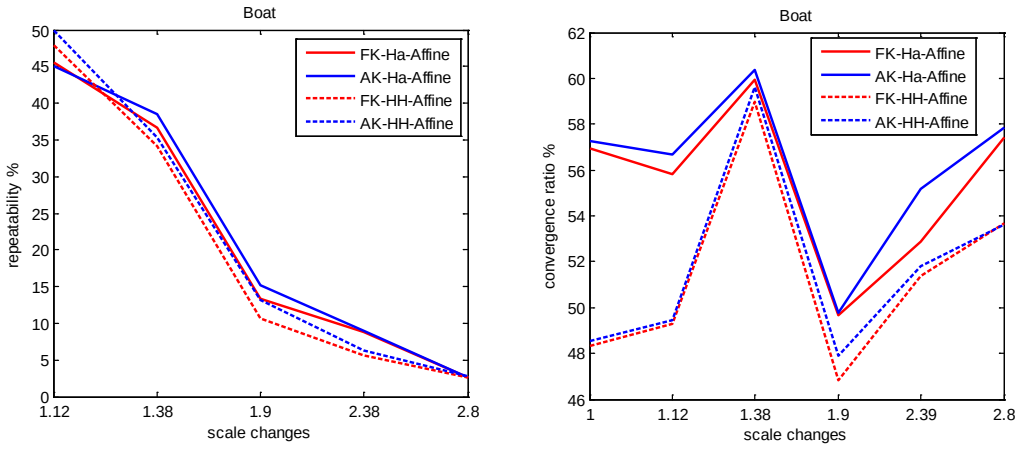
(a) **(b)**

Fig. 9. Viewpoint change for structured scene **(a)** repeatability **(b)** convergence ratio



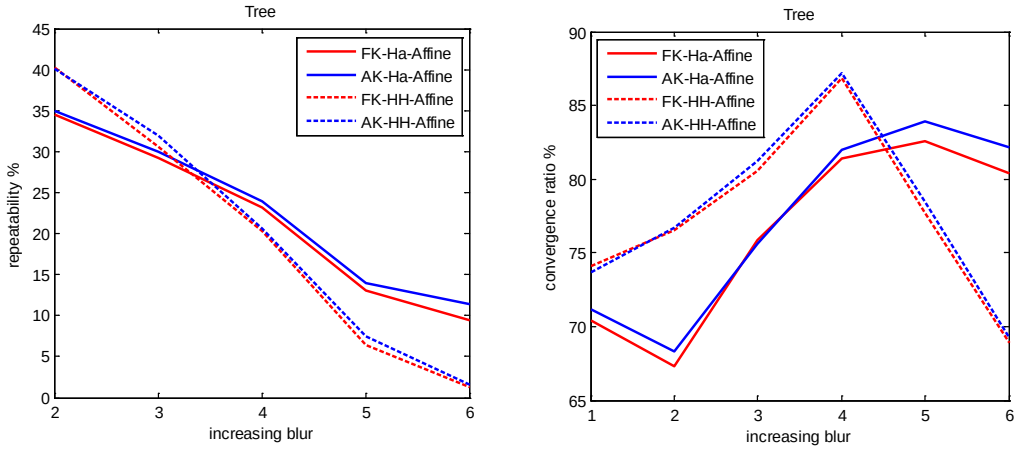
(a) **(b)**

Fig. 10. Viewpoint change for textured scene **(a)** repeatability **(b)** convergence atio



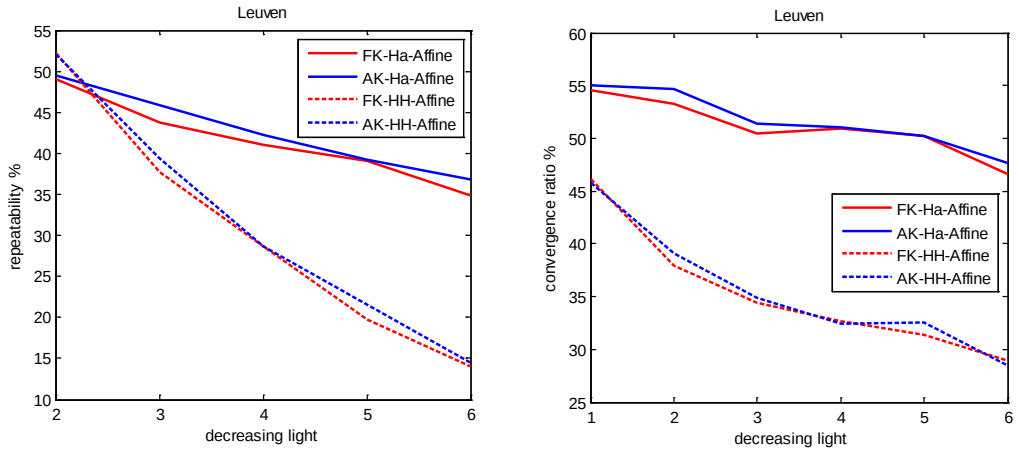
(a) **(b)**

Fig. 11. Scale rotation for structured scene **(a)** repeatability **(b)** convergence ratio



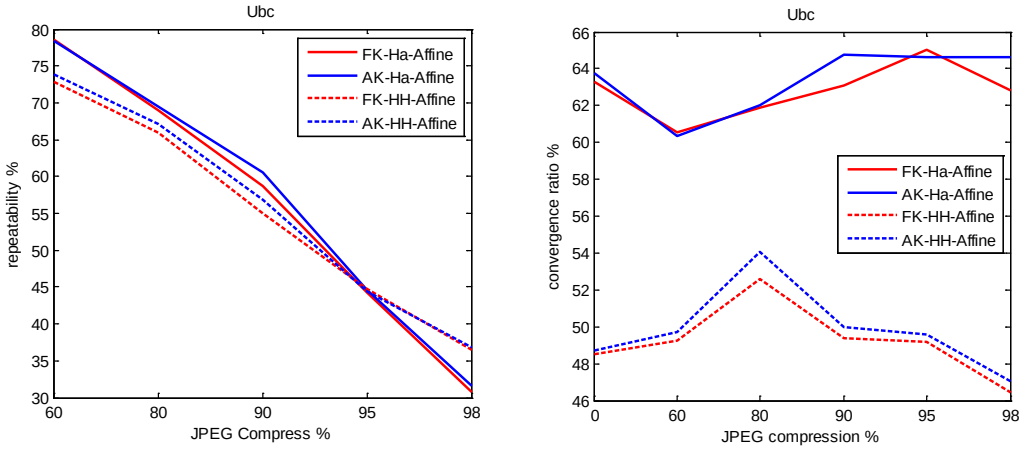
(a) **(b)**

Fig. 12. Blur for textured scene **(a)** repeatability **(b)** convergence ratio



(a) **(b)**

Fig. 13. Light change **(a)** repeatability **(b)** convergence ratio



(a) **(b)**

Fig. 14. JPEG compression **(a)** repeatability **(b)** convergence ratio

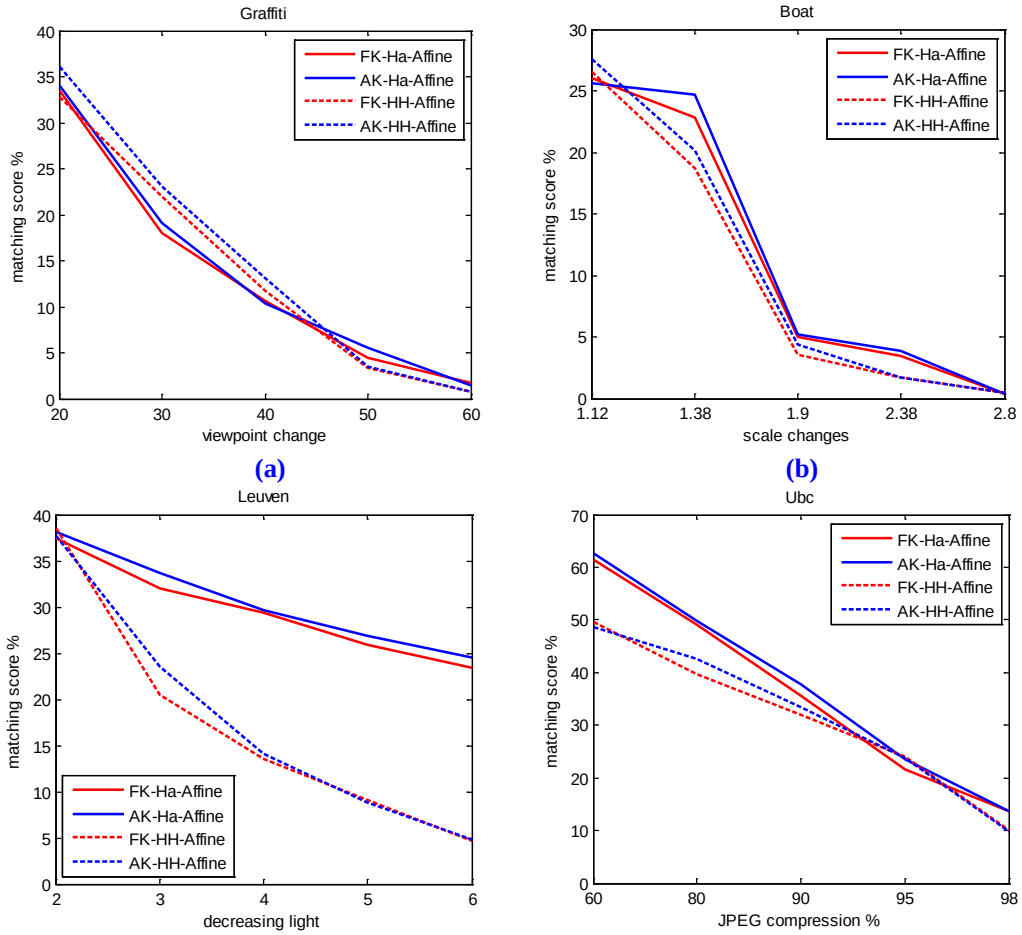


Fig. 15. Matching score for **(a)** viewpoint change **(b)** scale changes light change **(d)** JPEG compression

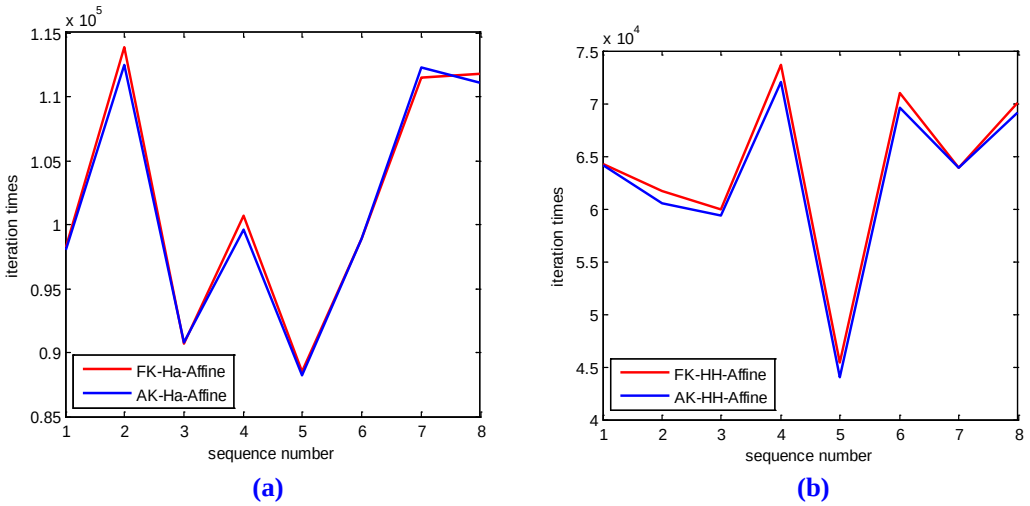


Fig. 16. Iteration times comparison **(a)** Ha-Affine **(b)** HH-Affine

5. Conclusions

In this paper, we have described how the overshoot phenomenon may arise when keeping the integration kernel shape linearly related to the shape determined by the second moment matrix. Especially when the anisotropy of the second moment matrix is larger, the overshoot phenomenon may lead to the oscillation of iterative process, thereby increasing the possibility of divergence of the iterative algorithm.

A method for reducing the problem is presented by introducing a new iterative scheme with adaptive integration kernel. In the proposed scheme, the shape of the integration kernel in each iteration is determined by the anisotropy of the second moment matrix. To this end, we define two criteria for estimating the integration kernel: first, if an integration kernel is in accordance with a second moment matrix which shares a large anisotropy, then the integration kernel is given a low confidence level. Second, the estimated integration kernel is beneficial to reduce the anisotropy of the second moment matrix. Based on the above criteria, two functions (the one is linear and the other is non-linear) have been proposed to estimate and modify the shape of the integration kernel in each iteration. Finally, the establishment condition of fixed-point property for the local image regions obtained by the proposed scheme is also presented. To satisfy the condition, we have also taken some specific measures embodied in the design of the estimation functions and the final convergence criteria. In synthetic image data, as can be observed, the overshoot phenomenon of second moment matrix has been effectively suppressed when our approach is adopted.

Whereas the discussion of the article is concerned with the second moment matrix, the underlying ideas can be extended to other structure metric matrices as long as they are affine invariant, i.e. affine Hessian matrix. Our method naturally fits in the framework that the feature detection matrix is same as the metric matrix. When the matrices are different, our method can still keep our performance advantage.

Experiment results on the standard dataset demonstrate that we obtain small improvement in convergence ratio and repeatability as compared with Harris-Affine and Hessian-Hessian-Affine. In the future, we will further explore how to further improve the convergence stability by building a reasonable switching system based on the location changes of feature points and the aspect alterations of the integration scale matrices.

Acknowledgment

We would like to thank Laptev for his valuable code.

References

[1] C. Harris and M. Stephens, “A combined corner and edge detector,” in *Proc. of Alvey Vision Conf.*, pp. 189-192, 1988. Article (CrossRef Link)

[2] C. Schmid, R. Mohr and C. Bauckhage, “Evaluation of interest point detectors,” *International Journal of Computer Vision*, vol. 37, no. 2, pp. 151-172, 2000. Article (CrossRef Link)

[3] K. Mikolajczyk and C. Schmid, “Scale & affine invariant interest point detectors,” *International Journal of Computer Vision*, vol. 60, no. 1, pp. 63-86, 2004. Article (CrossRef Link)

[4] T. Lindeberg and J. Garding, “Shape-adapted smoothing in estimation of 3-D shape cues from affine deformations of local 2-D brightness structure,” *Image and Vision Computing*, vol. 15, no. 6, pp. 415-434, 1997. Article (CrossRef Link)

[5] A. Baumberg, "Reliable feature matching across widely separated views," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, Hilton Head Island, South Carolina, USA, pp. 774-781, 2000. Article (CrossRef Link)
[6] T. Tuytelaars and K. Mikolajczyk. "Local invariant feature detectors: A survey," *Foundations and Trends in Computer Graphics and Vision*, vol. 3, no. 3, pp. 177-280, 2008. Article (CrossRef Link)
[7] T. Lindeberg, "Feature detection with automatic scale selection," *International Journal of Computer Vision*, vol. 30, no. 2, pp. 79-116, 1998. Article (CrossRef Link)
[8] D.G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91-110, 2004. Article (CrossRef Link)
[9] H. Bay, T. Tuytelaars and L.V. Gool, "SURF: speeded up robust features," in *Proc. of European Conf. on Computer Vision*, pp.404-417, 2006. Article (CrossRef Link)
[10] G. Dorkó and C. Schmid, "Maximally stable local description for scale selection," in *Proc. of European Conference on Computer Vision*, 2006. Article (CrossRef Link)
[11] Wei-Ting Lee and Hwann-Tzong Chen, "Histogram-based interest point detectors, " in *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, 2009. Article (CrossRef Link)
[12] Wolfgang F¨orstner, Timo Dickscheid and Falko Schindler, "Detecting interpretable and accurate scale-invariant keypoints," in *Proc. of the IEEE Conf. on Computer vision and Pattern Recognition*, Kyoto, Japan, 2009. Article (CrossRef Link)
[13] Wolfgang Förstner, "A framework for low level feature extraction," in *Proc. of European conf. on computer vision*, Stockholm, Sweden, vol. 3, pp. 383-394, 1994. Article (CrossRef Link)
[14] J. Matas, O. Chum and M. Urban, and T. Pajdla, "Robust wide baseline stereo from maximally stable extremal regions," *IVC*, vol. 22, no. 10, pp. 761-767, 2004. Article (CrossRef Link)
[15] T. Tuytelaars and L.V. Gool, "Matching widely separated views based on affine invariant regions," *International Journal of Computer Vision*, vol. 59, no. 1, pp. 61-85, 2004. Article (CrossRef Link)
[16] Timor Kadir, Andrew Zisserman and Michael Brady, "An affine invariant salient region detector," in *Proc. of European Conference on Computer Vision*, pp.228-241, 2004. Article (CrossRef Link)
[17] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir and L.V. Gool, "A comparison of affine region detectors," *International Journal of Computer Vision*, vol. 65, no. 1/2, pp. 43-72, 2005. Article (CrossRef Link)
[18] F. Schaffalitzky and A. Zisserman, "Multi-view matching for unordered image sets," in *Proc. of European Conference on Computer Vision*, pp. 414-431, 2002. Article (CrossRef Link)
[19] http://www.robots.ox.ac.uk/vgg/research/affine/

Appendix A. The affine invariance of affine Hessian matrix

Affine Hessian matrix defined in affine Gaussian scale space is as follows

$$H = \nabla(\nabla L(\mathbf{x};\Sigma_D)) \tag{25}$$

which can be used to describe the anisotropic shape of blob-like structures and is affine invariant.

Proof:

Assume that two patches are related by an affine transformation $\mathbf{x}_R = \mathbf{A}\mathbf{x}_L$, and then the relation between their affine Gaussian scale-space representations is shown as follows:

$$L(\mathbf{x}_L;\Sigma_L) = R(\mathbf{x}_R;\Sigma_R) \tag{26}$$

then:

$$\nabla L(\mathbf{x}_L;\Sigma_L) = \mathbf{A}^T\nabla R(\mathbf{A}\mathbf{x}_L;\mathbf{A}\Sigma_L\mathbf{A}^T) = \mathbf{A}^T\left(\nabla R_1(\mathbf{A}\mathbf{x}_L;\mathbf{A}\Sigma_L\mathbf{A}^T),\nabla R_2(\mathbf{A}\mathbf{x}_L;\mathbf{A}\Sigma_L\mathbf{A}^T)\right)^T$$

$$H_L = \nabla(\nabla L(\mathbf{x}_L;\Sigma_L)) = \mathbf{A}^T \begin{Bmatrix} \left(\mathbf{A}^T\nabla\left(\nabla R_1(\mathbf{A}\mathbf{x}_L;\mathbf{A}\Sigma_L\mathbf{A}^T)\right)\right)^T \\ \left(\mathbf{A}^T\nabla\left(\nabla R_2(\mathbf{A}\mathbf{x}_L;\mathbf{A}\Sigma_L\mathbf{A}^T)\right)\right)^T \end{Bmatrix} = \mathbf{A}^T \begin{Bmatrix} \left(\nabla\left(\nabla R_1(\mathbf{A}\mathbf{x}_L;\mathbf{A}\Sigma_L\mathbf{A}^T)\right)\right)^T \\ \left(\nabla\left(\nabla R_2(\mathbf{A}\mathbf{x}_L;\mathbf{A}\Sigma_L\mathbf{A}^T)\right)\right)^T \end{Bmatrix} \mathbf{A}$$

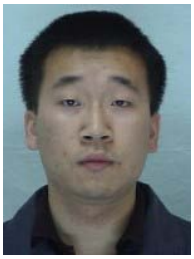
$$= \mathbf{A}^T\nabla(\nabla R(\mathbf{A}\mathbf{x}_L;\mathbf{A}\Sigma_L\mathbf{A}^T))\mathbf{A} = \mathbf{A}^T H_R \mathbf{A}$$

This proves that affine Hessian matrix is affine invariant.

Congxin Liu received a B.S. degree from Wuhan University of Hydraulic and Electrical Engineering (Yi Chang), China, in 1997, a M.S. degree from Three Gorges University, China, in 2004. He is currently a Ph.D. student in the Department of Electronic and Electrical Engineering in Shanghai Jiao Tong University. His research interests include local invariant feature and image matching.

Jie Yang received a Ph.D. degree in computer science from the University of Hamburg, Germany in 1994. Dr Yang is now the professor of Institute of Image Processing & Pattern Recognition in Shanghai Jiao Tong University. He has taken charge of many research projects (e.g. National Science Foundation, 863 National High Tech. Plan) and published one book in Germany and more than 200 journal papers. His major research interests are image retrieval, object detection and recognition, data mining, and medical image processing.

Yue Zhou is an associate professor at the Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University, China. His research interests include visual surveillance, object detection, and pattern analysis.



Deying Feng received a B.S. degree from Shandong University of Technology, China, in 2005, a M.S. degree from Shanghai Maritime University, China, in 2008. He is now a Ph.D. student in Shanghai Jiao Tong University, China. His research interests include image retrieval and similarity search.