

Image registration method based on improved Harris corner detector

Qi Zeng (曾琦)*, Liu Liu (刘浏), and Jianxun Li (李建勋)

Automation Department, School of Electronic & Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China

*E-mail: zqmark@gmail.com

Received November 2, 2009

Harris corner detector is a classic tool to extract feature. It is stable to illumination change and rotation but unstable to more complicated transform. In order to register images with different viewpoints, we extend Harris corner detector to scale-space to gain invariance to scale change, then we apply affine shape adaptation to the scale invariant point until convergence is reached, giving it invariance to affine transform. With these local features, we use general feature descriptor and matching algorithm to generate matches and then use the matches to calculate the geometric transform matrix, which enables the final registration. Result shows that our algorithm can get more accurate matches than scale invariant feature transform SIFT, and less difference exists between registered images.

OCIS codes: 100.2000, 100.5010.

doi: 10.3788/COL20100806.0573.

Image registration is a process of finding a proper geometric transform between two images that can align corresponding points in them. It is the foundation of applications, such as image fusion, medical image processing, and three-dimensional (3D) image reconstruction and is widely used in medical imaging and remote sensing. Recently, image registration has been a topic widely discussed, and methods with high efficiency and accuracy have been developed. For example, Guizar-Sicairos *et al.* proposed to use nonlinear optimization and matrix-multiply discrete Fourier transforms to register two-dimensional (2D) images with sub-pixel accuracy^[1]. In medical image registration, Modersitzki proposed to integrate the concept of local rigidity to the Flexible Image Registration Toolbox (FLIRT), giving extra constraint for rigid object in non-rigid optimization process^[2].

In medical image registration, many methods, like the maximization of mutual information (MMI) method^[3], treat the registration process as an optimization problem and use this measure as an object function to find the best transform. However, because of the diversity of image registration problem, one framework for image registration is to assume that the transform between images is of a certain kind (rotation, scaling, etc.), then to use local feature matches to calculate the transform matrix. For example, when translation, rotation, and scaling are present, a registration method was proposed to deal with that situation^[4]. In this framework, local feature matching is of utmost importance. A good local feature should generally have a clear mathematically well-founded definition, also, the local image structure around the local feature is rich in terms of local information contents, such as derivative information^[5], curvature information, etc. Most importantly, a good local feature should be tolerant to image noise, changes in illumination, scaling, rotation, as well as changes in viewpoint.

The most classic local feature is Harris corner^[6] detector, it is stable to illumination change and rotation but unstable to more complicated transform. Scale invariant

feature transform (SIFT)^[7] detector, which takes advantage of the scale invariant nature of scale-space representation, is stable to scale change but not very good at locating corners which often correspond to significant local structures. In this letter, we propose a method that can register images with affine and scale transform. Beginning with traditional Harris corner detector, we extend it to scale-space, giving the detector invariance to scale transform, and then we apply affine shape adaptation, giving the detector invariance to affine transform. We use default descriptor and matching algorithm to generate matches, and use those matches to calculate geometrical transform parameters. Finally, we transform one image using the geometrical transformation matrix to align with the other image.

Geometric transform between images is the foundation in our image registration. To describe the transform, we introduce projective geometry^[8]. Unlike traditional Cartesian coordinate, a point in projective coordinate is defined as a vector of three elements: the first two are x and y , and the third coordinate is introduced to deal with the situation of infinite point. The third element is 1 when the point is not infinite and 0 when the point is infinite. Because when divided by 0, any small value becomes infinite, using projective coordinate makes an infinite point homogeneous with any other point in the space. A geometric transform can be written in matrix form as

$$\begin{pmatrix} x' \\ y' \\ x'_3 \end{pmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{pmatrix} x \\ y \\ x_3 \end{pmatrix}. \quad (1)$$

A geometric transform can be further divided to isometries, similarity transforms, and affine transforms. Here we focus on affine transforms because of its ubiquity:

$$\begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} = \begin{bmatrix} a_{11} & a_{12} & t_x \\ a_{21} & a_{22} & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}. \quad (2)$$

Affine transform has 6 degrees of freedom (DOFs) (there are 6 variables in the transform matrix). To solve a linear system with 6 variables, we need at least 3 matches.

Local feature is the local description of a specified structure or characteristic of image. Sometimes local feature is referred to as interest point, key point, etc. Classical Harris corner point is defined as a point which has great variance of intensity in all directions.

Without loss of generality, we assume that a gray-scale 2D image is denoted as $I(x, y)$. Taking an image patch over the area (u, v) and shifting it by (x, y) , the weighted sum of square difference between these two patches, S , is given by

$$S(x, y) = (x, y) \mathbf{A} \begin{pmatrix} x \\ y \end{pmatrix}, \tag{3}$$

$$\mathbf{A} = \sum_u \sum_v w(u, v) \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix},$$

where the matrix \mathbf{A} is referred to as the second moment matrix, it consists of the first order partial derivatives of the intensity function of the image $I(x, y)$, denoted as

$$I_d = \frac{\partial}{\partial d} I(x, y) \quad (d = x, y).$$

Considering that the contribution of each point in the image patch surrounding the corner point should be different, a weighing function $w(u, v)$ is used to assign bigger weight to points which are closer to the corner point and smaller weight to those at the boundary. Typically a Gaussian function is chosen as the weigh function:

$$w(u, v) = e^{-(u^2+v^2)/2\sigma^2}.$$

According to Ref. [6], a measure for a point's likelihood of being a corner is introduced as

$$M_c = \lambda_1 \lambda_2 - \kappa (\lambda_1 + \lambda_2)^2 = |\mathbf{A}| - \kappa [\text{Tr}(\mathbf{A})]^2, \tag{4}$$

where \mathbf{A} is a nonsingular 2×2 matrix, thus it must have two eigenvalues λ_1 , and λ_2 . Using the determinant and trace of the matrix \mathbf{A} and an empirical value κ (usually κ is assigned a value between 0.04 and 0.15), we can calculate a measure for a point's likelihood of being a corner.

When we have calculated all the M_c measures for each point in an image, we can choose those with M_c measures above a certain threshold as Harris corner points. Classical Harris point detector is a reliable local feature extractor. It is stable to illumination change and rotation, but it is not stable to affine and scale transforms.

Because of its inability to deal with affine and scale transforms, we introduce an improved Harris detector.

Considering the scale transform, we choose scale as a new dimension in addition to spatial dimension. The motivation for this choice is that real-world objects are composed of different structures at different scales, in contrast to idealized mathematical entities such as points or lines, they appear in different ways depending on the scale of observation. For example, the concept of a tree is appropriate at the scale of meters, while concepts such as leaves and molecules are more appropriate at finer scales.

One way to represent an image $I(x, y)$ as its scale space representation $L(x, y, t)$ with the addition of the scale coordinate t is to take advantage of the low-pass nature of Gaussian filter:

$$g(x, y, t) = \frac{1}{2\pi t} e^{-(x^2+y^2)/2t}. \tag{5}$$

Thus we can generate a Gaussian scale-space representation, formulated as

$$L(x, y, t) = g(x, y, t) \times I(x, y). \tag{6}$$

In order to apply the traditional Harris corner detector at every scale of the Gaussian multi-scale representation $L(x, y, t)$, the second moment matrix is redefined as

$$\mu(x, y, t, s) = g(x, y, s) \times \begin{bmatrix} L_x^2(x, y, t) & L_x(x, y, t) L_y(x, y, t) \\ L_x(x, y, t) L_y(x, y, t) & L_y^2(x, y, t) \end{bmatrix}, \tag{7}$$

where $g(x, y, s)$ is a Gaussian function with s being the integration scale parameter, $L_x(x, y, t)$ and $L_y(x, y, t)$ are the first order partial derivatives of image $L(x, y, t)$. According to Ref. [9], the integration scale parameter s and the scale coordinate should maintain $s = \gamma^2 t$, where γ is an empirical value usually chosen in the interval $[\sqrt{2}, 2]$.

Using the second moment matrix given in Eq. (7) and the measure for a point's likelihood of being a corner in Eq. (4), we apply traditional Harris detector at each scale and get multi-scale Harris corner detector, naturally invariant to scale transform.

The multi-scale Harris corner detector is a scale-invariant detector. However, it is not invariant to arbitrary affine transform. According to Ref. [10], affine shape adaptation is a methodology for iteratively adapting the shape of the smoothing kernels in an affine kernel group to the local image structure in neighborhood region of a specific image point. Provided that this iterative process converges, the resulting fixed point will be affine invariant. The initial points of the algorithm come from multi-scale Harris corner detection. Hence, the convergent points are stable to both scale and affine transforms.

For convenience of description, in the following part, interest point (x, y, t) is denoted as x or x_w . Given an initial point $x^{(0)} = (x, y, t)$, the iteration process is as follows.

Step 1: Initialize a transform matrix $\mathbf{U}^{(0)}$ to identity matrix; let the number of iteration $k \geq 0$. A local window W is centered at the interest point x and transformed by transform matrix \mathbf{U} which is initialized by identity matrix and updated by concatenating an additional square root of the second moment matrix at each iteration step:

$$\mathbf{U} = \prod_k (\mu^{-\frac{1}{2}})^{(k)} \mathbf{U}.$$

Step 2: Normalize the window $W(x_w) = I(x)$ centered on $x_w^{(k-1)} = [\mathbf{U}^{(k-1)}]^{-1} x^{(k-1)}$. The iterative affine shape adaptation method works in the transformed image domain. The local window centered on the interest point $x^{(k)}$ is transformed by \mathbf{U} , and this operation is called U-transformation. At first, the matrix \mathbf{U} is initialized by the identity matrix, and then, through each iteration, it

is concatenated with an additional square root of the second moment matrix. In the k th iteration, $x_w^{(k-1)}$ refers to the interest point in the transformed domain, which is specified by the subscript w . Obviously, the initial value is $x_w^{(0)} = (\mathbf{U}^{(0)})^{-1}x^{(0)} = x^{(0)}$.

Step 3: Select integration scale:

$$\sigma_I^{(k)} = \arg \max_{\substack{\sigma_I = t\sigma_I^{(k-1)}, \\ t \in [0.7, 1.4]}} \sigma_I^2 [L_{xx}(x_w^{(k-1)}, \sigma_I) + L_{yy}(x_w^{(k-1)}, \sigma_I)],$$

$$L_{dd} = \frac{\partial^2}{\partial d^2} [g(x, y, t) * I(x, y)] \quad (d = x, y).$$

The normalized Laplacian^[10] is used here to find the integration scale by finding the maximum over scale. The finding process is also an iterative process: by multiplying t , the scale parameter is altered at each iteration at random direction until the maximum of normalized Laplacian is reached. Similar process is used in the next step, and we omit the implementation details for simplicity consideration.

Step 4: Select the differentiation scale:

$$\sigma_D^{(k)} = \arg \max_{\substack{\sigma_D = s\sigma_I^{(k)}, \\ s \in [0.5, 0.75]}} \frac{\lambda_{\min}[\mu(x_w^{(k-1)}, \sigma_I^{(k)}, \sigma_D)]}{\lambda_{\max}[\mu(x_w^{(k-1)}, \sigma_I^{(k)}, \sigma_D)]},$$

where $\lambda_{\min}[\mu(x_w^{(k-1)}, \sigma_I^{(k)}, \sigma_D)]$ and $\lambda_{\max}[\mu(x_w^{(k-1)}, \sigma_I^{(k)}, \sigma_D)]$ are the minimum and maximum eigenvalues of 2×2 matrix $\mu(x_w^{(k-1)}, \sigma_I^{(k)}, \sigma_D)$. The differentiation scale is determined by finding the maximum of normalized isotropy over scale, which is the ratio of the two eigenvalues of the second moment matrix.

Step 5: Spatial localization:

$$x_w^{(k)} = \arg \max_{x_w \in W(x_w^{(k-1)})} \left(\left| \mu(x_w, \sigma_I^{(k)}, \sigma_D^{(k)}) \right| - \kappa [\text{Tr}(\mu(x_w, \sigma_I^{(k)}, \sigma_D^{(k)}))]^2 \right).$$

According to the definition of the second moment matrix μ in Eq. (7), four parameters are needed to calculate it. Firstly, the spatial location, i.e., x and y are given by $x_w^{(k-1)}$; then the integration scale is determined by choosing the maximum over scale of the normalized Laplacian; at last, the differentiation scale is determined by choosing the maximum of normalized isotropy. Having μ , we detect the spatial localization $x_w^{(k)}$ by find the maximum of the Harris measure nearest to $x_w^{(k-1)}$.

Step 6: Transform back to the original reference frame:

$$x^{(k)} = x^{(k-1)} + \mathbf{U}^{(k-1)}(x_w^{(k)} - x_w^{(k-1)}).$$

Given the second moment matrix μ , the new point $x_w^{(k)}$ and the old one $x_w^{(k-1)}$, a new interest point $x^{(k)}$ can be found. We obtain a vector of displacement to the nearest maximum in the U-normalized window W . The location of the initial point is corrected with the displacement vector back-transformed to the original image domain.

Step 7: Compute the square root of the second moment matrix $\mu_i^{(k)} = \mu^{-\frac{1}{2}}(x_w^{(k)}, \sigma_I^{(k)}, \sigma_D^{(k)})$; prepare the square

root of the second moment matrix for \mathbf{U} update.

Step 8: Concatenate transformation $\mathbf{U}^{(k)} = \mu_i^{(k)}\mathbf{U}^{(k-1)}$ and normalize $\mathbf{U}^{(k)}$ to $\lambda_{\max}(\mathbf{U}^{(k)})=1$. In this step, the matrix U is updated for the next iteration.

Step 9: Go to Step 2 if $1 - \lambda_{\min}(\mu_i^{(k)})/\lambda_{\max}(\mu_i^{(k)}) \geq 0.05$; otherwise stop, so that convergence is achieved.

By checking the local isotropy of the second moment matrix $\mu_i^{(k)}$ and comparing it with a tolerance of error, we can decide whether convergence is achieved or go on iteration. The local isotropy can be measured by the smaller eigenvalue of the square root of the second moment matrix divided by the bigger one.

Until now, we have extended the traditional Harris corner detector to multi-scale Harris corner detector followed by an iterative search for affine invariant feature. Theoretically, the improved Harris corner detector should be able to detect local features in scale and affine transformed images. Next we present the experimental results.

We choose graffiti images to test our algorithm and compare it with traditional registration methods. The images were deliberately taken from different angles, creating a natural affine transform. We first took one of the images from straight on, then we moved to the left of the first position and took the second one which was to be registered. The corresponding points were shown in the same color. They all laid in areas with significant angle structure and were more or less different because of the shifted viewpoint.

Figure 1 show the Harris corner scattering patterns before and after the affine shape adaptation. In Fig. 1(a), five layers of Harris points are detected using the multi-scale Harris corner detector; however, if all five layers of points are projected into the bottom layer, points at different layers would stack upon each other and form several point clusters which are usually related to one single characteristic structure. This can cause inaccurate corner location as well as unnecessary computation.

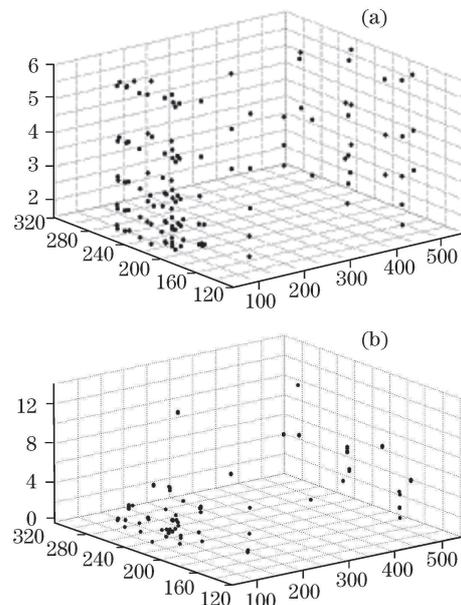


Fig. 1. Harris corner scattering patterns (a) before and (b) after affine shape adaptation.



Fig. 2. Matching results from (a) SIFT and (b) our algorithms.

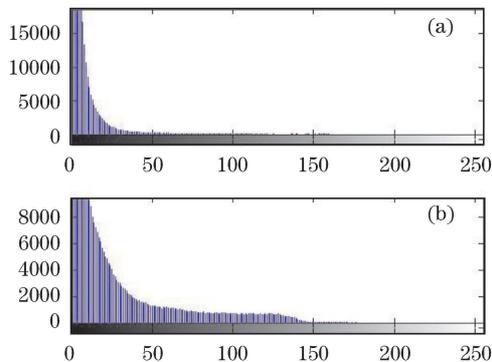


Fig. 3. Histograms of difference. (a) Our algorithm and (b) SIFT algorithm.

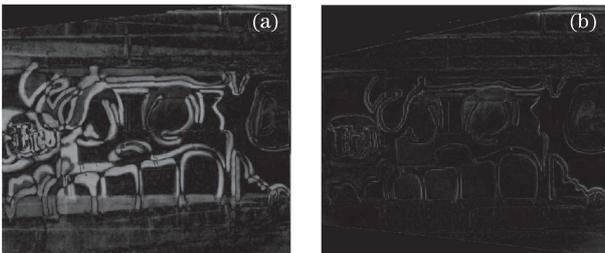


Fig. 4. Difference images of (a) SIFT and (b) our algorithms.

In Fig. 1(b), after applying the affine shape adaptation, many point clusters disappear and only several points are left, which is the result of the iterative process that leads to convergence. These convergent points correspond to the characteristic structure more accurately and also save computation burden.

Figure 2 show the matching points from SIFT algorithm and our algorithm. As we can see, SIFT gener-

ates too many interest points and matches and needs further processing. At the same time, our algorithm generates less key points and matches but with high correctness, which makes the subsequent calculation of transform very promising. Besides, our algorithm extracts interest points from corner areas while SIFT has no preference to corner point. This allow us to gain better accuracy than SIFT.

Figure 3 shows the histograms of difference of overlapping parts between the registered image and original image, Figure 4 give the difference images. As we can see, our algorithm has most of its elements lying in the lower half, while SIFT has some of its element spreading from 0 to over 150. Our algorithm has achieved better registration result.

In conclusion, we present an image registration method based on improved Harris corner detector. We assume the transform to be scale and affine transforms, and focus on improving the traditional Harris corner detector, giving it invariance to scale and affine transforms. Using the detector, we find the corresponding point match and achieve satisfactory results.

This work was jointly supported by the National Natural Science Foundation of China (Nos. 60935001 and 60874104), the Space Foundation of Supporting-Technology (No. 2008-HT-SHJD003), the Cultivation Fund of the Key Scientific Technical Project Ministry of Education of China (No. 706022), the National “973” Project of China (No. 2009CB824900), and the Shanghai Key Basic Research Foundation (No. 08JC1411800).

References

1. M. Guizar-Sicairos, S. T. Thurman, and J. R. Fienup, *Opt. Lett.* **33**, 156 (2008).
2. J. Modersitzki and W. M. Wells III, *Int. J. Computer Vision* **76**, 153 (2008).
3. P. Viola *Int. J. Computer Vision* **24**, 137 (1997).
4. L. Li, Q. Zeng, and F. Meng, *Chin. Opt. Lett.* **6**, 827 (2008).
5. C. Han, J. Li, X. Chen, and Z. Zhu, *Chin. Opt. Lett.* **6**, 334 (2008).
6. C. Harris and M. Stephens, in *Proceedings of the Fourth Alvey Vision Conference* 147 (1988).
7. D. G. Lowe, *Int. J. Computer Vision* **60**, 91 (2004).
8. R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision* (2nd edn.) (Cambridge University Press, Cambridge, 2003) chap. 2.
9. K. Mikolajczyk and C. Schmid, *Int. J. Computer Vision* **60**, 63 (2004).
10. T. Lindeberg and J. Garding, *Image and Vision Computing* **15**, 415 (1997).