A Feature Matching Algorithm Based on an Illumination
and Affine Invariant for Aerial Image Registration

Zhaoxia Liu, Yaxuan Wang, Yu Jing
School of Software, Dalian University of Foreign Languages, 6th South Lvshun Road,
Lvshun Kou District, Dalian, 116044, China
E-mail: liu.zhaoxia0524@gmail.com

Received: 29 October 2013 /Accepted: 9 January 2014 /Published: 31 January 2014

Abstract: This paper presents a new feature matching algorithm for registering aerial images with large
illumination distortion, similar patterns, affine transformation and low overlapping areas. In this algorithm,
affine invariant regions are obtained by a K-NN graph and its corresponding adjacent graph. Then an
illumination and affine invariant called Illumination Invariant MultiScale Autoconvolution (IIMSA), which is
the combination of MSA and MultiScale Retinex (MSR), is defined to describe the triangle regions. Based on
the IIMSA, outliers are removed by a self-adaptive feature matching strategy. The performance of the proposed
algorithm is evaluated by registering both visible and thermal infrared (thermal-IR) images. Compared with
MSA, the accuracy of the proposed algorithm is highly improved. Copyright © 2014 IFSA Publishing, S. L.

Keywords: Feature descriptor, Illumination and affine invariant, Feature matching.

1. Introduction

Nowadays, image registration is being widely
applied in remote sensing, computer vision, pattern
recognition, medical image analysis and so on [1]. In
recent years, many feature-based image registration
techniques have been proposed in remote sensing
[2-8]. However, Aerial images captured on the sea
with a visible or thermal camera often have the
following characteristics. Firstly, because of the
changes in appearance and temperature of the object,
infrared images of the same object might be different
in content and illumination. Second, the images
captured on the sea contain a lot of similar features,
so these images have the monotonous characteristics.
In feature matching, mismatches often occur in
registering them. Last but not the least, serial images
may have low overlapping area. In this condition, it is
inevitable that there are many outliers in non-
overlapping area. Therefore mismatches are still
challenging problems in registering aerial images. In
this paper, we focus on the ambiguity in feature
matching caused by large illumination changes,
similar patterns and low overlapping areas and affine
transformation between images. For feature point
based matching algorithm, both descriptor and
feature matching strategy are critical. The descriptor
is commonly used for establishing the
correspondence between feature point sets. The most
remarkable descriptor is Scale-Invariant Feature
Transform (SIFT), which has been shown to be
invariant to image rotation, scale and affine
distortion, addition of noise, and change in
illumination [9]. However, it is still difficult to
handle images with large illumination changes, low
overlapping areas and monotonous backgrounds.

It is necessary to take into account the region
information around the feature points to deal with the
ambiguity in feature point matching. For area-based
matching algorithm, affine invariants are usually
obtained by edge-based or intensity-based information. Well known methods include Harris-affine detector, Hessian-affine detector [10, 11], maximally stable extremal region detector (MSER) [12], intensity extrema-based region detector (IBR) [13], edge-based region detector (EBR) [14] has been measured and the comparison result is that MSER obtains the highest score in many cases [15]. However, MSER can not be applied to the aerial images with low overlapping areas and monotonous backgrounds or similar feature. Because there are few overlapping closed sub-regions in those images, it is hard to find similar regions from them. Another affine invariant descriptor Multiscale Autoconvolution (MSA), which applied standard point-based invariants and combined with probabilistic ideas to describe regions, is proposed by Rahtu et al. [15]. However, in MSA, nonuniform illumination distortion is not considered sufficiently.

Considering the aforementioned pros and cons, we propose a new feature matching algorithm for registering images with affine transformation and large illumination distortion accurately and automatically. In this paper, triangle region, instead of arbitrary geometric shape or a small window centered at some feature point, is used for registering images with low overlapping areas and monotonous backgrounds or similar features. These triangle regions are obtained by K-nearest neighbors (K-NN) graph of initial matched result by SIFT. Then combining MSA and multiscale retinex (MSR), affine and illumination invariant called illumination invariant multiscale autoconvolution (IIMSA) is defined to describe triangle regions. In the IIMSA, by removing the illuminance of the image, reflective image is obtained and replaces the original image of MSA. Finally, a filtering strategy is designed to remove outliers one by one.

2. Multiscale Autoconvolution

Rahtu et al. presented an affine invariant image transform Multiscale Autoconvolution (MSA) [16], which is based on the idea that the corresponding regions have the same intensity distribution and their mathematical expectations are invariant to affine transformation as defined below.

Define the affine transformation  by \( \chi = A(x) = Tx + t \), where \( t, x \in \mathbb{R}^2 \) and \( T \) is a \( 2 \times 2 \) nonsingular matrix. Suppose \( f(x) \) is an image intensity function corresponding to a gray-scale image in \( IR^2 \). Let \( x_0, x_1, x_2 \in \mathbb{R}^2 \) be three points from the support of \( f \). Now given these points, a point \( u_{(a, \beta)} \) can be denoted as

\[
u_{(a, \beta)} = \alpha(x_1 - x_0) + \beta(x_2 - x_0) + x_0
\]

Similarly, \( \nu_{(a, \beta)} = \alpha(x', 0) + \beta(x', 0) + x_0 \), where \( x_0, x_1, x_2 \) are the corresponding points of \( x_0, x_1, x_2 \) after the affine transformation \( A \). Using the relation \( x' = A(x) = Tx + t \), we have

\[
u_{(a, \beta)} = \alpha(Tx_1 - Tx_0) + \beta(Tx_2 - Tx_0) + Tx_0 + t = A(u_{(a, \beta)})
\]

indicating that the point \( u_{(a, \beta)} \) is also the affine point of the point \( u_{(a, \beta)} \).

Instead of using the fixed points \( x_0, x_1, x_2, x_0, x_1, x_2 \) are assumed to be independent random variables with probability density function \( pX(x) = f(x) \). Then \( u_{(a, \beta)} \) and \( u_{(a, \beta)} \) will become a random variable

\[
u_{(a, \beta)} = \alpha(x_1 - x_0) + \beta(x_2 - x_0) + x_0
\]

Like Eq. (2), \( U_{(a, \beta)} = A(U_{(a, \beta)}) = UTU_{(a, \beta)} + t \), to eliminate the affine transformation, \( f(U_{(a, \beta)}) = f(A^{-1}(U_{(a, \beta)})) = f(A^{-1}(A(U_{(a, \beta)}))) = f(U_{(a, \beta)}) \) is obtained. Hence the random variables \( f(U_{(a, \beta)}) \) and \( f(U_{(a, \beta)}) \) have the same distribution, which is independent of the affine transformation \( A \).

The MSA transform of \( f \) is defined as the expected value of \( f(U_{(a, \beta)}) \).

\[
Mf(\alpha, \beta) = E[f(U_{(a, \beta)})] = \int_{\mathbb{R}} f(u)p_{U_{(a, \beta)}}(u)du
\]

Suppose \( \gamma = 1 - \alpha - \beta \), \( U_{(a, \beta)} = \alpha X_1 + \beta X_2 + \gamma X_0 \), from the properties of the probability density, we can get the Eq. (5).

\[
p_{U_{(a, \beta)}} = (p_{a} * p_{\beta} * p_{\gamma})(u)
\]

So, Eq. (4) is denoted as

\[
Mf(\alpha, \beta) = \int_{\mathbb{R}} f(u)p_{U_{(a, \beta)}}(u)du = \int_{\mathbb{R}} f(u)(p_{a} * p_{\beta} * p_{\gamma})(u)du
\]

The double convolution in Eq. (6) is computationally expensive, so Eq. (6) can be alternatively expressed in terms of the Fourier transform as

\[
Mf(\alpha, \beta) = \frac{1}{\mathcal{F}(0)} \int_{\mathbb{R}^2} \mathcal{F}^{-1}(-\xi) \mathcal{F}(\alpha\xi) \mathcal{F}(\beta\xi) \mathcal{F}(\gamma\xi)d\xi
\]

where \( \mathcal{F} \) is the Fourier transform of \( f \).
3. Definition of IIMSA Descriptor

In this section, we propose a new illumination and affine invariant descriptor by combining MSA and MSR, which is called Illumination Invariant MultiScale Autoconvolution (IIMSA). The proposed descriptor is effective and robust to registering images with illumination changes, affine transformation and low overlapping areas. The details are as follows.

3.1. Affine Region Generation

In this subsection, affine region is generated by graph structure. It is known that graph structure can be constructed by Delaunay triangulation. However, the initial triangulation graph pair is similar only if the nonrigid transform is slight, so it is not suitable for large affine transformation. Here, for affine transformation, K-NN graph and the corresponding adjacent graph are constructed.

Suppose two point sets \( P = \{p_1, p_2, \ldots, p_n\} \) and \( Q = \{q_1, q_2, \ldots, q_n\} \) are obtained from two affinely transformed images respectively, where \( p_i \) and \( q_i \) are the matched points by comparing the distance of the closest neighbor with that of the second-closest neighbor of SIFT descriptor [9]. Two graphs \( G_r = (V_r, E_r) \) and \( G'_{q'} = (V_{q'}, E_{q'}) \) are established in the following way. Where \( V_r = \{v_1, v_2, \ldots, v_n\} \), vertex \( v_j \) is the point \( p_i \). A non-directed edge \( e(i, j) \in E_r \) exists when \( p_i \) is one of the K-nearest neighbours of \( p_j \), which is denoted by \( p_j \in N(p_i) \). Accordingly the corresponding graph \( G'_{q'} = (V_{q'}, E_{q'}) \) is constructed as follows: the corresponding feature point \( q_j \) exists as vertex \( v'_j \), then \( V_{q'} = \{v'_1, v'_2, \ldots, v'_n\} \). Because the distance might be changed after affine transformation, a corresponding non-directed edge \( (i, j) \in E_{q'} \) exists no matter whether \( q_j \) is one of the K nearest neighbours of \( q_i \). Therefore, \( G'_{q'} \) may not be a KNN graph, it can be named adjacent graph, which is different from that of GTM [18]. Fig. 1 shows the two graphs \( G_r \) and \( G'_{q'} \) established according to the method described above, where \( G_r \) is the K-NN graphs and \( G'_{q'} \) is \( G_r \)'s corresponding adjacent graph. They are invariant to affine transformation.

The vertexes of the graphs are local extreme points. The intensity distribution around the points is heterogeneous and contains a lot of information, so are the triangles regions constituted by those points. Thus, the triangle regions in the two graphs are used for evaluating the similarity of the corresponding points. Suppose two triangle regions \( Tr(i) \) and \( Tr'(i) \) are obtained from the graphs \( G_r \) and \( G'_{q'} \) respectively.

Since the graph structures are invariant to affine transformation, the two triangle regions \( Tr(i) \) and \( Tr'(i) \) are considered as affine regions.

3.2. IIMSA Descriptor

After applying illumination changes \( G \) and affine transformation \( A \) to an image \( f(x) \), a new grayscale image in \( IR^2 \) with the image function \( g(x') \) are generated, where \( g(x') = f(A(G(x))) \), we call \( g \) illumination and affine transformed version of \( f \).

Here, we will give the definition of affine and illumination invariant descriptor. Eq. (7) is the integral of image intensity function, so it is sensitive to large nonuniform illumination changes.

In general, an image \( I(x, y) \) is regarded as product of reflectance \( R(x, y) \) and illuminance \( L(x, y) \), i.e. \( I(x, y) = R(x, y)L(x, y) \) [17]. Computing the reflectance and the illuminance fields from real images is an ill-posed problem. Therefore, various assumptions and simplifications about \( L \), or \( R \), or both are proposed with the attempt to solve the problem. Jobson proposed a Multiscale Surround Retinex (MSR) to get the reflectance \( R \) from real image \( I \), which is introduced in [19]. The equation is given as follows:

\[
R_{MSR} = \sum_{n=1}^{N} w_n R_n ,
\]

where \( R_n = \log I(x, y) - \log[f_n(x, y) * I(x, y)] \) is the retinex output of \( nth \) scale and \( f_n(x, y) = Ke^{-\frac{x^2+y^2}{2}} \) is the surround function. \( w_n \) is the weight associated with the \( nth \) scale. \( R_{MSR} \) is the MSR output and invariant to illumination changes to some extent, so it can be used for replacing the original image.

Combining Eq. (7) and Eq. (8), we can obtain the following equation.

\[
F(\alpha, \beta) = \frac{1}{R_{MSR}(0)^{d\epsilon}} \int_{IR^2} R_{MSR}^{-\epsilon} R_{MSR}^{-\epsilon} d\epsilon
\]
where \( R_{MSR}(c) \) is the Fourier transform of \( R_{MSR} \).

By varying \( a, \beta \) and \( \gamma \), a vector \( D = \{ F(a_1, \beta_1), \ldots, F(a_n, \beta_n) \} \) is obtained, which should satisfy \( a + \beta + \gamma = 1 \). \( D \) is named as Illumination Invariant MultiScale Autoconvolution (IMSA) and used as an illumination and affine descriptor to describe the corresponding triangle regions \( Tr(i) \) and \( Tr'(i) \), which are denoted by \( D(Tr(i)) \) and \( D(Tr'(i)) \) respectively. If all the points are matched in pairs, the \( D(Tr(i)) \) and \( D(Tr'(i)) \) should be the same. Otherwise, the difference between \( D(Tr(i)) \) and \( D(Tr'(i)) \) can be used for evaluating the similarity of the two corresponding triangle regions. Here, the similarity of triangle regions is calculated according to the following Eq. (10).

\[
S(Tr(i), Tr'(i)) = e^{\frac{(D(Tr(i)) - D(Tr'(i)))^2}{2}} \tag{10}
\]

4. Feature Point Matching

An initial one-to-one correspondence between two point sets is obtained by comparing the distance of the closest neighbor with that of the second-closest neighbor of SIFT descriptor [9]. Due to there are similar features and visual ambiguity between sensed image and reference image, the two initial point sets are mismatched inevitably. Compared with the feature points, triangle regions, which are obtained from the K-NN graph of the initial feature points, contain plenty of information and can be used for removing outliers. In this section, a self-adaptive feature matching strategy is proposed to remove outliers by comparing the similarity between corresponding triangle regions as follows.

If all the corresponding points are matched points, the intensity distribution in the corresponding triangles is consistent. If not, their similarity can be used for evaluating the similarity of their vertices. If points \( p_i \) and \( q_i \) are a pair and all of their K-NN are corresponding points, the corresponding triangles will be similar. Otherwise, the triangles around the outliers will be much different. Therefore, in this algorithm the similarity of corresponding points are evaluated by the mean similarity of the triangles around it. For one corresponding triangles \( Tr(j) \) and \( Tr'(j) \), their similarity \( S(Tr(j), Tr'(j)) \) can be calculated by (10), and then the similarity of \( p_i \) and \( q_i \) is obtained by the following Eq. (11).

\[
S_{point}(p_i, q_i) = \frac{\sum_{j=1}^{N} S(Tr(j), Tr'(j))}{N_i}, \tag{11}
\]

where \( N_i \) is the number of triangles around \( p_i \) and is also the number of triangles around \( q_i \). The larger the value of \( S_{point} \) is, the more similar the points are.

Based on the similarity, a self-adaptive feature matching strategy is presented in Fig. 2.

1. \( E_{cost} = Transformation(P, Q, \theta) \)
2. while \( E_{cost} \geq \beta \)
3. \( M_p = GetDistanceMatrix( P ); \)
4. \( M_q = GetDistanceMatrix( Q ); \)
5. \( [ G_r, G_q ] = ConstructGraph( K, P, Q, M_p, M_q ); \)
6. \( [ D(Tr), D(Tr') ] = Descriptor( G_r, G_q ); \)
7. \( [ S_{point} ] = SimilarityPoint( I_s, I_s, D(Tr), D(Tr') ); \)
8. \( (P, Q) = \{ (p_i, q_i) | S_{point}(i) > \min(S_{point}) \} \)
9. \( E_{cost} = Transformation(P, Q, \theta) \)
10. end

Fig. 2. Outline of self-adaptation matching strategy.

Global transformation error is defined before the introduction of the matching strategy. If surviving point sets \( P \) and \( Q \) are the matched point sets, transformation parameters \( \theta \) can be calculated. For every \( p_i \) and its corresponding point \( q_i \), we suppose

\[
q_i \approx T(p_i, \theta) \tag{12}
\]

The transformation error can be defined as Eq. (13).

\[
E = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left\| T(p_i, \theta) - q_i \right\|} \tag{13}
\]

The self-adaptive feature matching strategy can be described as follows:

At the beginning, the transformation error \( E_{cost} \) is initialized by the initial one-to-one correspondence. The distance between any two points in \( P \) is calculated and stored in the distance matrices \( M_p \) and the distance between any two points in \( Q \) is stored in \( M_q \). Then, the K-NN graph \( G_r \) and the corresponding adjacent graph \( G_q \) are constructed. Each triangle in the graph \( G_r \) is described by IMSA descriptor, which is stored in the vector \( d(Tr) \). Similarly, the descriptor of each triangle in the graph \( G_q \) is stored in the vector \( d(Tr') \). The mean value \( S_{point} \) of similarities between all triangles around \( p_i \) and that of \( q_i \) in the two graphs are calculated by their descriptors, and is taken as the similarity between \( p_i \) and \( q_i \). The points that have
After the outliers are removed, $E_{\text{out}}$ are updated. The procedure needs to be executed iteratively until $E_{\text{out}} \leq \sigma$, which means the transformation error is stable and accurate. The surviving point sets are the final matching result.

5. Experiment and Analysis

In this section, we use five pairs of aerial images to evaluate the performance of the proposed algorithm. Firstly, dataset and parameter setting are given. Secondly, the importance of removing the illuminance of image is illustrated. Thirdly, experimental results are presented and the results of IIMSA are compared with those of MSA. Finally, we analyze the convergence of the IIMSA.

5.1. Dataset and Parameter Setting

Among the five pairs of images, image pair 2 is the visible sea lettuce images taken at TsingTao, China just one month before the sailing competition in 2008 Summer Olympic Games. Image pair 1, 3-5 are the thermal-IR images captured by an airborne camera (ThermCAM S65) after the accident, the Panama-flagged cargo ship ‘AGIOS DIMITRIOS 1’ ran aground near Zhuhai, Guangdong Province, China in September 2009. As these images are taken on the water or on the bank under different situations, most of the images have illumination changes, affine transformation and similar patterns or low overlapping area.

In IIMSA, to describe triangle area, $(\alpha, \beta)$ is equal to $(0, -0.5)$, $(0.4, 0.8)$, $(-0.2, 0.9)$ and $(0.3, -0.5)$. In the feature matching process, two parameters, i.e. $E_{\text{out}}$ and $K$, are designed to make the algorithm converge to an optimal solution. The $E_{\text{out}}$’s threshold $\sigma$ need to be determined. For exact affine transformed images, the transformation error is small. In practice, especially for aerial images taken on the sea, object appearance may change slightly, so there might be some local distortion in the images, which may make the global transformation error larger. In this paper, the value of threshold $\sigma$ is set to 3 pixels and $K$ is set to 4 empirically. Experimental results show that it is feasible to match point sets from different images and the matching results are insensitive to the values of these parameters.

5.2. Experimental Results

In Fig. 3, the effect of removing illuminance in images is demonstrated. Fig. 3 (a) and (b) show images with illumination changes. (c) and (d) depict reflective images after the MSR algorithm is applied on Fig. 3 (a) and (b). Those images are described by MSA descriptor respectively, and their similarities are calculated. With large illumination change, the difference between (a) and (b) is very large and the similarity between (a) and (b) is $1.4489E-33$. After the illuminance is eliminated, the similarity between (c) and (d) rises to 0.73522. The similarity of the image pair with illumination changes is greatly improved after MSR algorithm is applied. We can also see that the reflective images (c) and (d) are very similar.

In the following, the performance of IIMSA and MSA are compared. Fig. 4 shows the initial graph structures and the matching results of IIMSA and MSA for thermal IR images (image pair 1) and visible images (image pair 2), respectively. The first image pair is the oil slick images taken on the coast after the ‘AGIOS DIMITRIOS 1’ accident. The second pair is the sea lettuce images with low overlapping area and contains a lot of outliers in non-overlapping area. Both of them are with illumination changes. From Fig. 4, it can be seen that IIMSA successfully matched these feature points from images with monotonous backgrounds, low overlapping area and contains a lot of outliers in non-overlapping area. However, MSA failed in this situation even for images with few outliers in the first pair. That is because the distinctive power of MSA is limited to nonuniform illumination changes. Fig. 5 shows an example of the graph transformation process for the first group, from iteration 0 (initial graphs) to iteration 4 (final identical graphs), with $K = 4$. The symbol ‘*’ in Fig. 5 represents outliers that have been removed, which is circled in different color.
transformation and similar patterns as shown in Fig. 6. In Fig. 6, accurate matching results are obtained by IIMSA and the distribution of the matched points is reasonable to get an accurate transformation parameter. Although the matching result of MSA is accurate, there are fewer matched points than that of IIMSA. So it seems that more inliers may be taken for outliers and removed due to the influence of illumination changes.

The images in Fig. 6 are further distorted in illumination as shown in Fig. 7 (image pair $3'-5'$). From the matching result in Fig. 7, IIMSA still performs well and is free from interference of illumination changes. However, MSA fails in this situation.

Finally, the convergence of the proposed algorithm IIMSA is analyzed. Fig. 8 presented the RMSE tendency while registering image pair 1 and 4. From these images, it can be seen that the global transformation error will always converge to a minimum and the number of iteration depends on the number of outliers. There are only four outliers in image pair 1, so the RMSE in Fig. 8 (a) converges quickly. On the other hand, as there are many outliers in the point set from image pair 4, it needs more iterations for matching point sets. Although more time-consuming on removing the outliers in the image pair 4, the proposed algorithm can always converge to a minimum in the end.

In Table 1 the numbers of matched feature points of IIMSA and MSA is presented and compared. The symbol “F” means that the corresponding method fails to match those point pairs. It is clear that the IIMSA algorithm obtains more feature points and achieves better performance in accuracy, especially when the image pairs have large nonuniform illumination changes.

![Fig. 4.](image-url) Fig. 4. Initial graph and matching results of IIMSA and MSA. (a) initial graphs of two image pairs; (b) final graphs of IIMSA; (c) final graphs of MSA.

![Fig. 5.](image-url) Fig. 5. Graph transformation process example after removing the points.
Fig. 6. Matching results of IIMSA and MSA. (a) final graphs of IIMSA; (b) final graphs of MSA.

Fig. 7. Matching results of IIMSA and MSA for further illuminate changed version of Fig. 6. (a) final graphs of IIMSA; (b) final graphs of MSA.

Fig. 8. RMSE tendency of IIMSA. (a) the tendency of image pair 1; (b) the tendency of image pair 4.
Table 1. Real matching point pairs and matching results.

<table>
<thead>
<tr>
<th>Image Pair</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>3'</th>
<th>4'</th>
<th>5'</th>
</tr>
</thead>
<tbody>
<tr>
<td>Real matched points</td>
<td>25</td>
<td>22</td>
<td>11</td>
<td>20</td>
<td>13</td>
<td>15</td>
<td>18</td>
<td>11</td>
</tr>
<tr>
<td>IIMSA</td>
<td>25</td>
<td>22</td>
<td>11</td>
<td>10</td>
<td>6</td>
<td>13</td>
<td>10</td>
<td>4</td>
</tr>
<tr>
<td>MSA</td>
<td>F</td>
<td>F</td>
<td>11</td>
<td>4</td>
<td>5</td>
<td>F</td>
<td>F</td>
<td>F</td>
</tr>
</tbody>
</table>

6. Conclusions

In this paper, a new feature matching algorithm for registering aerial images with large illumination distortion, affine transformation, similar features and low overlapping areas is proposed. In this algorithm, affine invariant regions are obtained by a K-NN graph and its corresponding adjacent graph. Then IIMSA, an illumination and affine invariant descriptor, is defined to evaluate the similarity of these regions. Finally, a filtering strategy is designed to remove outliers one by one. The performance of the proposed matching algorithm is tested by five groups of aerial images, experimental results demonstrate that the matching algorithm can match point pair sets exactly even with many outliers. Compared with MSA, IIMSA works better on precision and is more robust to illumination changes.

Acknowledgements

This work is supported in part by the National Natural Science Foundation (No. 61201454) and in part by the Scientific Research Project of Liaoning Province Educational Department (No. L2013432) of P. R. China.

References

[8]. B. Li, W. Wang, H. Ye, Multi-sensor image registration based on algebraic projective invariants, *Optics Express*, 21, 8, pp. 9824-9838.