

Research on Infrared and Visible Images Registration Algorithm Based on Graph

Xiao-Lin ZHU^{1,a}, Ying-Guang HAO¹, and Hong-Yu WANG¹

¹*Dalian University of Technology, School of Information and Communication Engineering, Dalian, China*

Abstract. In this paper, we proposed a registration algorithm based on the combination of pyramid hierarchical idea and graph theory. It can solve the challenging problem of extracting consistent characteristics and matching the infrared and visible images. First, extracting the maximally stable extremal regions (MSERs) in the determining maximum down-sampling images and each MSER is represented by a polygon. Then constructing the graph features and the mapping relationship of MSERs between the infrared and visible images are determined by the graph matching method. Next we can construct the initial point set for matching according to the mapping relationship. Finally, using the random sample consensus (RANSAC) algorithm to obtain the optimal parameters and determine the error evaluation parameters. According to the idea of pyramid stratification, the above process is repeated in the high resolution images under the constraint condition of current matching error. The experiment results show that the algorithm can make full use of the visual similarity structures between images, and can achieve a smaller matching error under the premise of ensuring the robustness of the matching.

1 Introduction

Image registration is the process of aligning two or more images in the same scene of different times, different angles and different sensors. It has been widely used in military, medical, remote sensing, pattern recognition, computer vision and other fields [1-3]. The registration of infrared (IR) and visible images (VI) is an important branch in the field of multi-sensor registration. At present, the IR and VI registration algorithms mentioned in the literature can be divided into two categories: the method based on image gray and the method based on feature [4-5]. The gray level registration algorithm is to calculate the spatial transformation relation of the images using the gray information under certain similarity measure criteria. This method does not need to extract image features, but use the gray information of the image directly. The advantages of it are no need for image preprocessing and good versatility. And the disadvantages include the large amount of calculation and sensitivity to the change of image gray. Besides, it only reflects the statistical information of gray value about each pixel in the image while ignoring the image spatial structure information, so it is easy to cause false matches [6]. The registration method based on the image features is implemented by extracting the invariant features (such as corners, lines, contours, etc.) to obtain the transformation relation. However, because the IR and VI are in different bands, there is little relevance between them, making registration method based on feature points easily lead to a large amount of false matches. Since

^a Corresponding author: zxlcheer@163.com

registration method based on line features puts forward more stringent requirements on image content and deformation, it limits the application scope of the algorithm [7]. And the validity of the registration method based on contour features is largely dependent on the quality of extracted contour features, while stable and effective contour extraction is a classical problem in the field of computer vision [8].

In this paper, we proposed a registration algorithm which combines pyramid hierarchical idea and graph theory. First, extracting the maximally stable extremal regions (MSER) in the determining maximum down-sampling images and constructing the graph feature for each region. Then the regional mapping relationship between the IR and VI is determined by the graph matching method. Finally, we can construct the initial set of matching points based on the regional mapping and obtain the optimum matching parameters using the random sample consensus (RANSAC) algorithm and determine the error evaluation parameters. According to the idea of pyramid stratification, the exact matching in local range in the high resolution images is carried out under the constraint condition of current matching error. Good performance can be achieved under the premise of ensuring the robustness of matching.

2 Registration definition and transformation model

Image registration is to make two images $f_1(x,y)$ and $f_2(x,y)$ achieve the consistency on geometric position and gray level. The relationship between them can be expressed as follows:

$$f_2(x,y) = g(f_1(h(x,y))) \quad (1)$$

where g represents a one-dimensional gray transform function, and h represents a two-dimensional spatial position transform function. Image registration is to find the ideal transformation parameters. Before the registration, it is necessary to determine the transformation model. The commonly used models are rigid body transformation, similarity transformation, affine transformation, perspective transformation, etc. In this paper, we consider affine transformation, that is, there are not only existing translation, rotation and scaling between images, but also inversion and shear [9]. The form of the transformation matrix \mathbf{H} is as follows:

$$\mathbf{H} = \begin{pmatrix} m_0 & m_1 & m_2 \\ m_3 & m_4 & m_5 \\ 0 & 0 & 1 \end{pmatrix} \quad (2)$$

where $m_0, m_1, m_2, m_3, m_4, m_5$ represent six independent degrees of freedom, an affine transformation can be expressed as follows:

$$\begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = \begin{bmatrix} m_0 & m_1 \\ m_3 & m_4 \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} + \begin{bmatrix} m_2 \\ m_5 \end{bmatrix} \quad (3)$$

3 Algorithm

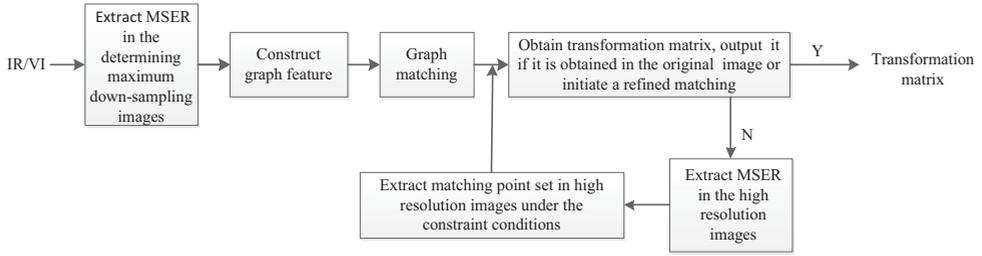


Figure 1. Algorithm block diagram of image registration.

3.1 Extraction of MSER

It is difficult to extract the common features for IR and VI due to the different imaging mechanisms. For the IR and VI, the same target may have different contrast and gray level; on the other hand, the same contrast and gray level may not belong to the same target. The purpose of extracting MSER is to solve the above difficulties, as many as possible to extract the common targets in IR and VI. And MSER is a kind of regional feature extraction operator which has characteristic of affine invariant. Firstly, the original images need to be converted into the gray images, and then the connected pixel sequences are obtained according to the pixel gray value. Finally, detecting the stability of connected regions according to the maximum rate of change V_{max} and the region area to obtain the MSERs [10-11]. For region Q and its border ∂Q , if gray value $I(p) < I(q)$ for $\forall p \in Q$ and $\forall q \in \partial Q$, then Q is called a minimum region; conversely, if $I(p) > I(q)$, then Q is called a maximum region. $Q_1, Q_2, \dots, Q_{i-1}, Q_i, \dots$ represent extremal regions that are nested with each other, that is $Q_{i-1} \subset Q_i$, the change rate on area Q_i is shown as follows:

$$q(i) = \frac{|Q_{i+\Delta} - Q_{i-\Delta}|}{Q_i} \tag{4}$$

if $q(i) \leq V_{max}$, then Q_i is called the most stable extremal region, Δ represents the change of gray value, $|\cdot|$ represents the region area, that is the number of pixels covered by the region. The gray values of the most stable extremal regions are less than (MSER-) or larger than (MSER+) the gray values of the pixels on the boundary. As shown in Figure 2, (a) is part of MSERs marked in VI, and (b) is the MSER extraction results of marked regions in (a). Because the connectivity is used in the extraction process and the constraint of region size, so the second word only retains the lower part. The goodness of MSER comes from its extraction process which is in accordance with human visual system, that is, the human eyes are more sensitive to the saliency regions (corresponding to the MSERs).

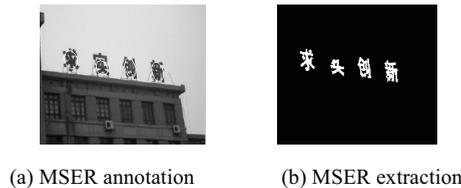


Figure 2. The extraction of MSER.

3.2 Construction of graph features

In order to describe the structure characteristics of the extracted MSERs, we make a polygon approximation for each MSER, and then construct the attribute graph [12-15]. For the matching of two polygons, let us consider two attribute graphs G_1 and G_2 , where $G_1 = (V_1, \varepsilon_1, A_1^V, A_1^E)$ with n_1 nodes and $G_2 = (V_2, \varepsilon_2, A_2^V, A_2^E)$ with n_2 nodes. V, ε, A^V, A^E represent a set of nodes, edges, node attributes and edge attributes, respectively. We only consider the edge attributes and they are expressed by 1×19 vectors. The first component represents the normalized distance, and the remaining 18 components represent the angle information of the edge and its adjacent edges. For the edge e_{ij} of point v_i to v_j , its first component represents the normalized length L_{ij} , and we use cyclic Gauss histogram for angle information, as shown in Figure 3.

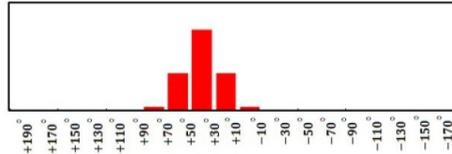


Figure 3. The representation of angle information.

We use uniform bins of size $\frac{2\pi}{n_p}$, where $n_p = 18$, that is, each interval represents 20 degrees. For angle θ_{ij} , the angle information \mathbf{P}_{ij} is shown as follows:

$$\mathbf{P}_{ij}(k) = f_p(k - m),$$

$$\text{s.t. } f_p(x) = N(0, \sigma_p) + N(\pm n_p, \sigma_p), \theta_{ij} \in \text{bin}_\theta(m) \quad (5)$$

where $N(\mu, \sigma)$ represents a discrete Gaussian window of size σ centered on μ , additional Gaussian terms in $f_p(x)$ induce the circular bins for angle. The final edge attribute \mathbf{a}_{ij} is composed by concatenating the normalized distance L_{ij} and angle information \mathbf{P}_{ij} . For edge e_{ij} : $\mathbf{a}_{ij} = [L_{ij}; \mathbf{P}_{ij}]$, which is asymmetric ($\mathbf{a}_{ij} \neq \mathbf{a}_{ji}$). In this work, we used a window size $\sigma_p = 5$, so that $N(0, 5) = 1.0$, $N(\pm 1, 5) = 0.4578$, $N(\pm 2, 5) = 0.0439$, and 0 otherwise.

3.3 Graph matching

3.3.1 Graph matching model

In this paper, one-to-one matching constraints are adopted so that every node in G_1 is mapped to at most one node in G_2 and vice versa. Graph matching identifies the subset of node correspondences between G_1 and G_2 among all of the possible correspondences, which best preserve the attribute relations under the matching constraints. We use binary indicator vector $\mathbf{y} = \{0, 1\}^{n_1 n_2}$ which has $n_1 \times n_2$ components to represent the correspondence between G_1 and G_2 , where n_1 and n_2 are the numbers of nodes in G_1 and G_2 respectively. If $v_i \in V_1$ matches $v_a \in V_2$, so that $y_{i,a} = 1$, and $y_{i,a} = 0$ otherwise. Affinity matrix M consists of the affinity values between edges and nodes. The

compatibility of two edge attributes $\mathbf{a}_{ij} \in A_1^e$ and $\mathbf{a}_{ab} \in A_2^e$ is encoded in the non-diagonal component $M_{ia;jb}$, and the compatibility of two node attributes $a_i \in A_1^V$ and $a_a \in A_2^V$ is encoded in the diagonal component $M_{ia;ia} \cdot M_{ia;jb}$ can be expressed as follows:

$$M_{ia;jb} = \exp(-\|\mathbf{a}_{ij} - \mathbf{a}_{ab}\| / \sigma_s^2) \quad (6)$$

where scaling factor σ_s^2 is chosen to be 0.1. In this paper, the diagonal components of M are set to zero, because no node attributes are used. The graph matching problem is to find the best matching relation \mathbf{y}^* between two graphs.

3.3.2 Construction of objective function

We can find the corresponding affinity values of edge attributes in M according to the components whose values are equal to 1 in match result \mathbf{y} . Then the affinity values are calculated as matching score, and the matching score is used to evaluate the matching result of the two graphs. Graph matching can be expressed as follows:

$$\begin{aligned} \mathbf{y}^* &= \arg \max_{\mathbf{y}} \mathbf{y}^T \cdot M \cdot \mathbf{y} \\ \text{s.t. } \mathbf{y} &\in C_{int} \cap C_{one} \end{aligned} \quad (7)$$

where $C_{int} = \{\mathbf{y} \mid \mathbf{y} \in \{0, 1\}^{n_1 \times n_2}\}$ enforces the integer constraint for the $n_1 \times n_2$ components in \mathbf{y} , namely the component in \mathbf{y} only can be 0 or 1. And $C_{one} = \{\mathbf{y} \mid \forall i, \sum_a y_{ia} \leq 1\} \cap \{\mathbf{y} \mid \forall a, \sum_i y_{ia} \leq 1\}$ enforces the one-to-one constraint. The affinity measure is typically restricted to be nonnegative, therefore the number of matches in the final solution is implicitly determined to the maximum possible number of matches under one-to-one constraint. The underlying assumption is that the solution that contains true matches would have a higher objective value than those which do not. However, this assumption poorly holds even with few outliers. The matching score S is composed of two parts, $S = S_{in} + S_{out} = \mathbf{y}^T \cdot M \cdot \mathbf{y}$, where S_{in} represents the matching score of the correct matching point pairs, and S_{out} represents the matching score of the outliers. The existence of the outliers will increase the matching score, so in this paper, we consider not only the matching score but also the number of matching points. By changing the objective function of formula (7) to the following:

$$\begin{aligned} \mathbf{y}^* &= \arg \max_{\mathbf{y}} \mathbf{y}^T \cdot M \cdot \mathbf{y} - \lambda^T \Omega(\|\mathbf{y}\|) \\ \text{s.t. } \mathbf{y} &\in C_{int} \cap C_{one} \end{aligned} \quad (8)$$

where $\Omega(\|\mathbf{y}\|) = [\|\mathbf{y}\|, \|\mathbf{y}\|^2]$, we approximate λ as $\lambda = [0, 0.7]$ in all the experiments. We adopt the Markov chain Monte Carlo (MCMC) sampling technique proposed in literature [16] to solve the optimization problem in formula (8).

3.4 Solve transformation matrix

We adopt the coarse-to-fine strategy, so the transform matrix needs to be obtained at each level of resolution, and the transformation matrix is obtained at low resolution plays a guiding role on the high resolution matching; finally using the matrix obtained in the original images to achieve registration. In this paper, the original images are processed by three times of down-sampling and the initial matching

point set is obtained on the maximum down-sampling images, and then the matching point set is used as the input of the RANSAC algorithm. According to the idea of Pyramid stratification, the above process is repeated in high resolution images to realize the precise matching in local area until the original resolution images according to the transformation matrix obtained in low resolution and with the constraint of current matching error.

3.5 Evaluation criteria

In this paper, we adopt the root mean square error (RMSE) as the basis for testing the accuracy of registration algorithms. The smaller the RMSE is, the higher is the accuracy of registration. The calculation formula for RMSE is defined as:

$$\text{RMSE} = \sqrt{\frac{1}{m} \sum_{i=1}^m (x_i - x'_i)^2 + (y_i - y'_i)^2} \quad (9)$$

where m is the number of matching points, x_i, y_i are the coordinates of reference image, x'_i, y'_i are the transformed coordinates in the image to be registered. At the same time, we also consider the subjective evaluation, if the registration accuracy is high, the common visual sensitive structure information in IR and VI can completely align, so it does not appear significant dislocation in the fused image and we can obtain the good visual effect.

4 Experiments

In this section we evaluate our method using real images of three scenes, and the scene 1 contains a building background, as shown in Figure 4 (a) and (b). There is a tiny deformation between images in scene 2, as shown in Figure 5 (a) and (b). And the image quality is poor in scene 3, as shown in Figure 6 (a) and (b). For each scene, we compare the result with two matching methods, namely, the registration method based on mutual information and the registration method based on line features. In the experiment, the VI is the image to be registered and the IR is the reference image. For scene 3, the registration method based on mutual information cannot be implemented due to the difference between VI and IR. Therefore, we only give the experimental results based on line features and the method proposed in this paper. The experimental results are shown as follows:

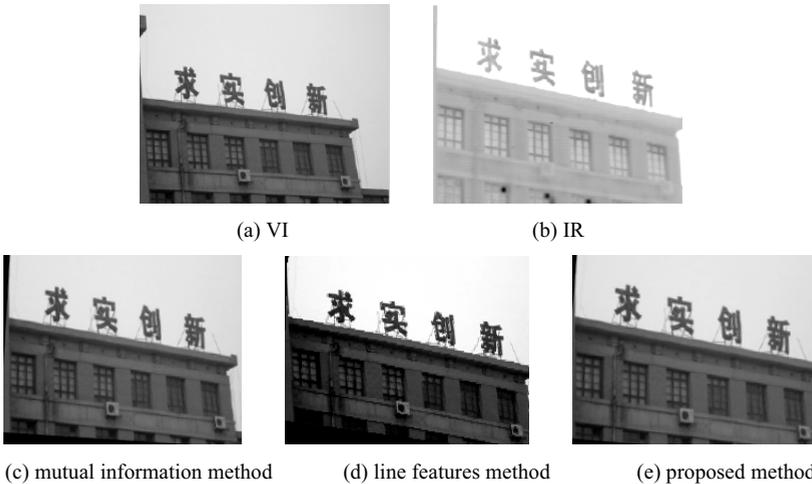


Figure 4. Comparison of results for the first scene.

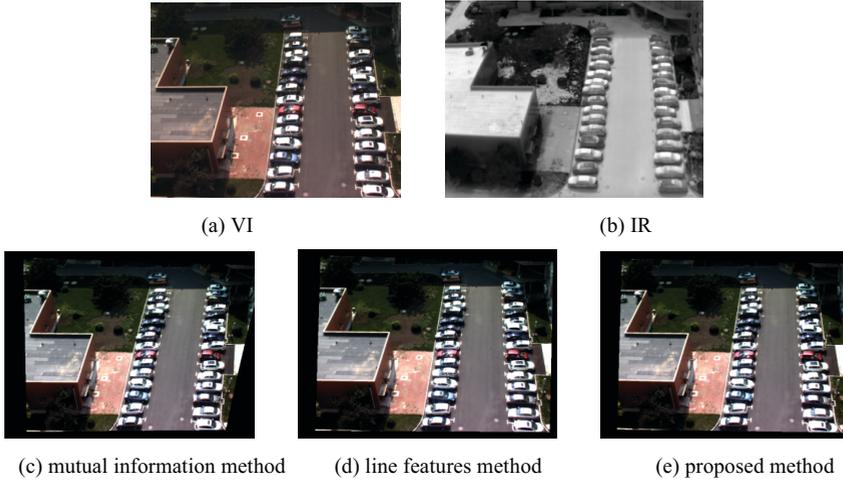


Figure 5. Comparison of results for the second scene.

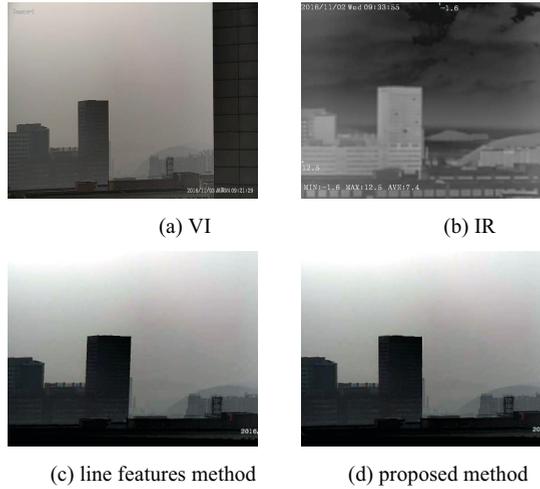


Figure 6. Comparison of results for the third scene.

4.1 Objective evaluation

In order to quantitatively evaluate the effect of registration, we calculate the RMSE of the registration method based on mutual information, the registration method based on line features and the proposed method respectively. The results are shown in Table 1:

Table 1. The RMSE of three registration methods.

Methods	RMSE (scene 1)	RMSE (scene 2)	RMSE (scene 3)
Mutual information	8.4524	11.462	---
Line features	4.769	2.769	4.683
Proposed	1.8797	1.536	2.320

It can be seen from Table 1 that our method achieves the best performance on three scenes compared with the registration methods based on mutual information and line features.

4.2 Subjective evaluation

In order to facilitate the observation of the effect of registration, we fuse the visible reference image and infrared image which has been registered, and the fusion results as shown in the following figures:

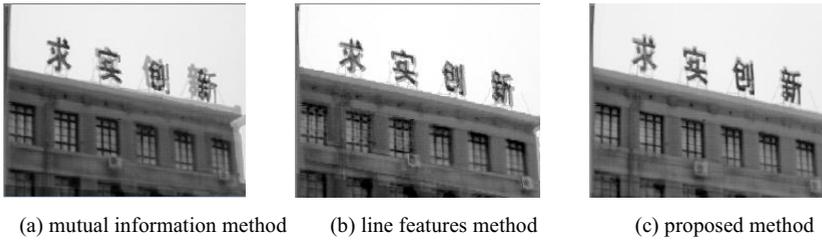


Figure 7. Comparison of fusion results for the first scene.

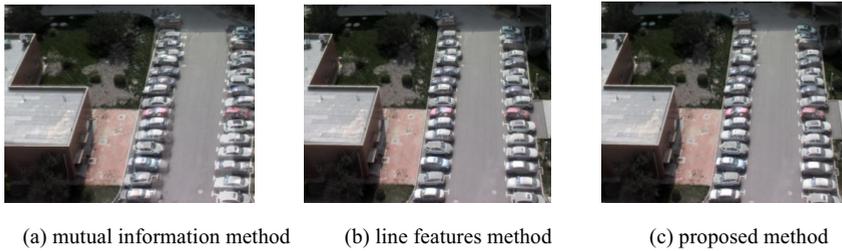


Figure 8. Comparison of fusion results for the second scene.

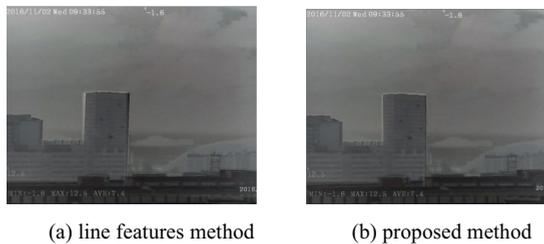


Figure 9. Comparison of fusion results for the third scene.

From the fusion results, we can see that the registration result based on mutual information method exhibits serious dislocation phenomenon for scene 1 and scene 2, and the mutual information method is not able to achieve the registration for scene 3 due to too poor quality, while there is a slight dislocation for registration method based on line features. After applying the registration algorithm proposed in this paper, the common visual sensitive structure information completed the alignment, and compared with the other two methods, there is no significant dislocation phenomenon and visual effect is good enough.

Conclusion

The gray characteristics are different for infrared image and visible images in the same scene due to different imaging mechanisms. Thus it causes the difference of extracted corners, lines and contours, especially for images of poor quality. So some of the existing registration algorithms are not ideal. In order to solve this problem, we proposed a registration algorithm based on the pyramid hierarchical idea and graph theory, and we consider the structure information of the graphs in the coarse-to-fine matching process. Experimental results show that the proposed method has a good adaptability to the image content and image deformation. Compared with other methods, the proposed algorithm can achieve state-of-the-art performance, merely with a small matching error under the premise of ensuring the robustness of the matching.

References

1. S.E. A Raza, Registration of thermal and visible light images of diseased plants using silhouette extraction in the wavelet domain, *Pattern Recognition*, **48** 7 2119-2128 (2015).
2. T. Chakravorty, G.A. Bilodeau, E. Granger, Automatic Image Registration in Infrared-Visible Videos using Polygon Vertices, *print arxiv* (2014).
3. H.R. Su, S.H. Lai, Non-rigid registration of images with geometric and photometric deformation by using local affine Fourier-moment matching, *CVPR*, 861-866 (2015).
4. Y.Y. Peng, W.J. He, G.H. Gu, Infrared and visible images registration based on feature points, *Laser and infrared*, **42** 6 713-717 (2012).
5. L. Lian, G.H. Li, J. Zhang, Automatic registration algorithm of infrared and visible images based on parameter estimation, *Electronic journal*, **40** 9 1829-1838 (2012).
6. J.L. Zhang, X. Jin, L. Hao, Infrared and visible images registration based on alignment and mutual information, *Modern electronic technology*, **12** 5-8 (2013).
7. C.L. Lu, L.H. Ma, M. Yu, Visible and infrared images registration algorithm based on parallel line group, *Journal of PLA University of Science and Technology Natural Science Edition*, **5** 439-446 (2015).
8. L. Lian, G.H. Li, J. Zhang, Automatic registration algorithm of infrared and visible images based on edge optimal mapping, *Acta Automatica Sinica*, **38** 4 570-581 (2012).
9. J. Zhang, J.S. Li, Y.W. Yang, Image distortion correction based on affine transformation and multilayer B spline registration, *Optics and precision engineering*, **23** 3 846-854 (2015).
10. M.M. Manafzade, A. Harati, Point cloud registration using MSSIR: Maximally stable shape index regions, *Electrical Engineering*, 1-6 (2013).
11. L. Liu, H. Tuo, T. Xu, Multi-sensor image registration using edge-enhanced maximally stable extremal region, (2012).
12. Y. Suh, K. Adamczewski, K.M. Lee, Subgraph matching using compactness prior for robust feature correspondence, *CVPR*, 5070-5078 (2015).
13. M. Cho, J. Lee, K.M. Lee, Reweighted Random Walks for Graph Matching, *ECCV*, **9** 492-505 (2010).
14. M. Cho, K. Alahari, J. Ponce, Learning Graphs to Match, *ICCV*, 25 – 32 (2013).
15. F. Zhou, F.D. L Torre, Deformable Graph Matching, **9** 4 2922-2929 (2013).
16. W.R. Gilks, Markov chain monte carlo, *Wiley Online Library*, (2005).