Survey on Registration Techniques of Visible and Infrared Images

Kyungsu Kim*
Agency for Defense Development, Yuseong P.O.Box 35, Daejeon, 305-600, Korea
kimkyungsu69@gmail.com

Abstract

This paper aims to present an overview of recent as well as classic multi-modal image registration methods, specially about registration of visible and infrared image. Image registration is the process of overlaying images (two or more) of the same scene taken at different times, from different viewpoints, and/or by different sensors. The registration geometrically aligns two images (the reference image and sensed input images). The reviewed approaches are classified according to their nature (pixel-based and feature-based). Registration algorithms compute transformations to set correspondence between the two images. The purpose of this paper is to provide a comprehensive review of the existing literatures available on registration algorithms of visible and infrared images. Moreover, we compare and analyze using real scene data the performance of Mutual information based algorithm and Fourier transform based algorithm which are representative algorithms in registration algorithms.

Keywords: Image Regestration Algorithm, Mutual information based algorithm, Fourier transform based algorithm

1 Introduction

Recent advances in imaging, networking, data processing and storage technology have resulted in an explosion in the use of multi-modality images in a variety of fields, including video surveillance, urban monitoring, cultural heritage area protection and many others. The integration of images from multiple channels can provide complementary information and therefore increase the accuracy of the overall decision making process. A fundamental problem in multi-modality image integration is that of aligning images of the same/similar scene taken by different modalities. This problem is known as image registration and the objective is to recover the correspondences between the images. Once such correspondences have been found, all images can be transformed into the same reference, enabling to augment the information in one image with information from the others.

The process of registering two or more images can be computationally intensive. The general procedure, however, is relatively straight-forward. The algorithm can be stated as follows:

1) Given two images to registered, one should denote as the reference image and the other should denote the mis-registered floating image. The objective is to transform the floating image until it looks like the reference image.

2) Choose a criterion function that will determine the degree of match/mismatch between the reference image and the transformed floating image. Also choose a stopping criterion that indicates the image is registered.

3) Optimize the transformation on the test image such that the stopping criterion is met.

IT CoNvergence PRActice (INPRA), volume: 3, number: 2 (June), pp. 25-35
*Corresponding author: Agency for Defense Development, Yuseong P.O.Box 35, Daejeon, 305-600, Korea, Tel: +82-(0)42-821-4584

The choice of criterion function is an essential part of the registration process. Comparison methods such as cross-correlation and mutual information are some of the more common techniques found in the literature. Correlation techniques perform well in mono-modal registration wherein there is a linear relationship between the measurements for the same spatial elements in the two image acquisitions. However, because of the non-linear relationship that can arise between the intensities of images across different modalities, correlation has been shown generally not to be a suitable candidate for a criterion function in multi-modal image registration. Since its introduction by Viola and Wells[15], mutual information has been one of the most discussed and acclaimed registration measures for multi-modal image registration. Mutual information is a statistical measure that assesses the strength of dependence between two stochastic variables. Even though mutual information has been shown to outperform other comparison methods used for registration, it is not a panacea[12]. The process can often be improved by incorporating spatial information when performing the alignment. In this paper, we limit our review to the work for registering visible and infrared images.

(a) Infrared image (b) Visible image

Figure 1: Two different modal images with same scene

2 Related Works

2.1 Classification of image registration algorithm

Several related survey papers for image registration have appeared over the years. [13], [3], [5] have provided a broad a broad overview of over three hundred papers for registering different types of sensors. Following most of literature, we also divide existing techniques into two categories: Pixel-based method and feature-based method. Pixel-based method first define a metric, such as the sum of square differences and mutual information[3], which measures the distance of two pixels from different images. The registration problem is them changed to minimize the total distance between all pixels on one image and the corresponding pixels on another image. The representatives are pixel-based methods are Mutual information based algorithm and Fourier transform based algorithm. Mutual information (MI) is a statistical measure that finds its roots in information theory. MI is a measure of how much information one random variable contains about another. The MI of two random variables A and B can be defined as

$$I(A,B) = \sum_{a,b} P_{A,B}(a,b) \log \frac{P_{A,B}(a,b)}{P_A(a) \bullet P_B(b)} \tag{1}$$

where $P_{A,B}(a,b)$ is the joint probability mass function (pmf) of the random variables A and B, and $P_A(a)$ and $P_B(b)$ are the marginal probability mass functions of A and B, respectively. In working with images, the functional form of the pmf is not readily accessible. The normalized histograms of the intensity values for each image serve as a good approximation of the pmf. The MI can also be written in terms of the marginal and joint entropy of the random variables A and B as follows

$$I(A,B) = H(A0) + H(B) - H(A,B) \tag{2}$$

where $H(A)$ and $H(B)$ are the entropies of A and B, respectively, and $H(A,B)$ is the joint entropy between the two random variables. They are defined as

$$H(A) = -\sum_{a} P_A(a) \log P_A(a) \tag{3}$$

$$H(A,B) = -\sum_{a,b} P_{A,B}(a,b) \log P_{A,B}(a,b) \tag{4}$$

One interpretation of entropy is as a measure of uncertainty of a random variable. A distribution with only a few large probabilities has low entropy value; the maximum value over a finite interval is achieved by a uniform distribution over that interval. The entropy of an image indicates how difficult it is to predict the gray value of an arbitrary point in the image. MI is bounded by cases of either complete dependence or complete independence of A and B, yielding values of I = H and I = 0, respectively, where H is the entropy of A and B. The strength of the mutual information similarity measures lies in the fact that no assumptions are made regarding the nature of the relationship between the image intensities in both modalities, except that such a relationship exists. This is not the case for correlation methods, which depend on a linear relationship between image intensities. For image registration, the assumption is that maximization of the MI is equivalent to correctly registering the images. It is clear that if the joint entropy of A and B are not affected by the transformation parameters, maximizing the MI is equivalent to minimizing the joint entropy. The joint entropy is minimized when the joint pmf of A and B contain few sharp peaks. This occurs when the images are correctly aligned. When the images are mis-registered, however, new combination of intensity values from A and B will cause dispersion in the distribution. This dispersion leads to a higher joint entropy value, which in turn decrease the MI.

Fourier transform based algorithm is preferred, if an acceleration of the computational speed is needed or if the images were acquired under varying conditions or they are corrupted by frequency-dependent noise. They exploit the Fourier representation of the images in the frequency domain. The phase correlation is based on the Fourier Shift Theorem and was originally proposed for the registration of translated images. It computes the cross-power spectrum of the sensed and reference images and looks for the location of the peak in its inverse. If a change of image scale is present too, the images can be registered using the combination of polar-log mapping of the spectral magnitude (which corresponds to the Fourier-Mellin transform) and the phase correlation or cepstrum filter. Some authors propose to compute the correlation in frequency domain to handle multimodal images when applied to the edge representations instead of the original gray level images[14]. Extension of phase correlation to sub-pixel registration by means of the analytic expression of phase correlation on down sampled images was introduced by Foroosh et al[6].

In feature-based method, interest points like Harris corners, scale invariant feature transform (SIFT), speed-up robust feature (SURF)[2], etc., are first extracted from images. Afterwards, these features are matched based on the metrics, such as cross correlation and mutual information. Once more than four feature correspondences are obtained, the transform can be computed. Over the last decade several variations of the original SIFT algorithm, as well as some novel approaches, have been proposed focusing on the perceived weakness of SIFT[2], [9]. Recently, applications that combine feature points from

different spectral band images are being developed. The difficulty in finding correspondences between feature points from visible and infrared images results from the nonlinear relationship results in a lack of correlation between their respective gradient. Furthermore, infrared images appear smoother, with loss of detail and texture, so that the detection of corners, as candidates for local descriptor points, is also poorly favored.

In principle, pixel-based method should be better than the feature-based method because the former considers the global minimization of the cost function, but the later one minimizes the cost function locally. In practice, however, feature-based method has better performance for many applications, because the interest point is supposed to be distinctive in a local area, thus leading to the better matching. On the other hand, the pixel-based method is much more expensive than the feature-based algorithm, because every pixel is involved in the computation. Most existing publications for image registration dedicate to solving four problems: 1) an efficient way to extract feature points, which guarantees the majority of features on both images is identical; 2) a better feature descriptor; 3) a suitable metric to measure the distance of two feature descriptors; 4) a proper transform model.

Among these four problems the detection of repeatable features and also the feature matching are more challenging when dealing with infrared and visual cameras. The main reason is that the electromagnetic wavelengths of visible sensor and IR sensor are quite different. Normally, the wavelength of IR sensor is from 4 to 12 microns, while the wavelength of visible sensor roughly lies between 0.4 to 0.7 microns. This leads to the fact that IR images have noticeably less texture in the area where temperatures are more homogeneous. However, the texture information is very important for both interest point detection and feature matching.

[1] presents a novel feature point descriptor for the multispectral image case : Far-infrared and Visible spectrum images. Initially, points of interest are detected on both images through a SIFT-like based scale space representation. Then these points are characterized using an Edge Oriented Histogram (EOH) descriptor. Finally, points of interest from multispectral images are matched vy finding nearest couples using the information from descriptor.

As an another approach, there is a method dealing with transforming a multi-modal registration to a mono-modal one used in medical image registration. These techniques that reduce a multi-modal to a mon-modal registration can be differentiated into two classes. The first one try to simulate one modality from the other[16]. Examples are X-ray to CT (Computer Tomography) registration with the creation of digitally reconstructed radiographs and ultrasound to CT registration with the simulation of ultrasound images. The second group consists of methods that transfer both images into a third, artificial modality.

2.2 Registration of visible and infrared image

Many approaches have been proposed for automatically registering Infrared (IR) and visible images. Edge/gradient information is one of the most popular features as their magnitudes [9] and orientation [4] may match between infrared and visible images.

In [5], authors first extract edge segment, which are then grouped to form triangles. The transform can be computed by matching triangles from the source to destination images.

Huang et al. [17] proposes a contour-based registration algorithm, which integrates the invariant moments with the orientation function of the contours to establish the correspondences of the contours in the two images. Normally it is difficult to obtain accurate registration by using contour-based method, because precisely matching all contours detected from two images is challenging. Moreover, this method drastically increases computation time compared to interest point-based registration. To improve this work, Han et al.[7] propose to find correspondences on moving contours. They extract silhouettes of moving humans from both images. Matching only the contours of humans significantly improves both the performance and the efficiency of the algorithm. An alternative[7] is to make use of the object moving

paths generated by object tracking algorithm. Finding correspondences between trajectories helps to align images. This type of algorithm work very well when moving objects can be precisely tracked from both channels. Unfortunately, the current tracking algorithm is not satisfactory in many applications. In [8], they try to align based on lines derived from edges of the images. These lines strongly relate to the boundaries of objects, which always appear on both images through IR sensor and visible sensor have significantly different properties. Moreover, they enable one-to-many matching based on a simple feature descriptor, which allows one feature on one image to have more potential correspondences in another image. This ensures that the majority of the initial matching is correct.

3 Experimental comparison

In this section, we implement two kinds of algorithms, mutual information based algorithm and Fourier transform based algorithm, to register visible image and IR image and compare their performance. We use Normalized mutual information[10] and SSD cost function based on linear regression[11] as similarity metrics. In these experiments, we consider only translational shift.

In order to register with the reference image, we apply original image without any special pre-processing for mutual information based algorithm. However, we get gradient images from image pairs, apply smoothing filter and modify a bit the algorithm[11] to consider translational shift for Fourier based algorithm. We acquire test images from one site including objects using MWIR camera (466X364) at regular intervals for all the day. Image sequences consist of 42 images which include most of normal images, some saturated images due to sunburn and some with occluded parts caused by regional fog. The reference image is CCD image taken in the afternoon on fine day at same place. Moreover, to evaluate the effect of contrast enhancing algorithm for registration, we apply pre-processing algorithm, CLAHE(Contrast Limited Adaptive Histogram Equalization), to input image in our experiments.

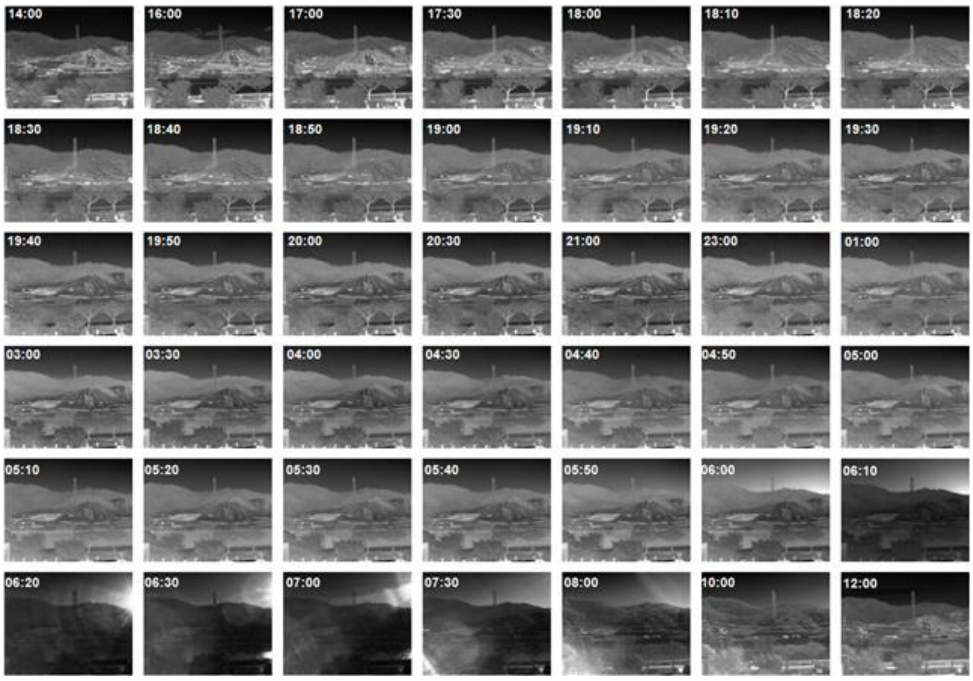


Figure 2: MWIR images

Figure 2 shows 42 images(466X364) collected for 24 hours.

```
1  input:reference image(g), input image(w), ROI image(f);
2  assign h1=1, h2=f, and h3=(∂f/∂θ);
3  precompute $\overline{F(h_i h_j)}$ for i, j = 1,2,3;
4  precompute $\overline{F(h_i)}$ for i = 1,2,3;
5  repeat
6      find bounding box for nonzero part of w';
7      shift g' and w' equally so that the bounding box is in the upper-left corner;
8      pad/crop the shifted g' and w' so that they are the same size as f;
9      compute F{w'} and F{g'w'};
10     compute $[A_{x,y}]_{i,j} = F^{-1}\{\overline{F\{h_i h_j\}} F\{w'\}_{x,y}\}$,;
11     i,j=1,2,3 for all integers shifts (x,y);
12     simultaneously (Note that $A_{x,y}$ is symmetric.);
13     compute $[b_{x,y}]_{i,j} = F^{-1}\{\overline{F\{h_i\}} F\{g'w'\}_{x,y}\}$,;
14     i,j=1,2,3 for all (x,y);
15     for each valid shift(x,y) do
16         solve $A_{x,y}s = b_{x,y}$ for $s$;
17         $C(x,y) = s^T A_{x,y}^T A_{x,y} s - 2b_{x,y}^T s$ ;
18         if C(x,y) is the lowest so far then
19             record $(x^*, y^*) = (x,y)$ and $s^* = s$;
20         end
21     end
22     adjust to compensate for the initial ;
23     shift of g' and w' (see step 7);
24 until completed more than 3 iterations;
25 The solution is $(x^*, y^*)$
```

Algorithm 1: SSD with linear remapping

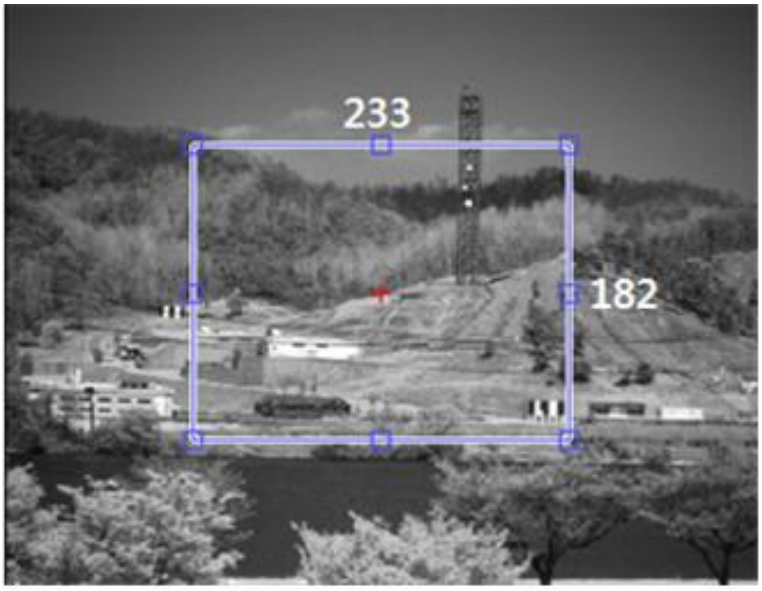


Figure 3: Reference image(CCD, 466 X 364)

Figure 3 is CCD image used for the reference image and the blue rectangle in the middle shows the region used for matching.

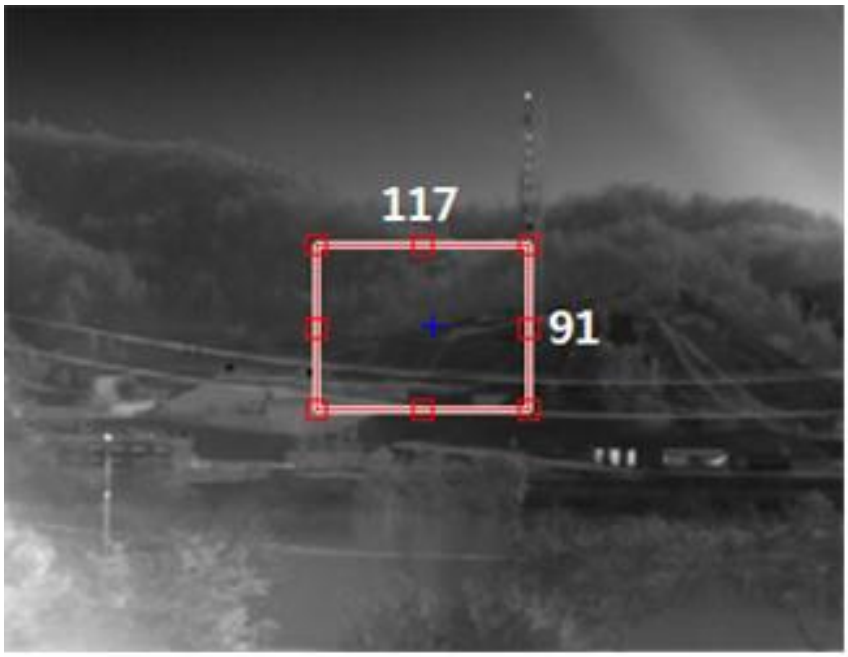
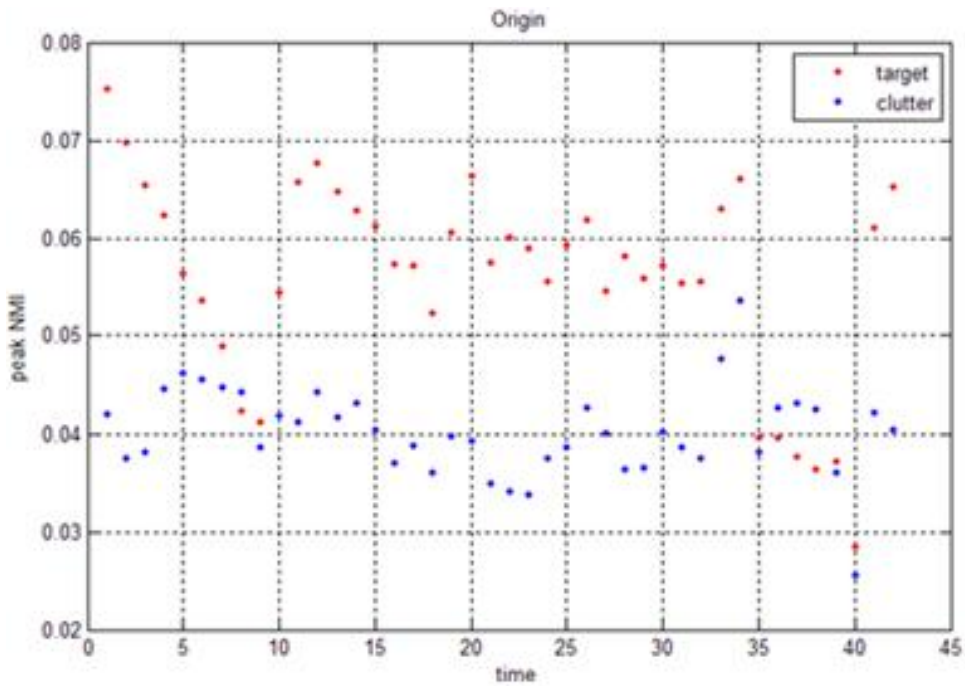


Figure 4: Input image (Infrared)
(Search region : 117 X 91)

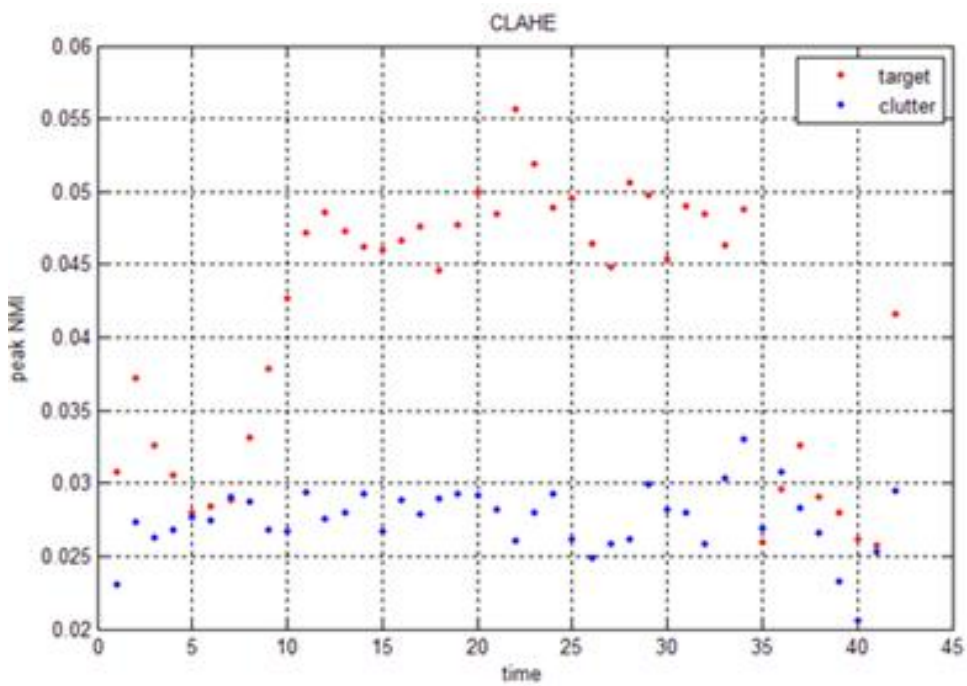
Figure 4 is the input image and the red rectangle in the middle shows the search region for matching.

Figure 5: Pre-processed input image by CLAHE

Figure 5 shows the pre-processed input image by CLAHE.



(a) Result of original image



(b) Result of contrast enhanced image(CLAHE)

Figure 6: Peak value of target and clutter after applying mutual information based algorithm (x axis : ♯ of input image, y axis : peak value of target and clutter)

Figure 6 is the result of the mutual information based algorithm and show the NMI peak values of target and clutter in 42 images. The NMI peak value of target is included in 10X10 regions around the center of input image. The NMI peak value of clutter is the maximum of the rest except for the maximum

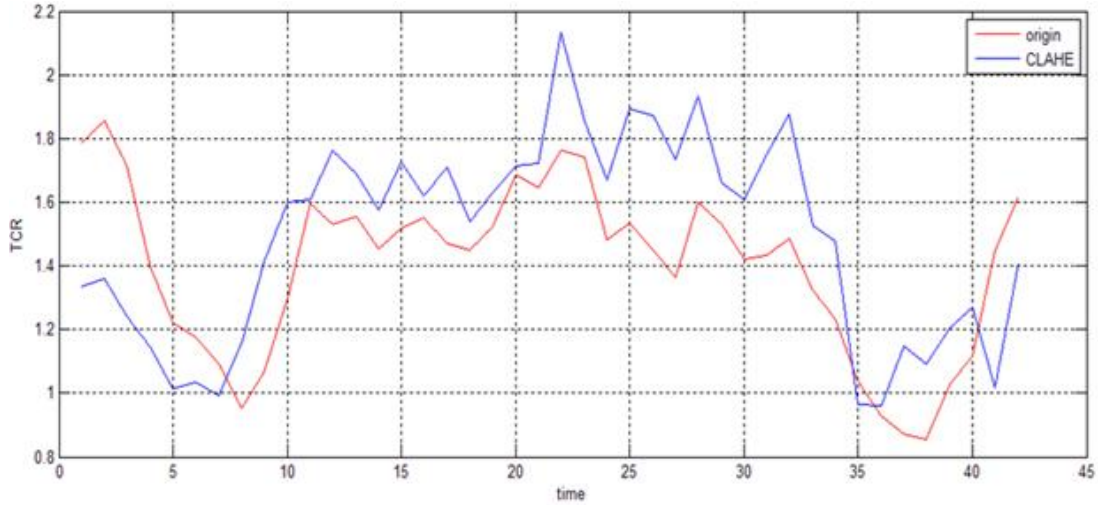


Figure 7: Comparison of TCR between original image and contrast enhanced image

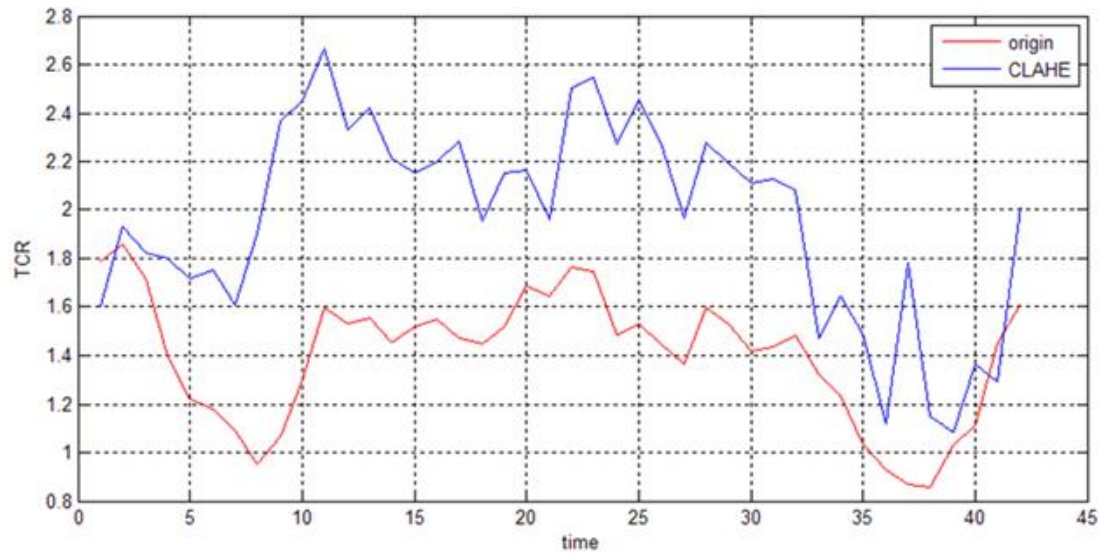


Figure 8: Comparison of TCR between original image and contrast enhanced image (in case that reference image is also contrast enhanced)

value considered as the target.

Figure 7 shows TCR about original image and contrast enhanced image. TCR is calculated by the following equation.

$$TCR(Target\ to\ Clutter\ Ratio) = Target\ peak\ NMI / Clutter\ peak\ NMI \tag{5}$$

We can see that the TCR of contrast enhanced image is higher than that of original image. Figure 8 is the result in case that the reference image (visible image) is also contrast enhanced. The TCR of contrast enhanced image has been improved quite differently from the Figure 7.

Table 1 and Table 2 show the registration result of Mutual information based algorithm and Fourier transform based algorithm. We consider that if the distance error of registration is inside of 5 pixels and the correlation surface becomes good curved surface, registration is success. As the results turned out,

we can see that the Mutual information based algorithm haven't always better performance than Fourier transform based algorithm and the latter has advantage in execution time and can have better performance than the former if the good pre-processing is applied to input image.

Table 1: Results of Mutual information based algorithm

Algorithm / Registration error (pixel distance)	Mutual Information based algorithm	
	Original image	Contrast enhanced image
inside of 5 pixels	37/42(88%)	39/42(93%)

Table 2: Results of Fourier transform based algorithm

Algorithm / Registration error (pixel distance)	Fourier transform based algorithm	
	Original image	Contrast enhanced image
inside of 5 pixels	37/42(88%)	41/42(100%)

4 Conclusion and Future Work

Image registration is one of the most important task when integrating and analyzing information from various sources. Among those, registration of visible image and infrared image is specially very important and has been studied for a long time in military application. This paper gives a survey of the classical and up-to-date registration algorithm of visible and infrared image, classifying them according to their nature. Specially, we compare the performance of representative algorithms, Mutual information based algorithm and Fourier transform based algorithm which are widely used in multimodal registration field. Mutual information based algorithm has good performance but disadvantage in execution speed. Fourier transform based algorithm has relatively lower performance but advantage in execution speed. As the experimental results say, Fourier transform based algorithm has better performance than Mutual information based algorithm if a proper pre-processing to input image is done, in addition to advantage in speed. Recently, not to apply a simple contrast enhancement algorithm to input image, the study on transformation to the other dimensional image to represent well common regions of multi-modal images has been done to improve the performance of registration.

Moreover, the future development in this field could pay more attention to the feature-based methods, where appropriate invariant and modality insensitive features can provide good platform for the registration. The demand for higher robustness and accuracy of the registration usually enforces solutions utilizing the iterations or backtracking, which also produces increase of computational complexity of the methods.

References

[1] C. Aguilera, F. Barrera, F. Lumbreras, A. D. Sappa, and R. Toledo. Multispectral image feature points. *Sensors*, 12(9):12661–12672, September 2012.

[2] H. Bay, T. Tuytelaars, and L. V. Gool. SURF: Speeded up robust features. In *Proc. of the 9th European Conference on Computer Vision (CV'06), Graz, Austria, LNCS*, volume 3951, pages 404–417. Springer Berlin Heidelberg, May 2006.

[3] L. G. Brown. A survey of image registration techniques. *ACM Computing Surveys (CSUR)*, 24(4):325–376, December 1992.

[4] Y. Caspi, D. Simakov, and M. Irani. Feature-based sequence to sequence matching. *International Journal of Computer Vision*, 68(1):53–64, June 2006.

[5] L. M. Fonseca and B. Manjunath. Registration techniques for multisensor remotely sensed imagery. *Photogrammetric Engineering & Remote Sensing*, 62(9):1049–1056, September 1996.

[6] H. Foroosh, J. B. Zerubia, S. Member, and M. Berthod. Extension of phase correlation to subpixel registration. *IEEE Tranasctions on Image Processing*, 11(3):188–200, March 2002.

[7] J. Han and B. Bhanu. Fusion of color and infrared video for moving human detection. *Journal of Pattern Recognition*, 40(6):1771–1784, June 2007.

[8] J. Han, D. Farin, and P. H. N. de With. Broadcast court-net sports video analysis using fast 3-d camera modeling. *IEEE Transactions on Circutts and Systems for Video Technology*, 18(11):1628–1638, November 2008.

[9] S. Lazebnik, C. Schmid, and J. Ponce. Semi-local affine parts for object recognition. In *Proc. of the 2004 British Machine Vision Conference (BMVC'04), Kingston University, London, UK*, pages 779–788, September 2004.

[10] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens. Multimodality image registration by maximization of mutual information. *IEEE Ttransactions on Medical Imaging*, 16(2):187–198, April 1997.

[11] J. Orchard. Efficient least squares multimodal registration with a globally exhaustive alignment search. *IEEE Transactions on Image Processing*, 16(10):2526–2534, October 2007.

[12] J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever. Image registration by maximization of combined mutual information and gradient information. *IEEE Transactions on Medical Imaging*, 19(8):809–814, August 2000.

[13] J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever. Mutual-information-based registration of medical images: a survey. *IEEE Transactions on Medical Imaging*, 22(8):986–1004, August 2003.

[14] Q. sheng Chen, M. Defrise, and F. Deconinck. Symmetric phase-only matched filtering of fourier-mellin transforms for image registration and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(12):1156–1168, December 1994.

[15] P. Viola and W. M. W. III. Alignment by maximization of mutual information. In *Proc. of the 5th IEEE International Conference on Computer Vision (ICCV'95), Cambridge, Massachusetts, USA*, pages 16–23. IEEE, June 1995.

[16] C. Wachinger and N. Navab. Structural image representation for image registration. In *Proc. of the 2010 Computer Vision and Pattern Recognition Workshops (CVPRW'10), San Francisco, California, USA*, pages 23–30. IEEE, June 2010.

[17] H. Xishan and C. Zhe. A wavelet-based multisensor image registration algorithm. In *Proc. of the 6th International Conference on Signal Processing Proceedings (ICSP'02), Beijing, China*, pages 773–776. IEEE, August 2002.

Author Biography

Kyungsu Kim received the B.S. in Computer Engineering in Chosun University in 1992. He received the M.S. in Computer Engineering in ChungAng University in 1994 and received the Ph.D. in Computer Engineering in Chosun University in 2000. He was a Post-doctoral researcher in Arizona State University and GIST(Gwangju Institute of Science and Technology) in 2000 and 2001 respectively. He is currently a principal researcher in ADD(Agency for Defense Development) from 2002. His current interests include multimodal image registration, multi-sensor target detection/-tracking, computer vision and computer graphics.