Multi-source Airborne IR and Optical Image Fusion and Its Application to Target Detection

Fenghui Yao and Ali Sekmen

Department of Computer Science, Tennessee State University

Abstract. This paper describes the fusion of the airborne optical image and IR (infrared) image generated from a moving platform, and the target detection from the fused images. The proposed algorithm first detects the object from optical image and IR image, respectively. Then it performs the object mapping to determine parameters for image fusion. And then it fuses the optical image and IR image and detects the target from the fused images. The real-world videos generated from a helicopter are used to test this algorithm. The experiment results validate the proposed algorithm.

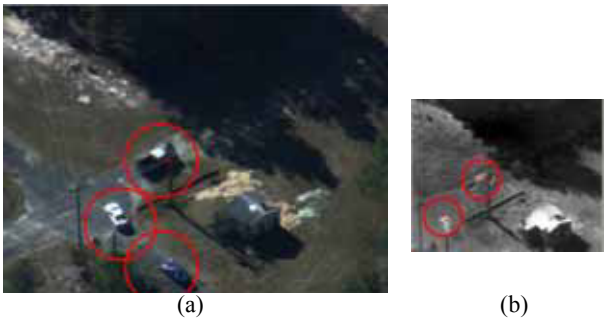
1 Introduction

Image fusion is a process of combining multiple images to form a single image by utilizing certain features from each image. The successful fusion of images acquired from different modalities or instruments is of great importance in many applications such as image analysis and computer vision, concealed weapon detection, and autonomous landing guidance. Image fusion can be performed at four levels of the information representation, which are signal, pixel, feature, and symbolic levels. Multi-scale transforms are widely used for analyzing the information content of images for image fusion. Several multiscale transforms have become very popular. These include the Laplacian pyramid transform [1], the contrast pyramid transform [2], the gradient pyramid transform [3], and the discrete wavelet transform (DWT) [4]. A comparative study of these methods is given in [5]. Recently, a new method that is based on trajectory association is proposed for image fusion [6]. Many of these works hand the still images. This paper describes a novel approach for fusing optical and infrared (IR) image sequences collected by an airborne platform and its application to target detection. A new algorithm is proposed for the effective fusion of airborne images from heterogeneous cameras. First, moving objects within the optical and IR images are detected. Second, an object mapping process is applied to match the objects in the optical images with the object in the IR images to find a relation between the images. Third, the optical and IR images are fused and finally moving targets are detected using the fused image sequences. The main contribution of this work is the development and evaluation of a novel algorithm for fusion of the airborne optical and IR images that results in more effective target detection. The foci of this algorithm are the object-based image fusion and target localization.

G. Bebis et al. (Eds.): ISVC 2008, Part II, LNCS 5359, pp. 651–660, 2008.

2 Algorithm Description

This work discusses the fusion of images generated by an optical camera and an IR camera mounted on a helicopter, and the target detection from the fused images. Fig. 1 shows the samples of an optical image and an IR image. The fusion of these two types of images faces the following problems.



(a) (b)

Fig. 1. Two different images. (a) 640×480 optical image; (b) 320×256 IR image. Red circles will be explained in Section 2.1.1.

(i) Everything in the scene including background appears to be moving since the cameras are mounted on a moving platform;
(ii) The optical image is color image, and IR image is grayscale image but recorded as pseudo color image, *i.e.*, IR signature is recorded to R-, G-, and B-channels. The resolution is different (640×480 for optical image and 320×256 for IR image), and the ratio of width to height is different;
(iii) There are some region overlaps, however, those regions are unknown;
(iv) There are multiple targets in images, and the number of targets may change (exit or reenter the field of view of a camera).

To address these problems, we designed an object mapping based image fusion and target detection algorithm. The entire processing flow is shown in Fig. 2. This algorithm consists of image registration, image fusion, and target detection. This research assumes that multiple cameras are mounted on the same helicopter or an unmanned aerial vehicle (UAV). Therefore, it is only necessary to perform the image registration once using certain number of image sequences to determine a relative motion relation between the optical and IR cameras. After performing the image registration, the registration parameters are used for image fusion and target detection. The following explains these three components in detail.

2.1 Image Registration

Image registration is the process of transforming the different sets of images into a common coordinate system. As shown in Fig. 2, the image registration in this system includes object detection from both optical and IR image, and object mapping. The

object detection is based on the algorithm developed in our previous work [7]. This section first summarizes the object detection algorithm in Section 2.1.1. Then it mainly discusses the object mapping in Section 2.1.2.

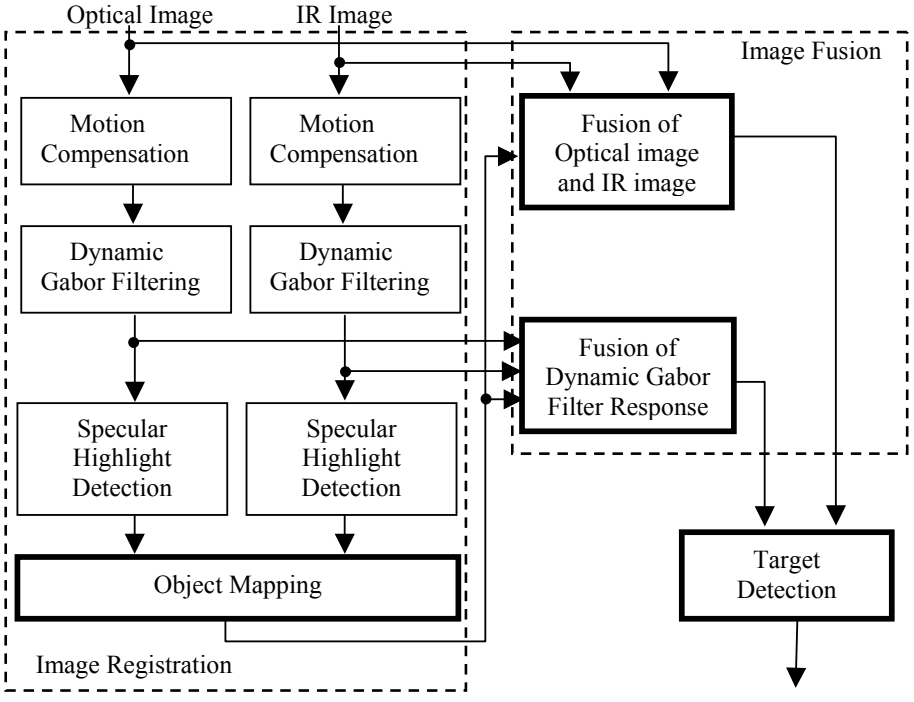


Fig. 2. Process flow of the entire algorithm

2.1.1 Object Detection

The object detection contains motion compensation, dynamic Gabor filtering (DGF), and specular highlight detection. Let F_i^τ denote the i-th image frame, where $\tau \in \{O, I\}$, and O and I represents the optical image and IR image, respectively. Then the object detection algorithm can be briefly summarized as follows. Details are referred to [7].

(i) For two consecutive frames, $F_{i-\Delta}^\tau$ and F_i^τ (Δ is the sampling interval), the feature points are detected by using Shi-Tomasi's method [8].

(ii) The optical flows between $F_{i-\Delta}^\tau$ and F_i^τ are detected by using Bouguet's algorithm [9]. The feature points are separated into inliers and outliers, where inliers are corresponding to the background, and outliers to the moving objects.

(iii) The inliers are used to estimate the affine transformation model between $F_{i-\Delta}^\tau$ and F_i^τ by using a RANSAC-like algorithm.

(iv) After the affine transformation model is determined, the frame difference is generated according to $F_{diff}^{\tau} = \left| F_i^{\tau} - \omega \times F_{i-\Delta}^{\tau} \right|$, where ω is the affine transformation model. Hence, the foreground can be separated from the background.

(v) DGF is applied to F_{diff}^{τ}, where the orientation of DGF is controlled by the optical flows corresponding to the inliers.

(vi) Specular highlights are detected. After DGF, the object detection becomes the detection of specular highlights. The detected highlights, after being filtered and merged, are considered as the objects.

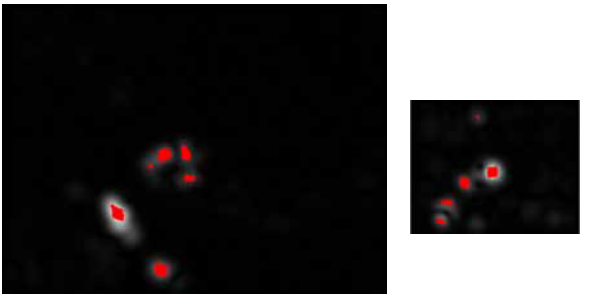


Fig. 3. Specular highlight detection results from images in Fig.1. (a) and (b), respectively

The detected highlights and objects from the input images in Fig. 1 are shown in Fig. 3, and in Fig. 1 by red circles. In the following, the objects detected from the optical image and IR image are denoted by $O^O = \{O_1^O, O_2^O, ..., O_M^O\}$ and $O^I = \{O_1^I, O_2^I, ..., O_N^I\}$, respectively, where M is the number of objects in optical image, and N in IR image.

2.1.2 Object Mapping

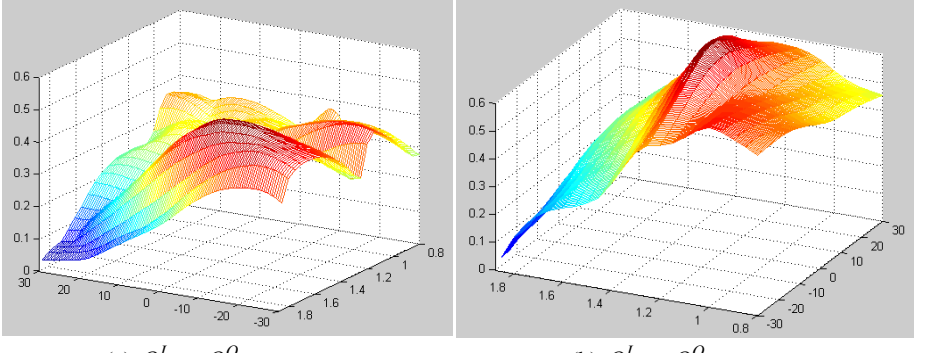
As shown in Fig. 1, the optical image and IR image are different in resolution, size, and width-to-height ratio. The registration/fusion of these two images can be defined as,

$$F_{fuse} = F_i^O \oplus \tilde{F}_i^I(s, \theta, \boldsymbol{\lambda}), \tag{1}$$

where $\tilde{F}_i^I(s, \theta, \lambda)$ is the output image of the IR image F_i^I after being enlarged by scaling factor s, translated by the translation vector $\boldsymbol{\lambda}$, and rotated by angle θ, and $\oplus$ is the image fusion operator. The task of image registration is to find s, θ, and $\boldsymbol{\lambda}$, which will be discussed in this section. The task of image fusion is to find fusion operator $\oplus$, which will be discussed in Section 2.2.

To find s, θ, and $\boldsymbol{\lambda}$, we employ the brute force algorithm which is described below. For $O_m^O \in O^O$ and $O_n^I \in O^I$ (m = 1, 2, …, M, n = 1, 2, …, N), we extract the grayscale sub-image $I_{m,sub}^{O,L_O}$ from the color image F_i^O centered at C_m^O, and the grayscale

sub-image $I_{n,sub}^{I,Li}$ from the pseudo color image F_i^I centered at C_n^I, respectively, where C_m^O is the center of the object O_m^O, and C_n^I of the object O^I, L_O is the size of $I_{m,sub}^{O,L_O}$, and L_I of $I_{n,sub}^{I,L_I}$. Note that L_I is smaller than L_O. Template matching for $I_{n,sub}^{I,Li}$ and $I_{m,sub}^{O,L_O}$ is performed in the following way. $I_{n,sub}^{I,Li}$ is shifted over $I_{m,sub}^{O,L_O}$ in the range $i \in [0, L_O - L_I]$ and $j \in [0, L_O - L_I]$. At each position (i, j) in $I_{m,sub}^{O,L_O}$, $I_{n,sub}^{I,Li}$ is enlarged by scaling factor $s \in [s_{\min}, s_{\max}]$, and rotated by angle $\theta \in [\theta_{\min}, \theta_{\max}]$ around (i, j) to generated the image $\tilde{I}_{n,sub}^{I,Li}(s,\theta)$. Then $\tilde{I}_{n,sub}^{I,Li}(s,\theta)$ is matched with $I_{m,sub}^{O,L_O}$. The correlation coefficient is adopted as the matching measure because it always ranges from -1 to +1, and is invariant to brightness and contrast. This brightness/contrast invariance can be explained as below [10].



(a) O_0^I to O_0^O (b) O_1^I to O_1^O

Fig. 4. Object matching measure map for the optical objects and IR objects detected in Fig. 3. (a) O_0^I to O_0^O, (b) O_1^I to O_1^O

Let $\boldsymbol{x}$ be the column-wise vector obtained by copying the grayscale pixels of $\tilde{I}_{n,sub}^{I,Li}(s,\theta)$, and $\boldsymbol{y}$ be the vector by copying the grayscale pixels in the region of $I_{m,sub}^{O,L_O}$ to be correlated with $\tilde{I}_{n,sub}^{I,Li}(s,\theta)$. Then the brightness/contrast correlation can be written as a least square problem:

$$\boldsymbol{y} = \beta \boldsymbol{x} + \gamma \boldsymbol{1} + \boldsymbol{\varepsilon} \tag{2}$$

where β and γ is the contrast correction factor and brightness correction factor, respectively, $\boldsymbol{1}$ is a vector of 1's, and $\boldsymbol{\varepsilon}$ is the vector of residual error. The problem is to find β and γ that minimizes $\boldsymbol{\varepsilon}^2$. This problem has a computationally fast solution. Let $\tilde{\boldsymbol{x}} = \boldsymbol{x} - \bar{x}$ and $\tilde{\boldsymbol{y}} = \boldsymbol{y} - \bar{y}$ be the mean-corrected vectors, where $\bar{x}$ and $\bar{y}$ is the mean of $\boldsymbol{x}$ and $\boldsymbol{y}$, respectively. Then,

$$\beta = \frac{\tilde{\boldsymbol{x}}\tilde{\boldsymbol{y}}}{\tilde{\boldsymbol{x}}^2}, \quad \gamma = \bar{y} - \beta \bar{x}, \text{ and } \varepsilon = \tilde{y} - \beta \tilde{x}. \tag{3}$$

The correlation coefficient r_{xy} can be calculated as,

$$r_{xy} = \frac{\tilde{\boldsymbol{x}}\tilde{\boldsymbol{y}}}{\|\tilde{\boldsymbol{x}}\|\|\tilde{\boldsymbol{y}}\|} = \frac{\beta\,\tilde{x}^2}{\|\tilde{\boldsymbol{x}}\|\|\tilde{\boldsymbol{y}}\|}. \tag{4}$$

This matching for $\tilde{I}^{I,Li}_{n,sub}(s,\theta)$ and $I^{O,L_O}_{m,sub}$ is performed at all position (i, j) for all $s \in [s_{\min}, s_{\max}]$, and $\theta \in [\theta_{\min}, \theta_{\max}]$, where $i \in [0, L_O - L_I]$ and $j \in [0, L_O - L_I]$. At each step, the matching measure in Eq. (4) is calculated. The above matching is repeated for all $O^O_m \in O^O$ and $O^I_n \in O^I$, where $m = 1, 2, \ldots, M$, $n = 1, 2, \ldots, N$. After this matching process, $M \times N$ measure maps are obtained. And next step is to search these $M \times N$ matching measure maps and find the maximal matching measure peak r_{maxpeak}. For the matching measure map for $O^O_m \in O^O$ and $O^I_n \in O^I$, if the matching measure takes r_{maxpeak} at the scale s_p, rotation angle θ_p, and position (i_p, j_p), then s_p, θ_p, and the translation vector $\boldsymbol{\lambda}_p = (x^O_m - x^I_n + i_p, y^O_m - y^I_n + j_p)^T$ are considered as the best scale, rotation angle, and the translation vector for the matched object pair O^O_m and O^I_n, which is also considered as the best scale, rotation angle, and translation vector for the IR image F^I_i to match the optical image F^O_i, where (x^O_m, y^O_m) and (x^I_n, y^I_n) is the center coordinates of O^O_m and O^I_n, respectively, and $s_p \in [s_{\min}, s_{\max}]$, and $\theta_p \in [\theta_{\min}, \theta_{\max}]$. The matching measure maps for mapping between O^I_0 and O^O_0, and O^I_1 and O^O_1 (others are omitted, here), detected in Fig. 3 (c) and (d), are shown in Fig. 4, where $s \in [0.8, 1.8]$, and $\theta \in [-30^\circ, 30^\circ]$, and at each step s is increased by 0.05, and θ by 0.5. The best matched object pair is O^O_1 to O^I_1, and the scale, rotation angle, and translation vector is 1.40, -0.6°, and $(73, 41)^T$, respectively.

2.2 Image Fusion

After the image registration parameters, s_p, θ_p, and $\boldsymbol{\lambda}_p$ are determined, the image fusion can be performed according to Laplacian pyramid transform [1], the contrast pyramid transform [2], the gradient pyramid transform [3], or DWT [4]. Currently, we are using weighted image average technique. First, the scaling, rotation, and translation operations are applied to the IR image F^I_i by employing parameters s_p, θ_p, and $\boldsymbol{\lambda}_p$, to generate the image $\tilde{F}^I_i(s_p, \vartheta_p, \lambda_p)$. Then $\tilde{F}^I_i(s_p, \vartheta_p, \lambda_p)$ and F^O_i are fused according to,

$$F^f_i = \kappa_1 \tilde{F}^I_i(s_p, \vartheta_p, \lambda_p) + \kappa_2 F^O_i, \tag{5}$$

where κ_1 and κ_2 are weighting coefficients, and superscript f on left hand side means *fuse*. Similarly, DGF response G^O_i and G^I_i of the optical image F^O_i and IR image F^I_i are also fused according to,

$$G^f_i = \kappa_1 \tilde{G}^I_i(s_p, \vartheta_p, \lambda_p) + \kappa_2 G^O_i. \tag{6}$$

2.3 Target Detection

Let us first clarify two technical terms, object detection and target detection. So far we use object detection, but in this section we start using target detection. They are basically the same, but this paper makes a difference in the following sense. The object detection means to detect objects from one or two image frames. The target detection means to detect objects from a short image sequence. This algorithm employs L frames, that is, F_{i-L}^f, F_{i-L+1}^f, …, F_i^f (currently L=10), to localize the target. The target detection algorithm is described as follows.

(i) Detect the specular highlights from the fused DGF response G_k^f ($k = i\text{-}L$, …, i) to locate the objects in the fused image F_k^f, by using the algorithm summarized in Section 2.1.1. The object detected from F_k^f is denoted by $O_{q,k}^f$, where q = 1, 2, …, Q, and Q is the number of objects. $O_{q,k}^f$ is represented by it center coordinates $C_{q,k}^f$, circumscribed rectangle $R_{q,k}^f$, and circumscribed ellipse $E_{q,k}^f$.

(ii) All objects detected from F_{i-L}^f, F_{i-L+1}^f, …, F_{i-1}^f are transformed to image frame F_i^f by using,

$$\tilde{C}_{q,k}^f = \omega_k^{k-1} \times \omega_{k-1}^{k-2} \times \cdots \times \omega_{i-1}^{i} \times C_{q,k}^f \tag{7}$$

where ω_{m-1}^m is the affine motion from frame m-1 to m, determined at the step of object detection.

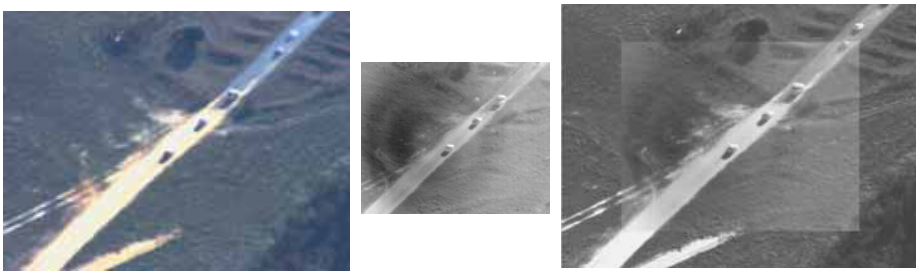
(iii) All objects are clustered by the grid-clustering method [11]. A filtering operation based on the cluster density with threshold d_{thres} is applied to the obtained clusters. Then each of the left clusters is considered as a target.

3 Experimental Results

The above algorithm is implemented by using MS-Visual C++ 6.0 and Intel Open CV on Windows platform. The frame interval Δ for object detection is set at 1, the searching range for s and θ in object mapping is set at [0.8, 1.8] and [-30˚, 30˚], respectively, and the increment for s and θ is 0.05 and 0.5˚, accordingly. The weighting coefficient κ_1 and κ_2 for image fusion are both set at 0.5. The image sequence length L for target localization is set at 10. The threshold d_{thres} for cluster filtering is set at 0.65.

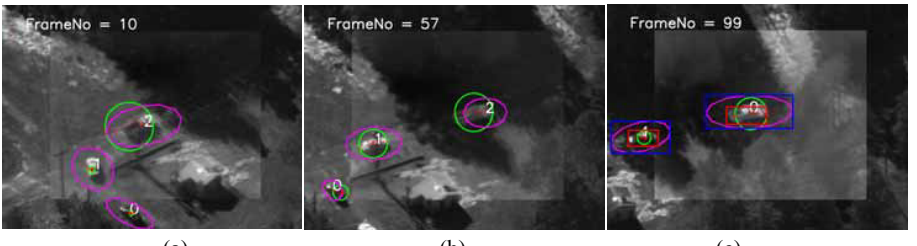
Fig. 5 shows some image fusion results. (a) and (b) shows an optical image and an IR image, respectively, and (c) is the fused image. From the fused image in (c) and (f), we can see the targets become clear and easy to detect.

Fig. 6 shows some targets detected at frame 10, 57, and 99, respectively. The system outputs the detection results from 10-th frame because the target localization employs 10 frames. The green circles mean the clustering results of the detected objects over 10 frames, and the purple ellipses the localized targets. Because the system



(a) (b) (c)

Fig. 5. Image fusion result. (a) Optical image; (b) IR image; (c) Fused image.



(a) (b) (c)

Fig. 6. Target detection result in frame 10, 57, and 99

Table 1. Processing time for image registration/fusion and target detection

Processing	Specific Task	Time (ms)
Image Registration/ Fusion	Optical object detection	1170
	IR object detection	437
	Object mapping	60403
	Fusion of optical and IR Image	16
	Total	62026
Target Detection (average)	Optical object detection	1171
	IR object detection	426
	Fusion of optical and IR Image	16
	Fusion of DGF response	16
	Target localization	98
	Total	1727

employs the object detection history (10 frames), the system can still detect the targets although they are lost shortly. This happens when objects enter the tree shade area, or go behind the tree leaves.

4 Performance Analysis

The algorithm described above is tested on Windows Vista machine mounted with an Intel Core 2 CPU and 2GB memory, running at 2.33GHz.We employed a 200-frame optical video sequence and a 200-frame IR video sequence to test the performance of the entire algorithm. The videos are sampled at interval $\Delta=2$, *i.e.*, totally it uses 200 optical and IR image frames. The resolution is 640×480 full color for optical image, 320×256 pseudo color for IR image. Each frame contains 2 to 3 objects, totally there are 270 objects. We use the ground truth data to evaluate this algorithm. As shown in Fig. 6 (c), the ground truth targets are shown by the red rectangle, and the detected targets are shown by blue rectangle (the circumscribed rectangle of the detected object). The detected objects are 204. From frame 59 to 70 (totally 36 objects), there are no objects detected because the moving displacement is too small. If we subtract these frames, the total object number becomes 224, the detection rate is 91%. The processing time is shown in Table 1. The image registration/fusion time is 62 seconds, and the average time for target detection is 1.7 seconds per frame.

5 Conclusions and Future Works

This paper proposed an algorithm for image registration/fusion and target detection from fused images, for the airborne optical and IR images. The image registration/fusion is based on object mapping. The technique for object mapping is invariant to rotation, scale, translation, brightness and contrast. The algorithm for target detection is based on the detection of specular highlights from fused DGF response and clustering technique. The experiment results show this algorithm is valid and efficient. The processing time for image registration/fusion is 62 seconds. This time is acceptable because this processing is executed only once (note that the optical camera and IR camera are mounted on the same moving platform). The average processing time for target detection is 1.7 seconds per frame. This time will be reduced to a half by resizing the optical image size to 320×240. Then the performance can be improved to 1.5 frames per second. This speed meets the requirements of many real-time applications. We performed some preliminary performance analysis. It needs to compare the target detection performance of this algorithm and those that use only the optical image or IR image. And further, the detection rate can be improved by using tracking technique. These are our current and future works.

Acknowledgement

This work was supported by a grant from the AFRL Information Institute, contract No. FA8750-08-1-0116.

References

1. Burt, P.J., Adelson, E.: The Laplacian pyramid as a compact image code. IEEE Trans. Communications 31(4), 532–540 (1983)
2. Toet, A.: Image fusion by a ratio of low-pass pyramid. Pattern Recognition Letters 9(4), 245–253 (1989)

3. Burt, P.J.: A gradient pyramid basis for pattern-selective image fusion, Society for Information Display. Digest of Technical Papers, 467–470 (1992)
4. Zhang, Z., Blum, R.S.: A categorization and study of multiscale-decompositionbased image fusion schemes. Proc. of the IEEE, 1315–1328 (August 1999)
5. Sadjadi, F.: Comparative Image Fusion Analysis. In: Proc. of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005 (2005)
6. Sheikh, Y.A., Shah, M.: Trajectory Association across Multiple Airborne Cameras. IEEE Trans. Pattern Anal. Mach. Intell (accepted)
7. Yao, F.H., Sekmen, A., Malkani, M.: A Novel Method for Real-time Multiple Moving Targets Detection from Moving IR Camera. In: Proc. of ICPR 2008 (accepted)
8. Shi, J., Tomasi, C.: Good features to track. In: Proc. of 9th IEEE Conference on Computer Vision and Pattern Recognition. Springer, Heidelberg (1994)
9. Bouguet, J.Y.: Pyramidal Implementation of the Lucas Kanade Feature Tracker Description of the algorithm, Intel Corporation (2003)
10. Kim, H.Y., Araújo, S.A.: Grayscale Template-Matching Invariant to Rotation, Scale, Translation, Brightness and Contrast. In: Mery, D., Rueda, L. (eds.) PSIVT 2007. LNCS, vol. 4872, pp. 100–113. Springer, Heidelberg (2007)
11. Schikuta, E.: Grid-Clustering: A fast hierarchical clustering method for very large data sets, CRPCTR93358 (1993)