Pattern Anal Applic (2006) 8: 375–389
DOI 10.1007/s10044-005-0020-8

THEORETICAL ADVANCES

Zheng Liu · Zhiyun Xue · Rick S. Blum
Robert Laganière

Concealed weapon detection and visualization in a synthesized image

Received: 11 February 2005 / Accepted: 17 May 2005 / Published online: 16 December 2005

Abstract Images acquired by heterogeneous image sensors may provide complementary information about the scene, for instance, the visual image can provide personal identification information like the facial pattern while the infrared (IR) or millimeter wave image can detect the suspicious regions of concealed weapons. Usually, a technique, namely multiresolution pixel-level image fusion is applied to integrate the information from multi-sensor images. However, when the images are significantly different, the performance of the multiresolution fusion algorithms is not always satisfactory. In this study, a new strategy consisting of two steps is proposed. The first step is to use an unsupervised fuzzy *k*-means clustering to detect the concealed weapon from the IR image. The detected region is embedded in the visual image in the second step and this process is implemented with a multiresolution mosaic technique. Therefore, the synthesized image retains the quality comparable to the visual image while the region of the concealed weapon is highlighted and enhanced. The experimental results indicate the efficiency of the proposed approach.

Keywords Concealed weapon detection · Unsupervised clustering · Multiresolution mosaic · Image fusion

This material is based on part of the work carried out at the SPCR laboratory of Lehigh University and the work is partially supported by the U. S. Army Research Office under grant number DAAD19-00-1-0431. The content of the information does not necessarily reflect the position or the policy of the federal government, and no official endorsement should be inferred.

Z. Liu (✉) · R. Laganière
School of Information Technology and Engineering
Faulty of Engineering, University of Ottawa, SITE-5025,
800 King Edward Ave, P.O. Box 450 STN A, Ottawa,
ON K1N 6N5, Canada
E-mail: zliu086@uottawa.ca
Tel.: +1-613-5625800
Fax: +1-613-5635664

Z. Xue · R. S. Blum
Signal Processing and Communications Research Lab,
Department of Electrical and Computer Engineering,
Lehigh University, 19 Memorial Drive West, Bethlehem,
PA 18015-3084, USA
E-mail: rblum@eecs.lehigh.edu
Tel.: +1-610-7583459
Fax: +1-610-7586279

1 Introduction

To address the emerging threats from terrorists, there is a need to develop an efficient technique for heightened security requirements and law enforcement. Currently, airport personnel examine passengers using metal detector, hand wands, and physical searches [1]. Hence, passengers with concealed objects may not be detected. Imaging systems with a radiation wavelength longer than 20 μm can penetrate clothing and thus have the potential to detect concealed weapons [2]. The enabling sensing mechanisms being studied include infrared (IR), acoustic, millimeter wave (MMW), X-ray sensors and so on [3]. Multiple image modalities of different radiation wavelengths can provide complementary information about the scene. Therefore, the use of multiple sensing modalities can increase the overall performance in comparison with single sensor systems. A technique, namely image fusion provides a solution to combine information from multiple images and generates a single image that gives a more accurate or complete description of the scene than any of the individual source images [4]. The application of image fusion techniques can be found in a wide range of applications including multi-focus imagery, concealed weapon detection (CWD), intelligent robot, surveillance system, medical diagnosis, remote sensing, non-destructive testing, etc. [3–14].

Based on requirements of the CWD application, there are different ways to implement the multi-modal image fusion process. The principle idea is illustrated in Fig. 1. The first purpose of fusion is to facilitate the detection process. Like the circle in Fig. 1, the fusion operation is to achieve an enhanced result for easing further analysis, recognition, or classification. Varshney

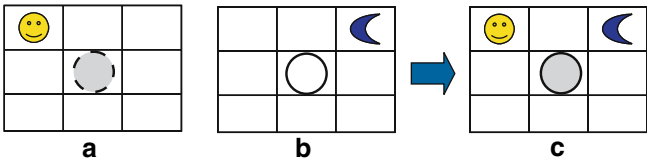


Fig. 1 The illustration of image fusion techniques for concealed weapon detection applications. **a**, **b** Input images results, **c** fusion result. The input images are assumed to be fully registered

et al. presented an automatic procedure to register and fuse IR and MMW images in [7]. However, the study on how further analysis can benefit from the fusion result is not available yet. The second purpose of CWD fusion is to locate human subjects with possible concealed weapons by fusing electro-optical (EO) and IR/MMW images [15]. Like the face and moon in Fig. 1, the fused image contains both the pieces of personal information, i.e., facial pattern, and the highlighted concealed weapon region. This fusion is carried out at the pixel level as well. A human operator is presented with a composite image, with which the operator can respond accurately and promptly [9, 15, 16]. Another important issue that has not been addressed yet is the "privacy rights". The multi-modal image device cannot be used as a tool for voyeurism [17]. Therefore, the fusion algorithm must be tuned to reveal only the concealed weapon's information instead of personal privacy to the operators. The work presented in this paper will focus on the second scenario, where a visual image is involved. From now on, we will use the terminology "CWD" to refer to the second topic described above.

The current study on multiresolution image fusion for CWD is to generate a composite image for the operator or an automated analysis procedure as show in Fig. 2a. We suggest a new image processing architecture in Fig. 2b. Each pixel from the IR and/or

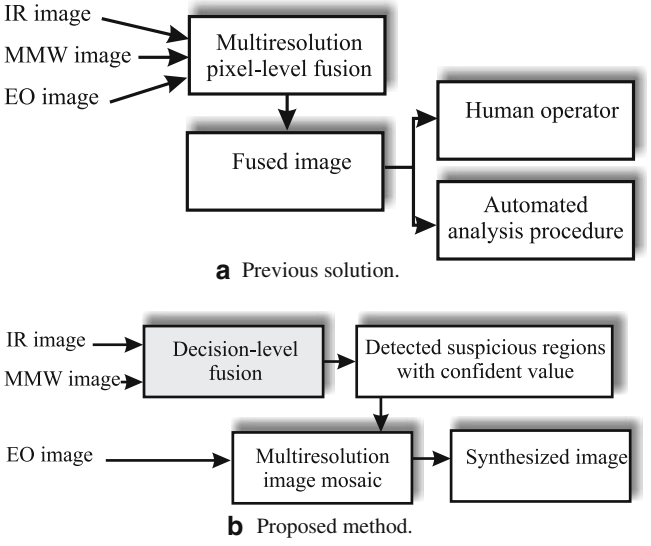


Fig. 2 The image processing architectures for CWD applications. **a** Previous solution. **b** Proposed method

MMW images is classified with a confident value as belonging to either a weapon or a non-weapon region. This can be implemented at a higher level (decision level instead of pixel level). The detected region is further segmented by a predefined confident threshold and embedded into the corresponding visual image by using a multiresolution image mosaic (MRIM) technique, which can achieve a seamless boundary between host image and embedded regions. In our work, only the selected (weapon region) parts are synthesized with the visual image, because other parts do not contribute to the weapon detection at all.

The philosophy of the proposed approach is different from previously published work, where a pixel-level fusion is carried out for the whole image. In our study, the weapon is first detected from an IR image by an unsupervised clustering algorithm namely fuzzy *k*-means clustering. The feasibility of the clustering algorithm on IR or MMW image is investigated. The detected region is used as a mask signal for the MRIM process. The steerable image pyramid is employed to decompose and reconstruct the two images. The reconstruction generates the final result. The rest of the paper is organized as follows: in Sect. 2, we first briefly review the preprocessing and image fusion techniques for CWD applications. A two-step scheme for synthesizing a composite image is described in Sect. 3. Experimental results can be found in Sect. 4. Discussion and conclusion are presented in Sects. 5 and 6, respectively.

2 Brief overview

2.1 Image fusion for CWD

The most popular solution to a CWD application is to fuse EO and IR/MMW images at pixel level with a so-called multiresolution analysis (MRA) technique [18–20]. The fusion process is carried out in the transform domain through the combination of the transform coefficients. A brief comparison of the algorithms is summarized herein in Table 1. Generally, the study of multiresolution image fusion (MRIF) is twofold, encompassing a multiresolution algorithm and a coefficient combination rule. A number of MRA algorithms have been investigated for the fusion of multi-sensor images so far. For detailed implementation, readers are referred to the relevant publications in Table 1. The choice of wavelet largely depends on the characteristics of the algorithm and the signal to be processed. Since an image is represented as a weighted sum of basis functions, choosing the basis function that resembles the signal will facilitate the analysis. The major steps of MRIF include: image decomposition, coefficient combination, and image reconstruction. The basic rule for coefficient combination is the absolute value maximum selection for high frequency bands and averaging for the low-pass band, i.e., the coefficients with larger absolute value from the high frequency bands will

Table 1 Comparison of multiresolution image fusion schemes

Multiresolution approaches		Contents				
		Researchers	Fusion rule	Contribution	Evaluation	Applications
Image pyramid	Laplacian pyramid	Adelson et al. [18]	Absolute value maximum selection (AVMS)	First study on MRIF		Multi-focused images
	Ratio of low pass pyramid	Toet [28, 29]	Maximum absolute contrast selection	Use of RoLP	Subjective evaluation	Simulation
	Gradient pyramid	Burt and Kolczynski [4]	Fusion based on match and salience measure	Image feature based fusion		Fusion of IR and visible image, multi-exposure, multi-focus images
		Wilson et al. [23, 24]	Weighted average	Perceptual-based fusion	SNR	Hyperspectral image
	Morphological pyramid	Matsopoulos et al. [11]	Maximum operation	Use of morphological pyramid	Cross-correlation	CT and MRI images
	Steerable pyramid	Liu et al. [30]	Apply Laplacian pyramid and AVMS rule for sub-images	Iterative fusion of sub-images	RMSE	Standard images for simulation, multi-sensor images
Wavelet	Orthogonal wavelet	Li et al. [20]	AVMS	Consistency verification, concept of region-based fusion	RMSE	Multi-focus images, multi-sensor images
	Steerable dyadic wavelet	Koren et al. [12, 26]	Maximum local oriented energy	Image feature represented with oriented energy	MSE	Different channels of landsat TM images
	Discrete wavelet frame	Rockinger [31–33]	Same as Burt and Kolczynski [4]	Studies on temporal stability and consistency		Image sequences
		Zhang and Blum [25]	Activity measure, region-based rule, grouping approach	Studies on region-based approach and grouping method	RMSE, mutual information, percentage of correct decision	Multi-focus images, MMW images, IR images
	Contrast-based wavelet	Pu and Ni [34]	Absolute value maximum selection	Present the concept of directive contrast	SNR	Infrared and visual images
	Complex wavelet	Nikolov et al. [35]	Chain representation fusion	Use of complex wavelet		Multi-focus images, CT and MR images
	Multiwavelet	Wang et al. [36]	Pixel selection based on the image's feature map	Use of multiwavelet	Subjective evaluation	SPOT images

be retained and used for reconstruction, because the larger values correspond to image features like edges, lines, or boundaries. More sophisticated rules will consider the area or region around the pixel and the corresponding areas or regions at the other frequency bands or resolutions [21, 22]. Image feature measurements in a region or across the frequency bands are generated. A selection rule is created or the weighting coefficients are derived from such measurements. The concept of match measure and salience measure originated from Burt's work on gradient pyramid based image fusion [4], where the match measure determined the selection or averaging operation while the salience measure chose the coefficients for the reconstruction in the selection mode. Wilson et al. [23, 24] introduced the contrast sensitivity measure to weigh the coefficient sets. Li's rule for coefficient selection was based on a 3×3 or 5×5 window [20], where the pixel with the maximum absolute value in the window represented the activity of the pixel located at the center. Li also introduced consistency verification as a rectification of the selection process. Zhang and Blum [25] used the average value in the region contoured and segmented by the Canny detector instead of the maximum pixel value to guide the fusion process. Thus, the approach is more robust to the noise. Koren et al. [26] used the local oriented energy as a metric of image feature and the coefficient selection was based on such measurement. Cross-band selection and coefficient grouping methods were proposed by Xydeas and Zhang in [27] and [25], respectively. This is actually another consideration for the region effect, since a single pixel at a lower resolution corresponds to several pixels (region) at a higher resolution.

The fusion of IR and MMW images has been studied by Salmani [37] and Varshney [7], respectively. In [5] and [3], Uner and Slamani fused multiple IR images with a discrete wavelet transform. In [9], Xue and Blum did an extensive study on fusion of visual and IR images with different MRIF algorithms. The fused results were evaluated by a number of quantitative metrics. However, the visual quality of the fused image was degraded in most of the experimental results. The problem is that the MRA algorithms try to keep the salient features of images no matter whether the substance is really useful or not. One disadvantage of the MRIF approach is that when the two source images have a great difference, the selecting or even the averaging of the low pass components will cause the "block" effect in the fused result. In other words, the reconstruction is not stable. Lately, Xue presented a new color-based fusion algorithm, in which the IR image was fused with color channels [16]. Yang et al. employed the expectation-maximization algorithm to estimate the optimal scene in [38].

2.2 Processing for object extraction

As described in [15], the further processing is toward an automatic weapon detection. Commonly used object extraction approaches are based on thresholding or segmentation techniques. In Slamani's mapping procedure A'SCAPE [3], homogeneous regions are separated by applying a series of threshold values followed by a low- and high-pass filtering operation. The basic idea is to group pixels in homogeneous regions. In [7], the authors suggested the use of Otsu's thresholding method [39] to the fused result of IR and MMW images. However, there is no study on assessing the performance of these approaches so far.

3 A two-step scheme for synthesizing a composite image

The objective of synthesizing a visual and a non-visual image is to retain the information of both the personal identification and the concealed weapons. It is obvious that the IR image contributes little to the facial identification in the case of being fused with a visual image. Therefore, a simple combination may degrade the quality of the fusion result for facial identification. The detection of concealed weapon depends on the operation of the IR sensor, because the pixel value of the IR image reflects the variations in temperature. If the IR sensor cannot locate the concealed weapon, the fusion with a visual image will not generate a useful result. The variation in temperature distribution of different objects, i.e., weapon, clothing, and body, can be identified by using an unsupervised clustering approach. A two-step scheme consisting of a detecting and an embedding operation is proposed next.

3.1 Concealed weapon detection

3.1.1 Fuzzy k-means clustering

Fuzzy k-means clustering assigns a membership grade to a data point belonging to certain cluster [40]. It is an unsupervised approach for data clustering through seeking a minimum of heuristic global cost function [40]:

$$J = \sum_{i=1}^{c} \sum_{j=1}^{n} \left[\hat{P}\left(\omega_i \middle| x_j, \hat{\theta}\right) \right]^b (x_j - \mu_i)^2, \tag{1}$$

where the probability $\hat{P}\left(\omega_i \middle| x_j, \hat{\theta}\right)$ stands for the fuzzy membership of pixel x_j ($j = 1 \ldots n$) in a cluster ω_i ($i = 1 \ldots c$), and there are in total c clusters in the data set. μ_i indicates the mean value for each cluster ω_i. b is a free parameter chosen to adjust the blending of different clusters, while $\hat{\theta}$ is the parameter vector for the membership functions. The probabilities of cluster membership for each pixel are normalized as:

$$\sum_{i=1}^{c} \hat{P}(\omega_i | x_j) = 1, \quad j = 1, \cdots\cdots, n. \tag{2}$$

The minimization of the cost function in Eq. 1 leads to the solutions [40]:

$$\mu_j = \frac{\sum_{j=1}^{n} \left[\hat{P}(\omega_i|x_j)\right]^b x_j}{\sum_{j=1}^{n} \left[\hat{P}(\omega_i|x_j)\right]^b} \tag{3}$$

and

$$\hat{P}(\omega_i|x_j) = \frac{(1/d_{ij})^{1/(b-1)}}{\sum_{r=1}^{c} (1/d_{rj})^{1/(b-1)}} \quad \text{and } d_{ij} = (x_j - \mu_i)^2. \tag{4}$$

The cluster means and point probabilities are estimated iteratively until there is only small change in μ_j and $\hat{P}(\omega_i|x_j)$.

By applying the fuzzy k-means clustering algorithm to the IR images, a set of clustered images are obtained. The idea is similar to Slamani's SMP algorithm [37] in grouping pixels in homogeneous regions. It is observed that the cluster corresponding to the highest center value is the collection of the points in the concealed weapon region. By applying a proper threshold value, a binary mask image is obtained and used for the mosaic operation.

The fuzzy k-means clustering algorithm needs the number of clusters as an input parameter, which can be determined empirically. Calculating validity measure indexes can help estimate the goodness of the fuzzy clustering algorithm and find the optimal number of clusters [41]. Herein, four validity indexes are employed, i.e., partition index (SC), separation index (S), Xie and Beni's index (XB), and Dunn's index (DI) [41]. Readers are referred to [41–44] for detailed description and implementation of these metrics.

A small cluster number is better for computational efficiency. In Fig. 3, SC and S index hardly decrease at point 8 while XB and DI reach their local minimum at the same point. Therefore, in our experiments, we select eight as the initial number of clusters for the IR images.

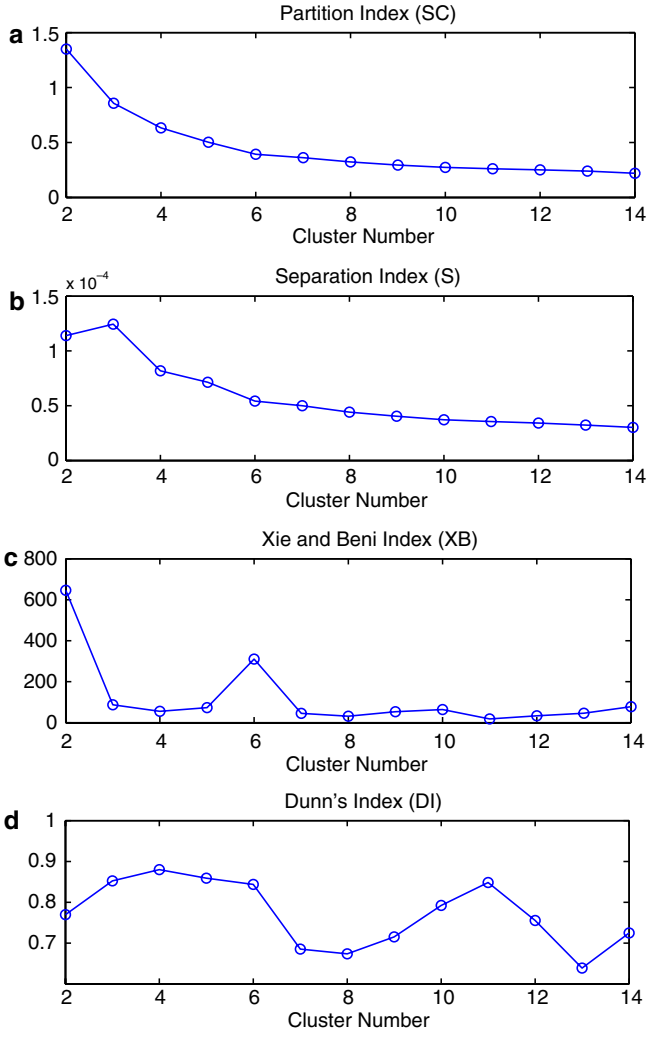


Fig. 3 The clustering indexes with different cluster numbers

3.1.2 *Region-of-interest (ROI) enhancement*

The aforementioned approach provides another advantage in that particular processing can be applied to the ROI partitioned by the mask image. On one hand, the synthesized image is evaluated by the operator; on the other hand, in further processing, different algorithms can be applied to the different ROI regions, respectively. For example, if we again apply the fuzzy k-means clustering algorithm to the ROI of an IR image, the shape of the weapon can be detected through finding out the cluster with the highest center value. With this information, the weapon in the IR image can be enhanced. If only the shape is enhanced, we can simply multiply the IR image with a gain map in which the value in the weapon region is larger than 1. Another enhancement scheme is to use the corresponding membership map from the IR image:

$$I_{\mathrm{IR}}(x,y) = I_{\mathrm{IR}}(x,y)(1 + \alpha \cdot F_{\mathrm{ROI}}(x,y)). \tag{5}$$

$F_{\mathrm{ROI}}(x, y)$ is the corresponding ROI fuzzy membership map. The pixel with the higher membership value is emphasized more by the parameter α. The next step is to follow previously descibed procedure to mosaic the visual image and the enhanced IR image.

3.2 Embedding in a visual image

3.2.1 *Steerable pyramid*

The multiresolution representation of an image is given as:

$$I(x,y) \rightarrow \left(LI_N(x,y), BI_i^j(x,y)\Big|_{i=1\ldots N}^{j=1\ldots K}\right) \tag{6}$$

or

$$I(x,y) \rightarrow \left(LI_N(x,y), HI_i^j(x,y)\Big|_{i=1\ldots N}^{j=1\ldots K}\right), \tag{7}$$

where $LI_N(x, y)$ denotes the low-pass component at the decomposition level N while $BI_j^i(x, y)$ and $HI_j^i(x, y)$ stand for the band-pass components or high-pass components at decomposition level i and orientation j, respectively. For Laplacian pyramid, there is only one band-pass image at each decomposition level, therefore $K=1$. For Daubechies wavelet, K equals to 3. In the experiments, we use another representation named steerable pyramid proposed by Simoncelli et al. [45]. The image will be presented as:

$$I(x,y) \rightarrow \left(LI_N(x,y), BI_i^j(x,y)\Big|_{i=1\ldots N}^{j=1\ldots K}, HI(x,y)\right). \tag{8}$$

The representation in the transform domain consists of three parts: one low-pass component, one high-pass component, and K band-pass components. Each band-pass component corresponds to an orientation angle $(i-1)\pi/4$, where $i=1 \ldots K$. Although the representation is overcompleted, it has the advantage that sub-bands are both translation and rotation invariant. Oriented features can be extracted by using the steerable pyramid representation. The structure of the steerable pyramid is shown in Fig. 4, where one high-pass filter $H_0(\omega)$, two low-pass filters $L_0(\omega)$ and $L_1(\omega)$, and a set of band-pass filters $B_k(\omega)$ are involved. k ranges from 1 to K.

To eliminate aliasing, avoid amplitude distortion, and cascade the system recursively, the following conditions should be satisfied:

$$Ł_1(\omega) = 0 \text{ for } |\omega| > \frac{\pi}{2} \tag{9}$$

$$|H_0(\omega)|^2 + |L_0(\omega)|^2 = 1 \tag{10}$$

$$|L_1(\omega)|^2 + \sum_{k=1}^{K} |B_k(\omega)|^2 = 1. \tag{11}$$

For more information about the steering theory and details of filter design, readers are referred to references [45, 46].

3.2.2 Procedure for image mosaic

The idea of MRIM is to combine two or more images into a composite one with an invisible seam [4, 47]. The general procedure is shown in Fig. 5. Like the multi-resolution image fusion process, the input images are decomposed by a certain multiresolution algorithm Ψ. Meanwhile, the Gaussian pyramid of the binary mask image is constructed $GI_N(x, y)$, ... $GI_2(x, y)$, $GI_1(x, y)$, where N is the decomposition level. The new image components can be formed by the weighted sum with the Gaussian image components. There are several ways to achieve this.

The first implementation is achieved by the weighted summation of every image component. The formulae are given below:

$$HI(x,y) = GI_1(x,y) \cdot HI_{\mathrm{IR}}(x,y) + (1 - GI_1(x,y)) \cdot HI_{\mathrm{V}}(x,y) \tag{12}$$

$$BI_i^j(x,y) = GI_i(x,y) \cdot BI_{\mathrm{IR}i}^j(x,y) + (1 - GI_i(x,y)) \cdot BI_{\mathrm{V}i}^j(x,y) \tag{13}$$

$$LI_N(x,y) = GI_N(x,y) \cdot LI_{\mathrm{IR}N}(x,y) + (1 - GI_N(x,y)) \cdot LI_{\mathrm{V}N}(x,y). \tag{14}$$

The new image components will be used to reconstruct the composite image. The second implementation uses the edge information of the mask image map. The original edge map can be easily obtained by the Canny edge detector. Instead of generating a Gaussian image pyramid, through the down-sampling operation, we can get a set of edge images $E_N(x, y)$, ... $E_2(x, y)$, $E_1(x, y)$ and mask images $M_N(x, y)$, ... $M_2(x, y)$, $M_1(x, y)$. Now, the combination formulae become:

$$\begin{cases} I_{\mathrm{IR}}(x,y) & M(x,y)=1, E(x,y)=0 \\ (I_{\mathrm{IR}}(x,y) + I_{\mathrm{V}}(x,y))/2 & E(x,y)=1 \\ I_{\mathrm{V}}(x,y) & M(x,y)=0, E(x,y)=1 \end{cases} \tag{15}$$

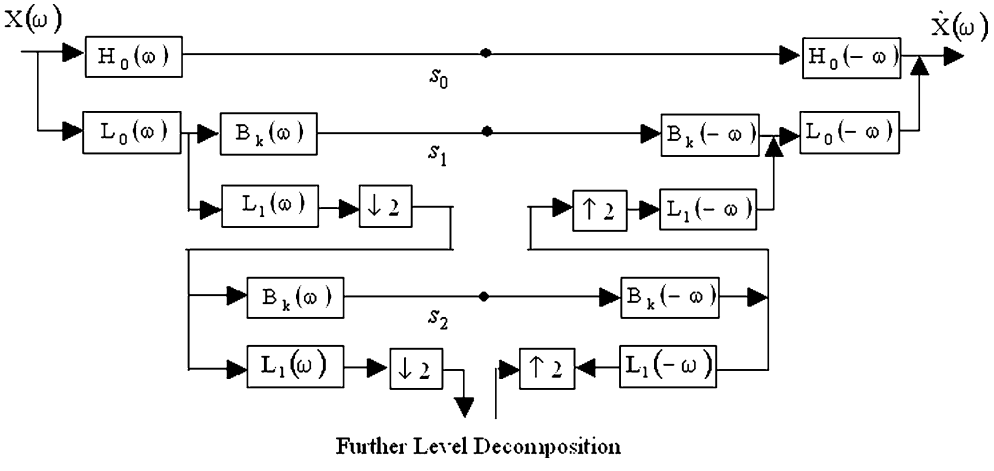


Fig. 4 The architecture of the steerable pyramid

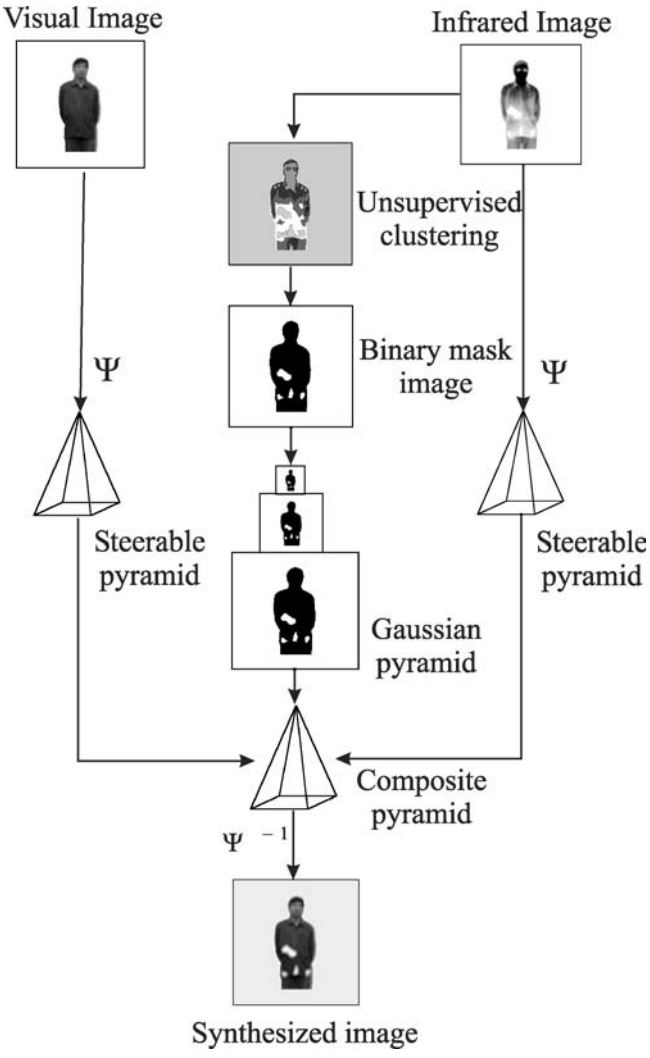


Fig. 5 The procedure for multiresolution image mosaic

$$LI(x,y) = M_N(x,y) \cdot LI_{\mathrm{IR}N}(x,y) + (1 - M_N(x,y)) \cdot LI_{\mathrm{V}N}(x,y). \quad (16)$$

The operation will copy the corresponding regions from the visual and IR images to the new image component, i.e., "cut and paste". At the edge between the two regions, an average operation is applied. In the above equations, I_{IR} (x, y), I_{V} (x, y), and $I(x, y)$ stand for the high- and band-pass image components of IR, visual, and new images, respectively. For the low-pass component, we do not use the edge to smooth the transition zone. The discussion can be found in Sect. 5. The third implementation differs from the first in the combination of low pass components. For the high-pass and band-pass components, Eq. 12 and 13 are applied. The low pass component from the visual image is retained as the new low pass component for reconstruction; or a weighted summation is implemented in the marked weapon region by the mask image map. Such operations can also be applied for texture mapping [47].

3.3 Result assessment

An ideal solution for evaluating the fused image is to compare it with a reference image, which is assumed to be perfect. However, such a reference image is not available in advance for the CWD application. The success of the application largely depends on whether the suspicious regions can be detected or not. Therefore, the classification metrics, accuracy and reliability, are employed herein. An illustration to interpret this concept is given in Fig. 6. Suppose A is the ground truth (true weapon region), B is the detected result (detected weapon region) and C is the overlap between A and B. The accuracy is defined as the ratio between the positively true and all pixels that are used as the ground truth of this class, i.e., $(C/A)\times 100\%$ while the reliability is expressed as $(C/B)\times 100\%$, i.e., the ratio between the positively true and all pixels classified as this class. A large accuracy value together with a higher reliability indicates a good classification result.

4 Experimental results

The multi-sensor image data was collected at the Signal Processing and Communication Laboratory of Lehigh University. There are nine pairs of visual and IR images shown in Fig. 7. In the following experiments, we assume: (1) the visual image and IR image are fully registered; (2) both the visual and IR image are background subtracted; and (3) there is a concealed weapon in each scene.

In the first part of the experiment, the first pair of images in Fig. 7 was integrated by image fusion algorithms. Figure 8 presents the results obtained by applying Laplacian pyramid, Daubechies wavelet and Simoncelli steerable pyramid based fusion algorithms, respectively. The coefficient combination rule is: averaging the low pass image components and applying the maximum selection rule to the high pass components. More sophisticated rules and algorithms were implemented in [4, 20, 21]. The steerable pyramid based algorithm was presented in [30] and applied to the image pair in Fig. 7a, b. Figure 8c, d give the results. The facial pattern is obscure in the pixel-level fusion results, although the weapon region can be observed to some extent.

In the second part of the experiment, the MRIM was implemented. As described in Sect. 3.2.2, there are three approaches that come with the multiresolution mosaic

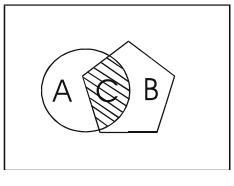


Fig. 6 Illustration for accuracy and reliability assessment

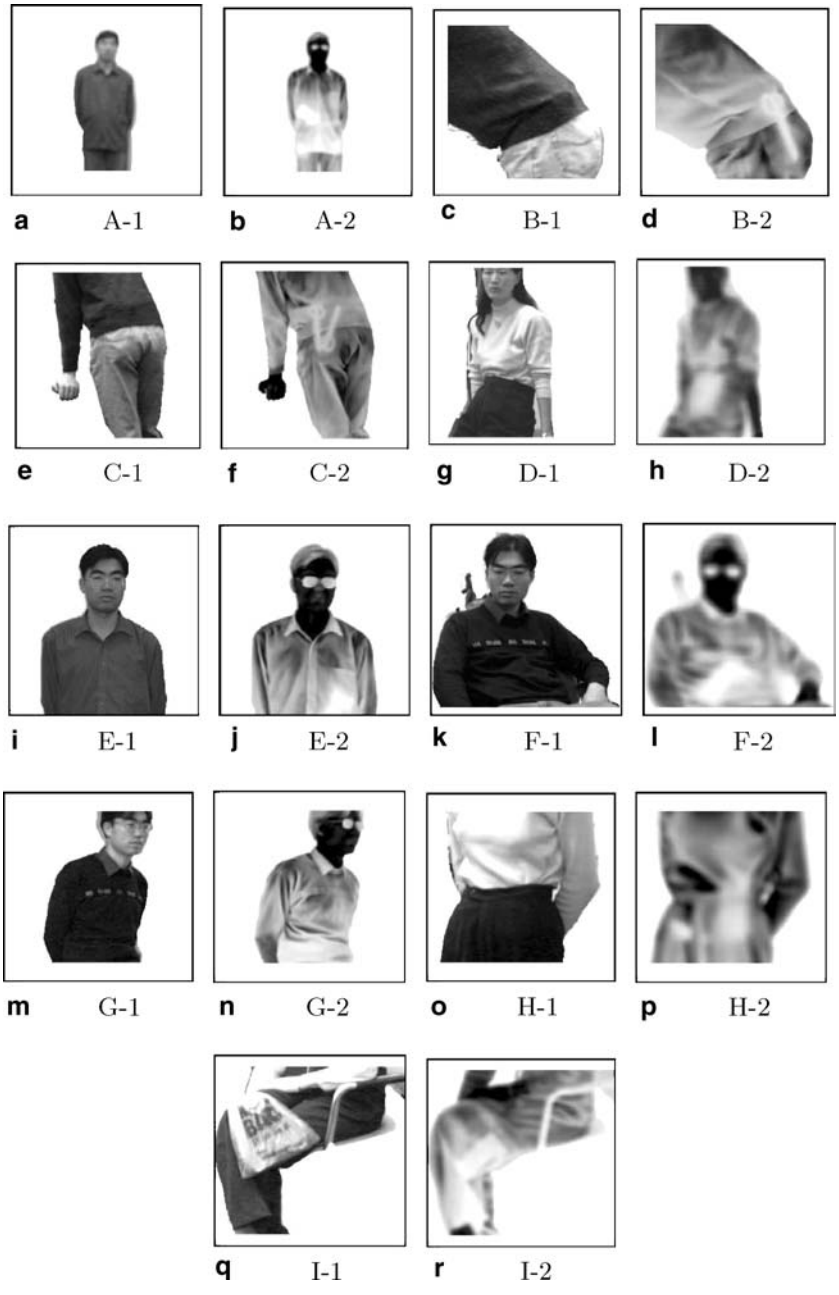


Fig. 7 Multi-sensor images used for testing in the experiment: totally eight groups are involved (A–I)

scheme. To apply the mosaic algorithm, the mask signal needs to be extracted. In Fig. 9a, the segmented result by applying fuzzy *k*-means clustering algorithm is shown. By selecting the cluster with the highest center value and applying a proper threshold value, the binary image map was obtained and given in Fig. 9b. In the experiment, the points in this cluster with a value larger than 0.1 were collected and averaged. The averaged value was selected as the threshold. With the binary mask image, the visual and IR images were synthesized by the proposed algorithms. The decomposition level of the multiresolution representation does affect the results. We gave the results with two, three, and four level decomposition in Fig. 10.

To see how the number of clusters affects the detection of weapon region in terms of accuracy and reliability measurements, we used a set of numbers in Table 2 to cluster IR image of Fig. 7b and compared the detected results with a manually generated reference image. Figure 11 shows the curve. A larger cluster number can achieve a higher reliability at the cost of losing accuracy; meanwhile, a larger number will introduce computational loads. For the CWD application, a higher accuracy has priority over reliability in most cases.

In addition, we compared the fuzzy *k*-means clustering method with the expectation-maximum (EM) clustering and *k*-means clustering methods. The three clustering algorithms were applied to the nine groups of multi-sensor images with the same cluster number 8. The results of accuracy and reliability assessments are listed

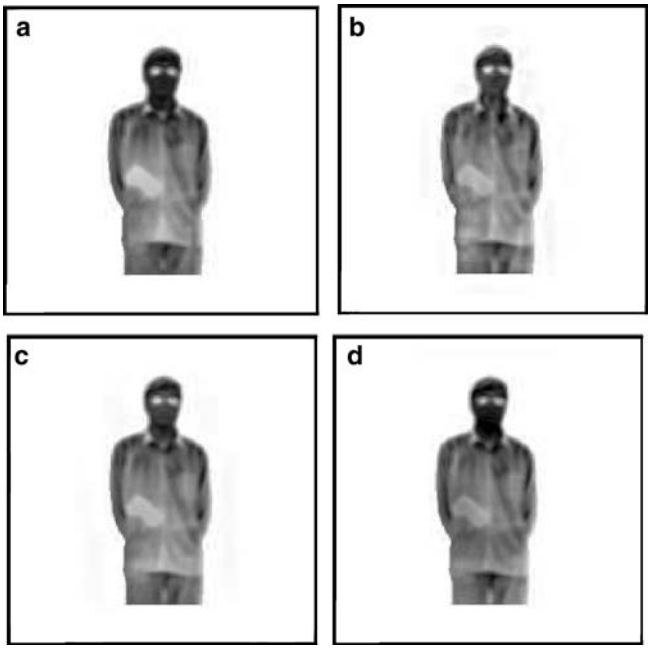


Fig. 8 Image fusion results achieved by **a** Laplacian pyramid; **b** Daubechies wavelet four; **c** Simoncelli steerable pyramid (averaging for low-pass component and maximum selection for band- and high-pass components); and **d** Simoncelli steerable pyramid with sub-band images integrated by Laplacian pyramid)

in Table 3 and illustrated in Fig. 12. In terms of classification rate, the fuzzy clustering does not show obvious advantages over the other approaches. Nevertheless, the outputs of fuzzy clustering can be used to enhance the ROI in the IR image. The concealed weapons in Fig. 4b, c have explicit shapes. The enhancement may facilitate further processing. First, we used the binary mask image to extract the ROI of the IR image. Then, the ROI was segmented again by the clustering algorithm. The region of the concealed weapon was further refined. By using the fuzzy membership map of the ROI, the IR image can be enhanced according to Eq. 16. The visual image was then synthesized with the enhanced version of the IR image. Figures 13 and 14 show the results.

From the above experiments, we can see that the third multiresolution mosaic approach with a decomposition level two achieved a better result in terms of human perception. Eventually, we applied this approach to the other images and have given the results in Fig. 15.

5 Discussion

The advantages of pixel-level fusion of IR and MMW images are not explicitly identified; therefore, a decision-level fusion for classification is suggested. In this study, we did not implement the shaded block in Fig. 2b, which

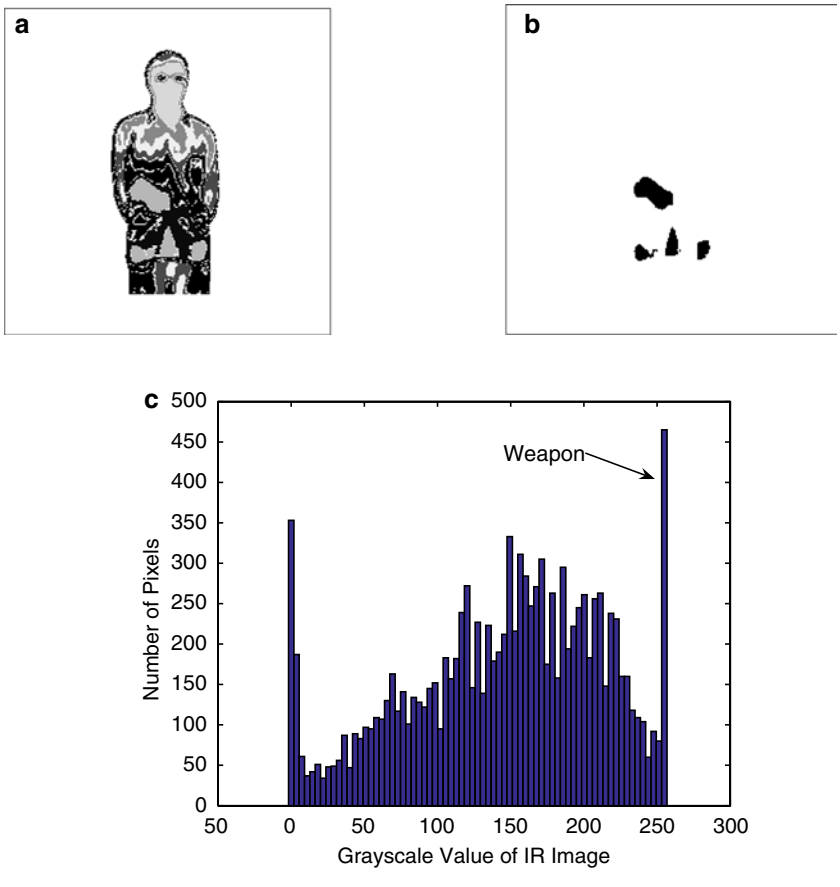


Fig. 9 **a** Clustered image by fuzzy *k*-means clustering algorithm; **b** binary mask image obtained from the clustered result; and **c** histogram of IR image

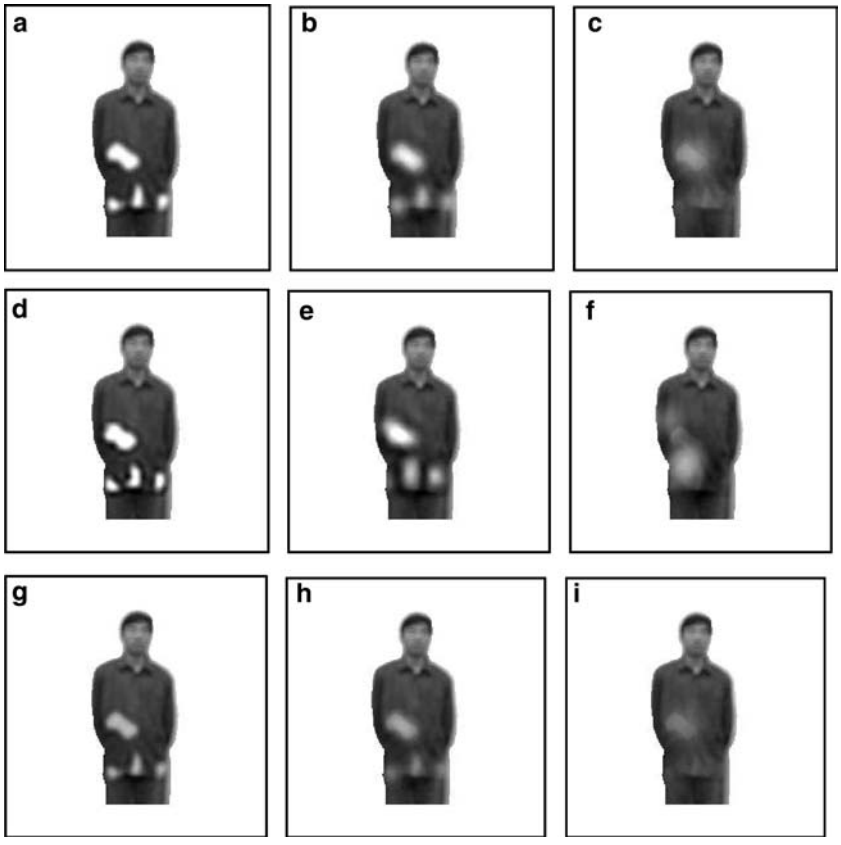


Fig. 10 Mosaic results achieved by applying the multiresolution approach one at different decomposition level **a** 2, **b** 3, and **c** 4; approach two at decomposition level **d** 2, **e** 3, and **f** 4; approach three at decomposition level **g** 2, **h** 3, and **i** 4

Table 2 Comparison of the fuzzy *k*-means clustering results with different initial cluster numbers

Cluster number	8	10	13	16	19	22	25	30	40
False positive	0.5146	0.4314	0.3718	0.3444	0.3014	0.3014	0.2819	0.2552	0.2552
True positive	1	1	0.9721	0.9604	0.9249	0.9249	0.9106	0.901	0.901

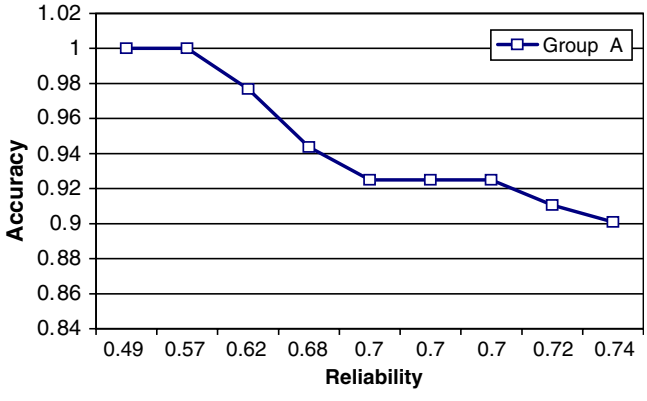


Fig. 11 The effect of cluster number for IR image of group A in Fig. 4

may involve two or more long-wavelength sensors for a decision-level fusion. Following the procedure in Fig. 2b, we investigated the detection of concealed weapons from the IR image and creating a composite image with visual information for an operation or avoiding privacy offense. As far as the second scenario is concerned, the idea is to detect the concealed weapon from the IR, MMW image, or their fusion result and embed the weapon region in the visual image. Since the most important information provided by IR or MMW image is the region of the concealed weapon, the other parts will not make any contribution to the specific

Table 3 Comparison of multiresolution image fusion schemes

	Fuzzy *k*-means clustering		EM clustering		*k*-Means clustering	
	Accuracy	Reliability	Accuracy	Reliability	Accuracy	Reliability
Group A (1)	1	0.4854	1	0.2344	1	0.3917
Group B (2)	1	0.4569	1	0.3546	1	0.3828
Group C (3)	0.9529	0.4868	0.9532	0.4455	0.9540	0.3329
Group D (4)	0.4336	0.4117	0.4946	0.4077	0.5373	0.4090
Group E (5)	0.8618	0.6695	0.8618	0.5431	0.8618	0.6217
Group F (6)	0.9254	0.5066	0.9254	0.3414	0.9254	0.3290
Group G (7)	0.9776	0.8104	0.9776	0.6150	0.9776	0.6539
Group H (8)	0.4767	0.5556	0.5100	0.5055	0.8211	0.4980
Group I (9)	0.2248	0.2412	0.3222	0.2898	0.3895	0.3117

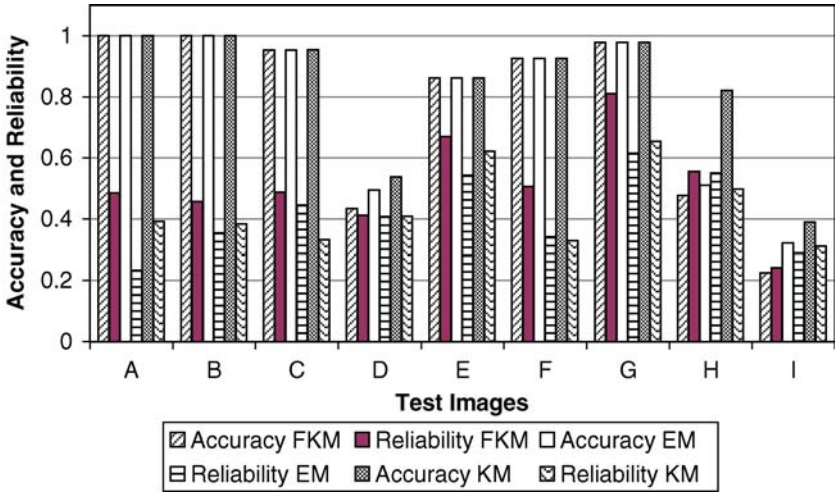


Fig. 12 The effect of cluster number for IR image of group A in Fig. 7

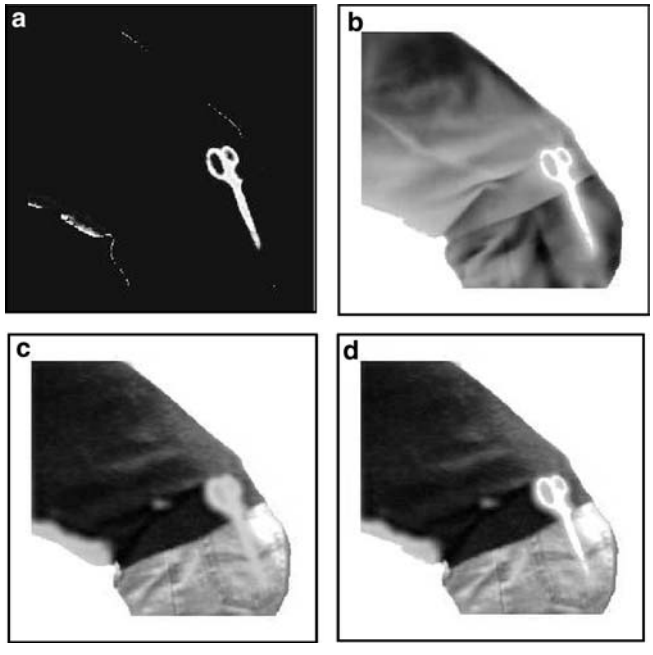


Fig. 13 Enhancement of ROI: **a** clustered result on the ROI of IR image; **b** enhanced IR image; **c** mosaic result with original IR image; and **d** mosaic result with enhanced IR image

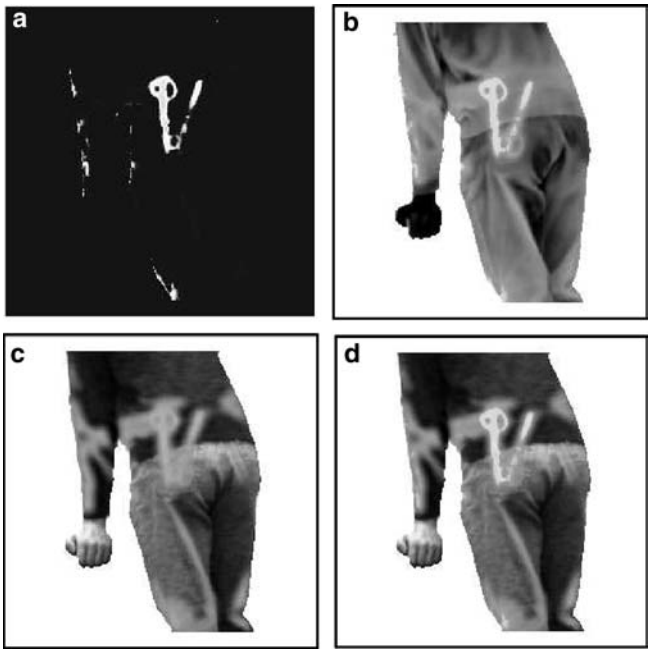


Fig. 14 Enhancement of ROI: **a** clustered result on the ROI of IR image; **b** enhanced IR image; **c** mosaic result with original IR image; and **d** mosaic result with enhanced IR image

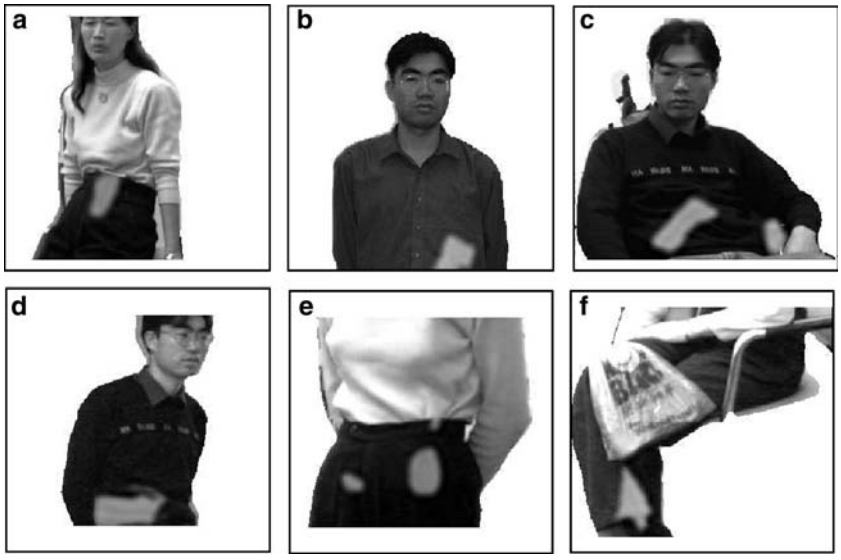


Fig. 15 Experimental results achieved by applying the third multiresolution mosaic scheme

analysis. The critical issue is the detection of weapon from IR images. If the weapon cannot be identified, it does not make any sense to fuse it with the visual image.

From the above experiments, we find that the multiresolution-based fusion approaches do not always generate a good result. This is due to the variations in image formation and intensity map. Furthermore, the fusion operation degraded the quality of the results due to the integration of useless information. The face is hard to identify in the fused image although the concealed weapon region is highlighted to some extent. Quantitative evaluation of image fusion results is performed by comparing with a reference. The metrics for comparison of two images like root mean square error, correlation, and signal to noise ratio are employed in [9], but these values do not assure the fidelity of the fused image. The quality of the fused image can be tested by further processing, such as face recognition or weapon template matching, if applicable. A better fusion result should facilitate the further processing. With the mosaic technique, one hundred percent of the visual image's quality can be preserved. The objective assessment of the results is accomplished by using the accuracy and reliability measurements once the threshold value is selected.

The advantage of using fuzzy k-means algorithm is that the clustered pixels are accompanied with a membership value ranging from 0 to 1, which provides additional information, i.e., to what extent we can trust the results. As shown in the experiment, the membership map can also be used for enhancement of the detected ROI region. The clustering does introduce the false alarm due to the "noise" in the IR image, which may come from the background. The detection of foreground object is not a difficult problem to solve. One solution is to use the technique for background subtraction as described in [48]. A camera calibration procedure is given in [49]. Thus, the processing can be focused on the derived target object. An example is shown in Fig. 15c, where the region under that person's left arm should be the backrest of the chair. However, this region is also detected and embedded in the corresponding visual image (see Fig. 15c), although this does not affect the subjective evaluation. Besides, the IR imager distinguishes the weapon from the other part of the body based on the temperature distribution. The bottom edge of the pants and the shirt or some other parts may have the same temperature as the weapon. The clustering algorithm may cluster those regions too as concealed weapon. This does happen to most of the images in Fig. 7. The IR imager has its limitation and does not assure a hundred percent detection. Therefore, in order to improve the probability of detection, other image sensors like millimetre wave imager or ultrasound imager can be employed to decrease the uncertainty with more complementary information. The study on probability of detection (POD) should be carried out and higher level fusion can be considered.

So far as the MRIM is concerned, one observation is that the process with a larger decomposition level degrades the mosaic results. This is not always true and largely depends on the size of the region (image) to be embedded. When the region is relatively small, as the test images in this report, at a lower resolution the image components will be blurred by the weighted summation with the Gaussian components of the binary mask image. This also happens to the edge-based mask image. Nevertheless, a lower-level decomposition is good for improving the computational efficiency.

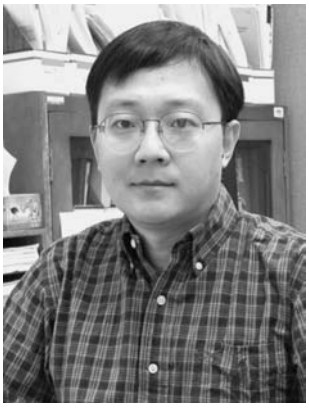
6 Conclusion

In this paper, a scheme based on multiresolution mosaic for concealed weapon detection is presented. The technique will enhance the portal detection for potential

threats at the airport or other sensitive locations. The procedure includes two steps: (1) weapon region detection from the IR image and (2) the ROI (detected weapon) mosaic on the visual image. This strategy clarifies the task for each stage, i.e., what to detect and how to combine the results. The multiresolution mosaic technique provides a way to combine two images seamlessly. In the synthesized image, the fidelity of the visual image is preserved well while the concealed weapon is highlighted. An enhancement of the ROI will further facilitate the process. The disadvantage is that the detection algorithm may introduce false positive or false negative error. This is partly due to the limitation of the IR image sensor itself. To improve the probability of detection, information fusion with other image sensors like a MMW imager is the work for the future.

7 Originality and contribution

This paper proposes a two-step scheme to generate a synthesized image from a visual and an IR image. The personal identification, i.e., facial pattern is retained with comparable fidelity to the visual image while the region of concealed weapon is enhanced and highlighted. In the detection process, an unsupervised clustering method, namely fuzzy k-means clustering, is employed to segment the suspicious objects from the IR image. The detected region is then embedded in the corresponding visual image with a multiresolution mosaic technique. This work provides an efficient solution to operator-assistant weapon detection and avoidance of privacy offense at the portal security check for sensitive locations.



Zheng Liu completed his BE degree in Mechanical and Automation from the Beijing Institute of Chemical Fibre Technology (P.R. China) in 1991 and obtained a ME degree in Automatic Instrumentation from the Beijing University of Chemical Technology (P.R. China) in 1996. He earned a Doctorate in Engineering from Kyoto University in 2000 (Japan). His research interests include data fusion, computer vision, and pattern recognition. He is a member of IEEE and CIPPRS.

Zhiyun Xue received the BSc degree and MSc degree in Electrical Engineering from Tsinghua University, Beijing, China, in 1996 and 1998, respectively. She worked as a research assistant in the Biomedical Engineering Research Center in Nanyang Technological University, Singapore, and received the ME degree from there in 2000. She is currently a PhD candidate in Lehigh

Rick S. Blum received a BS in Electrical Engineering from the Pennsylvania State University in 1984 and his MS and PhD in Electrical Engineering from the University of Pennsylvania in 1987 and 1991. From 1984 to 1991, he was a member of technical staff at General Electric Aerospace in Valley Forge, Pennsylvania and he graduated from GE's Advanced Course in Engineering. Since 1991, he has been with the Electrical and Computer Engineering Department at Lehigh University in Bethlehem, Pennsylvania where he is currently a Professor and holds the Robert W. Wieseman Chair in Electrical Engineering. His research interests include signal detection and estimation and related topics in the areas of signal processing and communications. He is currently an associate editor for the IEEE Communications Letters and he is on the editorial board for the Journal of Advances in Information Fusion of the International Society of Information Fusion. He was an associate editor for the IEEE Transactions on Signal Processing and edited special issue for this journal. He was a member of the Signal Processing for Communications Technical Committee of the IEEE Signal Processing Society.Dr. Blum is a Fellow of the IEEE, an IEEE Third Millenium Medal winner, a member of Eta Kappa Nu and Sigma Xi, and holds several patents. He was awarded an ONR Young Investigator Award in 1997 and an NSF Research Initiation Award in 1992.

Robert Laganière is an associate professor at the School of Information Technology and Engineering of the University of Ottawa. He received a PhD degree from INRS-Telecommunications in Montreal in 1996. His research interests are computer vision and image processing with applications to augmented reality, visual surveillance, 3D reconstruction and image-based rendering. He is also the co-author of a book on Object-oriented software development, published by McGraw Hill Editors.

Acknowledgements Mr. D. S. Forsyth is acknowledged for his valuable comments and discussions.

References

1. Klock BA (2003) Interface and usability assessment of imaging systems. IEEE AESS Syst Mag 18(3):11–12
2. McMillan RW, O Milton J, Hetzler MC, Hyde RS, Owens WR (2000) Detection of concealed weapons using far-infrared bolometer arrays. In: Conference digest on 25th infrared and millimeter waves, pp 259–260
3. Slamani MA, Ramac L, Uner M, Varshney P, Weiner DD, Alford M, Derris D, Vannicola V (1997) Enhancement and fusion of data for concealed weapons detection. In: SPIE, vol 3068, pp 20–25
4. Burt PJ, Kolczynski RJ (1993) Enhanced image capture through fusion. In: Proceedings of 4th international conference on image processing, pp 248–251
5. Uner MK, Ramac LC, Varshney PK, Alford M (1996) Concealed weapon detection: an image fusion approach. In: SPIE, vol 2942, pp 123–132
6. Varshney PK, Ramac L, Slamani MA, Alford MG, Ferris D (1998) Fusion and partitioning of data for the detection of concealed weapons. In: Proceedings of the international conference on multisource-multisensor information fusion
7. Varshney PK, Chen H, Uner M (1999) Registration and fusion of infrared and millimetre wave images for concealed weapon detection. In: Proceedings of international conference on image processing, vol 13, pp 532–536
8. Aggarwal JK (1993) Multisensor fusion for computer vision, vol 99 of NATO ASI series F: computer and systems science
9. Xue Z, Blum R, Li Y (2002) Fusion of visual and ir images for concealed weapon detection. In: Proceedings of ISIF 2002, pp 1198–1205
10. Foresti GL, Snidaro L (2002) A distributed sensor network for video surveillance of outdoors. In: Foresti GL, Regazzoni CS, Varshney PK (eds) Multisensor surveillance systems. Kluwer, Dordrecht, pp 7–27
11. Matsopoulos GK, Marshall S, Brunt JNH (1994) Multiresolution morphological fusion of mr and ct images of the human brain. IEE Proc Vis Image Signal Process 141(3):137–142
12. Koren I, Laine A, Taylor F (1998) Enhancement via fusion of mammographic features. In: Proceedings of international conference on image processing, pp 722–726
13. Pohl C, Genderen JLV (1998) Multi-sensor image fusion in remote sensing: concepts, methods and applications. Int J Remote Sens 19(5):823–854
14. Gros XE, Liu Z, Tsukada K, Hanasaki K (2000) Experimenting with pixel-level ndt data fusion techniques. IEEE Trans Instrum Measure 49(5):1083–1090
15. Chen HM, Lee S, Rao RM, Slamani MA, Varshney PK (2005) Imaging for concealed weapon detection. IEEE Signal Process Mag 22(2):52–61
16. Xue Z, Blum RS (2003) Concealed weapon detection using color image fusion. In: Proceedings of 6th international conference of information fusion, vol 1, pp 622–627
17. Loftus P (2005) Camera detects concealed weapons. Wall Street J (online)
18. Adelson EH, Anderson CH, Bergen JR, Burt PJ, Ogden JM (1984) Pyramid methods in image processing. RCA Eng 29(6):33–41
19. Chipman LJ, Orr TM (1995) Wavelet and image fusion. In: Proceedings of international conference on image processing, pp 248–251
20. Li H, Manjunath BS, Mitra SK (1995) Multisensor image fusion using the wavelet transform. Graph Models Image Process 57(3):235–245
21. Zhang Z (1999) Investigations of image fusion. PhD Thesis, Lehigh University
22. Piella G (2003) A general framework for multiresolution image fusion: from pixels to regions. Inf Fusion 4(4):259–280
23. Wilson TA, Rogers SK, Myers LR (1995) Perceptual-based hyperspectral image fusion using multiresolution analysis. Opt Eng 34(11):3154–3164
24. Wilson TA, Rogers SK, Kabrisky M (1997) Perceptual-based image fusion for hyperspectral data. IEEE Trans Geosci Remote Sens 35(4):1007–1017
25. Zhang Z, Blum RS (1998) Image fusion for a digital camera application. In: Proceedings of 32nd Asilomar conference on signals systems, and computers, Monterey, pp 603–607
26. Koren I, Laine A, Taylor F (1995) Image fusion using steerable dyadic wavelet transform. In: Proceedings of international conference on image processing, pp 232–235
27. Petrovic V, Xydeas C (1999) Multiresolution image fusion using cross band feature selection. In: SPIE, vol 3719, pp 319–326
28. Teot A (1989) Image fusion by a ratio of low-pass pyramid. Pattern Recognit Lett 9:245–253
29. Toet A (1992) Multiscale contrast enhancement with application to image fusion. Opt Eng 31(5):1026–1031
30. Liu Z, Tsukada K, Hanasaki K, Ho YK, Dai YP (2001) Image fusion by using steerable pyramid. Pattern Recognit Lett 22:929–939
31. Rockinger O (1996) Pixel level fusion of image sequences using wavelet frames. In: Proceedings of the 16th leeds annual statistical research workshop. Leeds University Press, pp 149–154
32. Rockinger O (1997) Image sequence fusion using a shift-invariant wavelet transform. In: Proceedings of international conference on image processing, vol 3, pp 288–301
33. Rockinger O, Fechner T (1998) Pixel-level image fusion: the case of image sequences. In: SPIE, vol 3374, pp 378–388
34. Pu T, Ni GQ (2000) Contrast-based image fusion using discrete wavelet transform. Opt Eng 39(8):2075–2082
35. Nikolov S, Hill P, Bull D, Canagarajah N (2001) Wavelets for image fusion. In: Petrosian A, Meyer F (eds) Wavelets in signal and image analysis, computational imaging and vision series. Kluwer, Dordrecht, pp 213–244
36. Wang H, Peng J, Wu W (2002) Fusion algorithm for multisensor images based on discrete multiwavelet transform. IEE Proc Vis Image Signal Process 149(5):283–289
37. Slamani MA, Varshney PK, Rao RM, Alford MG, Ferris D (1999) Image processing tools for the enhancement of concealed weapon detection. In: Proceedings of ICIP, vol 3, Kobe, pp 518–522
38. Yang J, Blum RS (2002) A statistical signal processing approach to image fusion for concealed weapon detection. In: Proceedings of ICIP, vol 1, pp 513–516

39. Otsu N (1979) A threshold selection method from gray level. IEEE Trans Syst Man Cybern 9:62–66
40. Duda RO, Hart PE, Strok D (2000) Patten classification, 2nd edn. Wiley Interscience, New York
41. Balasko B, Abonyi J, Feil B Fuzzy clustering and data analysis toolbox. Department of Process Engineering, University of Veszprem, Veszprem
42. Bensaid AM, Hall LO, Bezdek JC, Clarke LP, Silbiger ML, Arrington JA, Murtagh RF (1996) Validity-guided (re)clusting with applications to image segmentation. IEEE Trans Fuzzy Syst 4:112–123
43. Xie XL, Beni GA (1991) Validity measure for fuzzy clustering. IEEE Trans Pattern Anal Mach Intell 13(8):841–847
44. Maulik U, Bandyopadhyay S (2002) Performance evaluation of some clustering algorithms and validity indices. IEEE Trans Pattern Anal Mach Intell 24(12):1650–1654
45. Siomoncelli E, Freeman W (1995) The steerable pyramid: a flexible architecture for multi-scale derivative computation. In: Proceedings of 2nd IEEE international conference on image processing. Washington DC, pp 444–447
46. Siomoncelli EP, Freeman WT, Adelson EH, Heege D (1992) Shiftable multiscale transform. IEEE Trans Inf Theory 38(2):587–607
47. Hsu CT, Wu JL (1996) Multiresolution mosaic. IEEE Trans Consumer Electron 42(4):981–990
48. Toyama K, Krumm J, Brumitt B, Meyers B (1999) Wallflower: principles and practice of background maintenance. In: Proceedings of international conference on computer vision, pp 255–261
49. Yasuda K, Naemura T, Harashima H (2004) Thermo-key human region segmentation from video. Comput Graph Appl 24(1):26–30