AFRL-IF-RS-TR-2001-92
Final Technical Report
May 2001

SENSOR FUSION ALGORITHMS AND PERFORMANCE LIMITS

Syracuse University

Pramod K. Varshney, Mucahit K. Uner, Liane C. Ramac and Hua-Mei Chen

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

AIR FORCE RESEARCH LABORATORY
INFORMATION DIRECTORATE
ROME RESEARCH SITE
ROME, NEW YORK

20010706 110

This report has been reviewed by the Air Force Research Laboratory, Information Directorate, Public Affairs Office (IFOIPA) and is releasable to the National Technical Information Service (NTIS). At NTIS it will be releasable to the general public, including foreign nations.

AFRL-IF-RS-TR-2001-92 has been reviewed and is approved for publication.

Mark G. Alford

APPROVED: MARK G. ALFORD
Project Engineer

FOR THE DIRECTOR: JOSEPH CAMERA, Chief
Information & Exploitation Division
Information Directorate

If your address has changed or if you wish to be removed from the Air Force Research Laboratory Rome Research Site mailing list, or if the addressee is no longer employed by your organization, please notify AFRL/IFEA, 32 Brooks Road, Rome, NY 13441-4114. This will assist us in maintaining a current mailing list.

Do not return copies of this report unless contractual obligations or notices on a specific document require that it be returned.

REPORT DOCUMENTATION PAGE	Form Approved OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	May 01	Final Jul 95 - Jul 98

4. TITLE AND SUBTITLE

SENSOR FUSION ALGORITHMS AND PERFORMANCE LIMITS

5. FUNDING NUMBERS

C - F30602-95-1-0027
PE - 62702F
PR - 4600
TA - AA
WU - XP

6. AUTHOR(S)

Pramod K. Varshney, Mucahit K. Uner, Liane C. Ramac, and Hua-Mei Chen

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Syracuse University
Electrical Engineering and Computer Science Department
121 Link Hall
Syracuse NY 13244

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

AFRL/IFEA
32 Brooks Rd
Rome, NY 13441-4114

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

AFRL-IF-RS-TR-2001-92

11. SUPPLEMENTARY NOTES

AFRL Project Engineer: Mark G. Alford, IFEA, 315-330-3802

12a. DISTRIBUTION AVAILABILITY STATEMENT

Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

The concealed weapons detection program involves auto-detection and recognition of weapons under clothing. Single sensor methods are insufficient for detection in many situations. This program develops methodologies for detecting weapons under clothing using multiple dissimilar imaging sensors. Several signal processing steps are considered starting with image acquisition and concluding with a binary or tertiary decision about the presence or absence of a weapon. Image registration, noise removal filtering, image fusion and shape recognition are all discussed. Individual attributes are used to register images which are then fused in the wavelet domain.

14. SUBJECT TERMS

law enforcement, remote transmission, wireless image transmission, wavelet transform, saliency measures, morphological filtering

15. NUMBER OF PAGES

104

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

Standard Form 298 (Rev. 2-89) (EG)
Prescribed by ANSI Std. 239.18
Designed using Perform Pro, WHS/DIOR, Oct 94

ABSTRACT

The Concealed Weapons Detection (CWD) problem involves the automatic detection and recognition of weapons hidden underneath a person's clothing. One type of sensor by itself may not be able to detect a concealed weapon in different situations. Therefore, use of different imaging sensors for this task may provide more information than using a single imaging sensor. CWD includes several signal processing steps beginning with the initial step of image acquisition and concluding with the decision of whether or not a weapon is present. In this report, we discuss image registration, filtering for noise removal, image fusion, and shape recognition. The image registration method presented here uses attributes of the individual source images to find corresponding areas in the images. The image fusion technique described here is based on the wavelet transform where the most salient features from the source images are placed in the fused image. For law enforcement applications that may require transmission of images from remote locations, image fusion is utilized for wireless image transmission.

Keywords: Concealed weapons detection, image registration, morphological filtering, image fusion, shape recognition

TABLE OF CONTENTS

1. INTRODUCTION

The Concealed Weapons Detection (CWD) problem involves the automatic detection and recognition of weapons hidden underneath a person's clothing. One type of sensor by itself may not be able to detect a concealed weapon in different situations. Use of different imaging sensors for this task may provide more information than using a single imaging sensor. This is because dissimilar sensors can provide different and possibly complementary information. One example of dissimilar sensors is IR and MMW cameras. We have worked with images from both of these sensors where the IR sensors tend to have better spatial resolution but do not penetrate clothing well and the MMW sensors have poor spatial resolution but do a much better job at penetrating clothing. When the individual image information is fused into a composite image, this image will contain more information and increase the probability of recognition and correct interpretation.

CWD includes several signal processing steps beginning with the initial step of image acquisition and concluding with the decision of whether or not a weapon is present. Figure 1 shows a block diagram of the entire CWD process. The source images can be from multiple sensors of the same type but placed at different viewing angles, or the source images may be from different sensor types that provide complementary scene information. The first step after the acquisition of images from multiple sensors may involve several types of pre-processing. This can include tasks such as image registration, noise removal or contrast enhancement. All of these pre-processing tasks can help improve detection performance and recognition of a concealed weapon. The next step is fusion where the images from multiple sensors are combined to form one composite image. After fusion, the

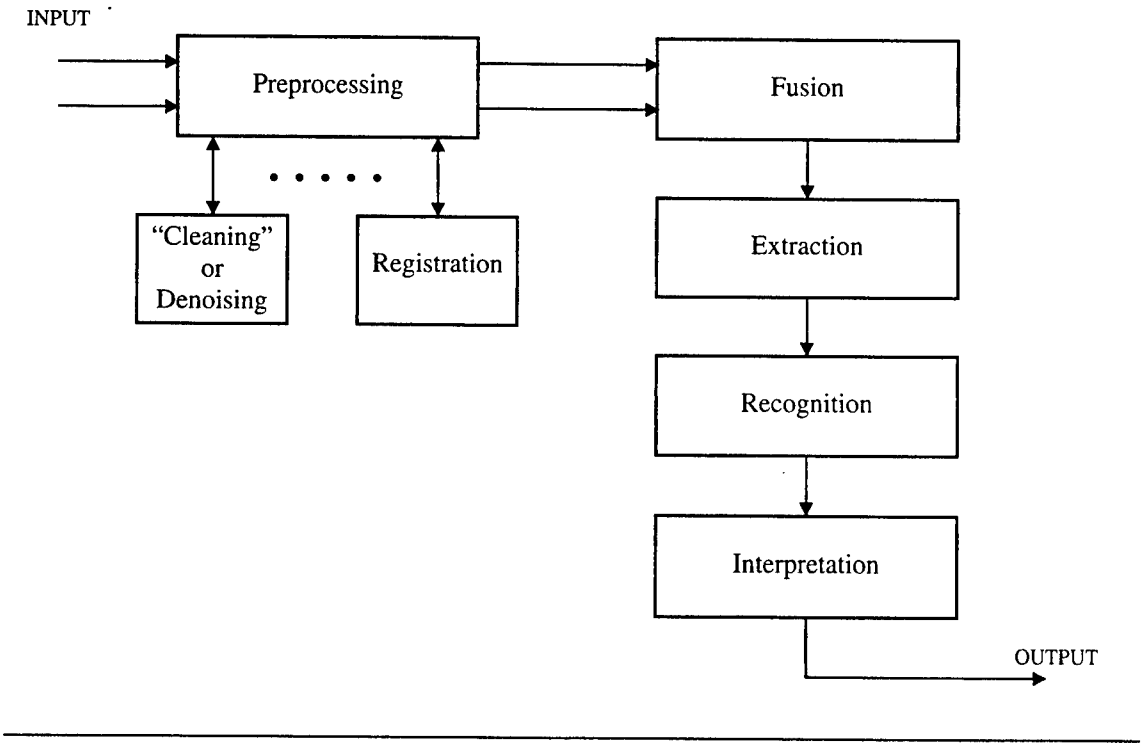


Figure 1: Processing steps for CWD

objects of interest are extracted from the composite image. Next, recognition algorithms are applied to the extracted objects and shape descriptors that characterize different object features are computed. Finally, the shape descriptors are interpreted to decide whether the object is a weapon or a non-weapon.

The preprocessing stage is important to all of the following stages of CWD. In general, imaging sensors are not located in the same physical location. Therefore, the source images need to be registered before fusion. Original sensor images may also contain unwanted details such as shadows, wrinkles, imaging artifacts, etc., which are not needed in the final fused image. These details can adversely affect the performance of the recognition stage of CWD. Therefore, before fusing the images it is desirable that most of these unwanted details are removed from the source images. This will help improve performance by removing details that may cause false recognition.

The CWD fusion stage encompasses the steps of image transformation (or representation),

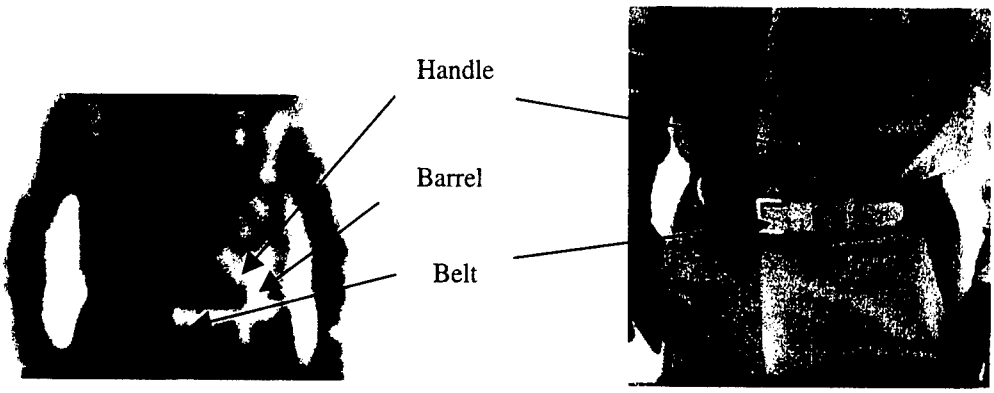
image fusion, and inverse image transformation. Image representation involves deciding what type of transform to implement and is an important issue for fusion. If the source images are fused without transforming to another domain, many of the essential details needed for recognition may not be included in the fused image. At different resolutions, the details of an image generally characterize different physical structures of the scene. For example, at coarser resolutions, the details correspond to the larger structures. As the resolution gets finer, the details correspond to the smaller structures. Depending upon the features being sought, multiresolution analysis such as the one based on the wavelet transform provides a very useful tool for image analysis. The actual fusion rules for combining the source images can utilize this resolution information to select the important details for the composite image.

Extraction (or segmentation), recognition, and interpretation are represented as different stages in the CWD block diagram but they are all related to one another. The extraction stage attempts to isolate the possible weapon objects from the rest of the image based upon the characteristics of different areas of the composite image. Several algorithms have been proposed to accomplish this but, in general, the segmentation algorithms are image dependent. The recognition stage includes defining measures that quantitatively describe the objects found in the previous stage. Typical recognition measures include Fourier descriptors and moments. The interpretation stage determines what group (weapon or non-weapon) the objects fall into based on prior knowledge.

In addition to the CWD stages discussed above, there may be a need to transmit images from remote locations in certain law enforcement situations. Portable communications units are increasing in popularity and wireless image transmission is one desired feature of these systems. However, wireless communications systems are susceptible to fading phenomenon that cause bursty channel errors. Image fusion will be discussed as a method to combat errors on these channels.

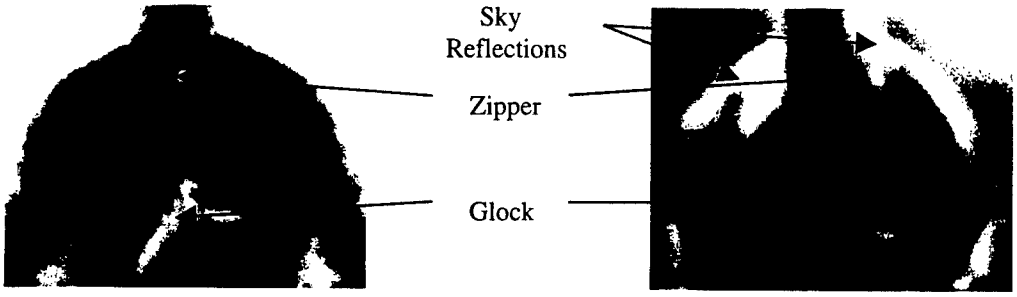
Section 2 describes the image registration algorithms for infrared (IR) and millimeter-wave (MMW) images. Section 3 describes morphological filtering as a preprocessing step to remove

unwanted image details. Section 4 discusses the details of the steps used for image fusion. The recognition and interpretation stages of CWD are described in Section 5. Section 6 shows the results of applying these algorithms to IR and MMW image pairs. Section 7 describes the use of image fusion for the transmission of images over wireless channels. Finally, Section 8 provides conclusions about this research.



(a) Indoor Images Beretta

(b) Outdoor IR and MMW



(c) Indoor/Outdoor comparison

Figure 2.1. Three IR and MMW image pairs that demonstrate the features of IR and MMW

2. IMAGE REGISTRATION

The IR imaging system used to detect concealed weapons was a quantum well longwave camera, which was built by Martin Marietta Labs. The MMW imager consists of a single detector that mechanically scans the scene in a raster pattern. It was built by Intelligent Machine Technology of Irving, TX. Figure 2.1 shows typical IR and MMW image pairs and Table 2.1 lists some specifications of the two imaging systems [CDF96].

Table 2.1: Some specifications of the IR and MMW imaging systems used to collect data

	IR	*MMW*
Operation band	8 ~ 12 μm	3.2 mm
Image size (in pixels)	256 × 256	192 × 254
Field of view	8.6 degree × 8.6 degree	18 degree × 24 degree
Lens	f/1.7	30 cm diameter aperture with f/1
Minimum resolvable temperature	0.007 Kelvin	0.2 Kelvin
Image time	Not available[1]	3 min ~ 90 min [2]

1. An IR image can be taken instantly.
2. This range corresponds to the resolvable temperature range 1.25 to 0.2 degrees K.

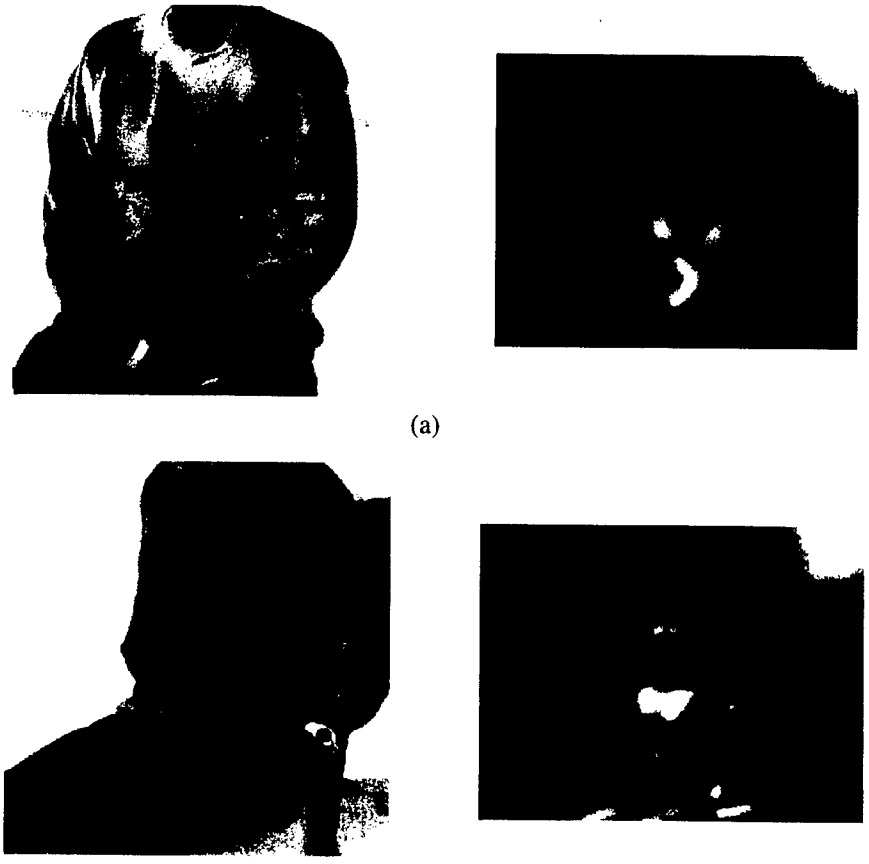


Figure 2.2: Two typical outdoor IR and MMW image pairs.

2.1 Characteristics of IR and MMW images

Three image pairs that demonstrate the features of both types of imagers are shown in Figure 2.1. Figure 2.1 (a) shows an image pair taken from IR and MMW imagers indoors. Figure 2.1 (b) shows an IR and MMW image pair taken outdoors. Figure 2.1 (c) shows a MMW image taken indoors and an equivalent outdoor MMW image. From Figure 2.1 (a) we can see the superior angular resolution of the IR image and the better penetration of the MMW imager. The Beretta barrel, which can not be seen in the IR image, can be seen clearly in the MMW image. In Figure 2.1 (b), similar characteristics of IR and MMW imagers can be seen, but the MMW image in Figure 2.1 (b) looks quite different from the one in Figure 2.1 (a) because of sky reflections. A further comparison of

indoor and outdoor MMW images is shown in Figure 2.1 (c)

From Figure 2.1 (c) we can find that the Glock is clearly detected in the indoor image but is not shown clearly in the outdoor image. The reason for this is possibly the fact that the cold sky reflections associated with the outdoor image have obscured the Glock. Nevertheless, the sky reflection can not be detected by IR imagers as shown in Figure 2.1 (a). In the following paragraphs we try to give a theoretical explanation of these phenomena.

Both the IR and MMW images are passive since they do not require any form of illumination to perform the required imaging. Both sensors detect emitted energy based on Planck's Law, which is stated below:

$$W_\lambda = 2\pi c^2 h \lambda^{-5} (e^{hc/\lambda kT} - 1)^{-1} \qquad (2.1)$$

where W_λ is the black body spectral energy density, c is the speed of light, λ is the wavelength, h is Planck constant, k is Boltzmann constant, and T is the absolute temperature of the object being imaged. For a given spectral band Equation (2.1) implies that the spectral energy density is a function of temperature only. That is, warmer objects will emit greater energy/unit time than cooler objects. In our case, the ratio of the emitted energy in the operation bands of IR and MMW imagers is about 10^8. Because of this, it is hypothesized that the MMW imagers "sees" an image that is a composite of reflected and emitted spectral energy depending on the emissivities and reflectivities of the objects within the image, while the IR imager "sees" only the thermal emissions and does not detect reflections. Therefore, the location of the person being observed may greatly affect the ability to detect the concealed weapon. More specifically, there is minimal cold sky reflection when a picture is taken indoors, while the sky reflection can dominate the MMW image if it is taken outdoors.

Another important issue is the differences in angular resolution between MMW and IR imagers. If the diffraction limited case is considered for both the IR and MMW imagers, we find that the IR imager has far superior angular resolution. This is not unexpected since the equation for

diffraction limited angular resolution is:

$$\Theta_R = \lambda / D \tag{2.2}$$

where λ is the wavelength and D is the aperture diameter. This equation implies that the IR imager's superior angular resolution (for comparable aperture diameters) results from the infrared's much shorter wavelength.

Thus, it is the difference of the operation band that results in the superiority of IR imagers in terms of angular resolution and the superiority of MMW imagers in their ability to penetrate heavy clothing. These two imagers provide complementary information that can be fused to improve CWD performance. Based on available IR and MMW images, we observed the following three characteristics:

1. Body portion is darker than the background in IR images.
2. Body portion is either darker or brighter than the background except for the transition part in MMW images and the boundary of body is preserved quite well.
3. The background is smoother than the body portion in MMW images.

The first two properties can be easily seen from Figure 2.1. In order to see the third property clearly, we convolve each MMW image with [1, -1] and then take the absolute value. The resulting images are shown in Figure 2.3. The gray level of each pixel in the resulting images represents the difference of the gray levels of the pixel and the pixel to its right in the original MMW images. Therefore, we can easily see that the background is smoother than the body portion in MMW images because the background is darker than the body portion. All of the three properties are used while developing our registration algorithm.

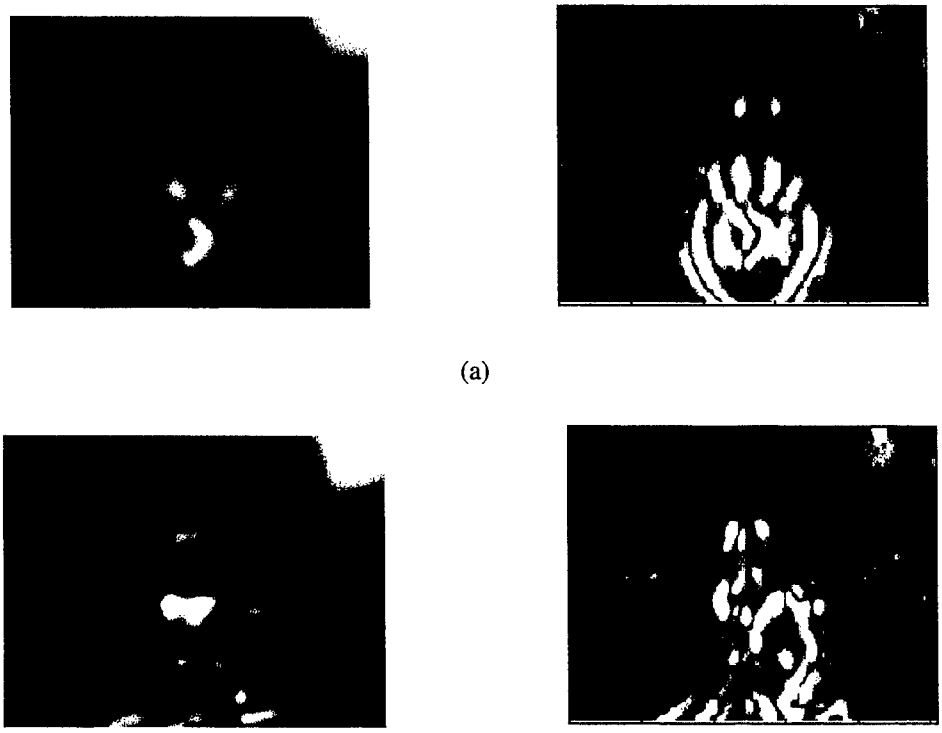


Figure 2.3: Smoothness of background in MMW images

2.2 Image Registration Review

Image registration is a procedure that determines the best fit between the objects in two or more images. The main issue of image registration is to establish the correspondence between images of the same scene. A broad range of image registration techniques have been developed for a wide variety of imaging problems. Remote sensing [TBC86], biomedical imaging [SFF], and computer vision [KJ91, Hor89] are typical application areas. According to the chosen feature space for performing image registration, the methods can be divided into two categories: area-based and feature-based methods. Excellent surveys of these techniques can be found in [Bro92, MF93].

2.2.1 Area-based registration methods

The area-based registration methods are also called 'block matching' methods. This method attempts to match a reference image (or block) with a region in the test image. We review some existing area-based registration techniques that utilize different similarity metrics.

Correlation method [Bro92]

Cross-correlation is often used for template matching or pattern recognition. For a template T and an image I, where T is small compared to I, the two-dimensional normalized cross-correlation function measures the similarity at each translation, (u, v), and is given by:

$$C(u,v) = \frac{\sum_x \sum_y T(x,y) I(x-u, y-v)}{\sqrt{\sum_x \sum_y I^2(x-u, y-v)}}. \qquad (2.3)$$

If the template matches the image at translation *(u, v)*, except for an intensity scale factor, the cross-correlation will have its peak at *C (u, v)*. Thus, by computing C over all possible translations, it is possible to find the region in the test image that is most similar to the template. Note that the cross-correlation must be normalized since local image intensity would otherwise influence the measure. If the transformation between the two images includes rotation and scaling, the cross-correlation between the image and the template is computed for each allowable transformation of the template. The transformation whose cross-correlation is the largest indicates the location of optimal registration. However, the search space needs to be kept small. Otherwise, the computational cost quickly becomes unmanageable.

Sequential similarity method

An algorithm more efficient than the traditional cross-correlation method is the sequential similarity detection algorithm (SSDA) proposed by Barnea and Silverman [BS72]. They suggested a similarity measure that is based on the absolute differences between the pixels in the two images:

$$E(u, v)=\sum_x \sum_y |T(x, y)-I(x-u, y-v)|. \quad (2.4)$$

In addition to improving computational efficiency, Barnea and Silverman introduced a sequential search strategy. In the simplest case of translation registration, for each window of the image the absolute differences between pixel pairs are accumulated and summed until a pre-determined threshold is exceeded. The number of points examined before the threshold was exceeded is recorded for each window. The window that examined the most points is assumed to have the lowest measure and the best registration.

Fourier methods

The Fourier methods search for a match using information in the frequency domain instead of the space domain. A basic technique utilizing the Fourier transform that can efficiently find the translation between two images is phase correlation [KH75]. Let f_1 and f_2 be the two images that differ only by a displacement (x_0, y_0) i.e.,

$$f_2(x, y) = f_1(x-x_0, y-y_0) \quad (2.5)$$

Their corresponding Fourier transforms F_1 and F_2 are related as follows:

$$F_2(\xi,\eta) = e^{-j2\pi(\xi x_0+\eta y_0)} * F_1(\xi,\eta) \quad (2.6)$$

The cross–power spectrum of the two images f_1 and f_2 with Fourier transforms F_1 and F_2 is defined as

$$\frac{F_1(\xi, \eta)F_2^*(\xi, \eta)}{|F_1(\xi, \eta)F_2(\xi, \eta)|} = e^{j2\pi(\xi x_0+ \eta y_0)} \quad (2.7)$$

where F^* is the complex conjugate of F. The shift theorem guarantees that the phase of the cross-power spectrum is equivalent to the phase difference between the images. By taking the inverse Fourier transform of the representation in the frequency domain, we will have an impulse function whose location indicates the displacement needed to optimally register the two images.

Now if f_1 is the translated, rotated, and scaled replica of f_2 [RSC96], that is:

$$f_2(x,y) = f_1(ax\cos\theta_0 + ay\sin\theta_0 - x_0,\ -ax\sin\theta_0 + ay\cos\theta_0 - y_0) \tag{2.8}$$

then, their Fourier magnitude spectra in polar representation are related by

$$M_1(\rho,\ \theta) = \frac{1}{a} M_2(\rho/a,\ \theta-\theta_0) \tag{2.9}$$

where (x_0, y_0) is the displacement, θ_0 is the rotated angle, and a is the scale.

Furthermore, by converting the first axis to logarithmic scale, scaling can be reduced to a translational movement (ignoring the multiplication factor $1/a$), i.e.,

$$M_1(\log\rho,\theta) = M_2(\log\rho - \log a, \theta-\theta_0) \tag{2.10}$$

Then by using the phase correlation technique introduced previously, *log a and* θ_0 can be found. The scale a can be determined accordingly.

2.2.2 Feature-based registration methods

Feature-based methods extract common features from the images to be registered, and then they attempt to match the common points. Therefore, these methods are also called point-matching methods. The general method for point matching consists of three stages. In the first stage, feature points in the image are computed. In the second stage, feature points in the reference image, often referred to as control points, are matched with feature points in the data image. In the last stage, a spatial mapping, usually two 2D polynomial functions of a specified order (one for each coordinate) is determined using these matched feature points. Mapping of one image onto the other is performed by applying the spatial mapping and interpolation.

Feature point extraction

Corners and vertex are typical feature points. Deriche and Giraudon [DG93] identify two broad groups of corner and vertex detectors. The first group extracts corners/vertices from edges represented as chain codes, by searching for points having curvature maxima or a significant change in direction. Rutkowski and Rosenfeld [RR78] report the results of a comparison among five different methods for determining corners from a chain code representation of detected, closed edges. The second group detects corners/vertices directly from the gray-level signal by applying interest operators. Several detectors for identifying and localizing interest points have been developed and Deriche and Giraudon [DG93] review many of them. Schmid [Sch96] compares several gray-level signal-based detectors based on several criteria and shows that the corner detector of Harris and Stephens [HS88] gives the best results.

In addition to corners and vertices, line intersections [SKB82], centers of gravity for closed-boundary regions [Gos86], and centers of windows having locally maximum variances [Mor81] also have been used as feature points.

Point matching

The problem of point matching uses two sets of points extracted from the two images to be registered. The first set, P, from one image contains m points. The second set, Q, from another image contains n points. Set Q is similar to set P, except that some points in P are missing and some new points, not in P, are present. The positions of the remaining points in Q are the points common with set P (within a given tolerance). The multi-point matching task is to eliminate all the points in Q (or P) which do not have a match in P (or Q) and then to find the correct match between the common points.

Most methods for point feature matching utilize geometric constraints based on the point positions, neighboring matches and their disparities. Translation matrix methods [KRD80], relaxation methods [RR80], clustering methods [SKB82], and center of gravity methods [Gos86]

belong to this category. Another category utilizes radiometric constraints [SM95] based on neighborhood characteristics of the feature, such as local image statistics. Combined radiometric and geometric matching are also proposed in literature [GB85, TH94, Hei92].

Determination of the spatial transformation

Once a sufficient number of corresponding points are found, the parameters of any transformation can be found by approximation or interpolation. Approximation uses transformation parameters that are typically based on statistical methods such as least square regression analysis or clustering. The transformation to be found does not match the control points exactly but finds the best approximation. The number of matched points must be sufficiently greater than the number of parameters of the transformation so that sufficient statistical information is available to make the approximation reliable. Merickel [Mer] registers successive serial sections of biological tissue for their 3D reconstruction using a linear least squares fitting of feature points.

For manual control points, there are usually fewer but more accurate matches, suggesting that interpolation may be more applicable. Interpolation finds the transformation that matches the control points of the two images exactly. There must be precisely one matched point for each independent parameter of the transformation to solve the system of equations. The resulting transformation defines how the image should be resampled. Bernstein [Ber76] uses this method to correct satellite imagery with low frequency sensor-associated distortions as well as for distortions caused by earth curvature and camera attitude and altitude deviations. Maguire et al. [MNR90] used this method for registration of medical images from different modalities.

2.3. The Registration Algorithm

As indicated earlier, our goal is to develop an automatic procedure to register IR and MMW images prior to fusion for the CWD application. These two images provide complementary information that can be fused to enhance the overall information available and to better detect

concealed weapons. It is a difficult problem because the two phenomenologies are different and they generate quite different images. In both IR and MMW imaging, images are generated due to temperature difference. However, light reflections also play an important role in forming the MMW image and make it quite different from the corresponding IR image. Although the body shape can be seen in both IR and MMW images, the two shapes are not exactly the same since the body shape in MMW images is affected by light reflections. Therefore, feature matching in terms of points is not feasible. In this section, we present the details of our new approach. We make the following assumptions while developing our registration algorithm:

1. The distances between the object and the sensors are large enough so that the object can be considered to be a planar object and its depth can be neglected.
2. The scale factor between the two images can be calculated based on distance information and sensor parameters like the field of view (FOV).
3. The two sensors are placed in such a manner that no rotation is required for registration.

The purpose of making the above assumptions is to ensure that the only pose parameters to be found are x-displacement and y-displacement. The details of our feature-based registration procedure are given in the following sections.

2.3.1 Idea behind our algorithm

We illustrate our idea by the following experiment. In Figure 2.4, (a) and (b) are two complete extracted images of certain object. (c) is boundary of the extracted image. (d) is part of the boundary. We correlate image (a) with images (b), (c) and (d) respectively and find the position of the peak in each correlation function to register images (b), (c) and (d). Figure 2.5 shows the

registered images. The result shows that the images (b)-(d) can be registered correctly. This experiment demonstrates the following:

1. It is possible to employ binarized images to compute correlation for registration.
2. It is sufficient to have one high quality silhouette and the other one could be the boundary of the silhouette.
3. The procedure works even when complete boundary is not available.

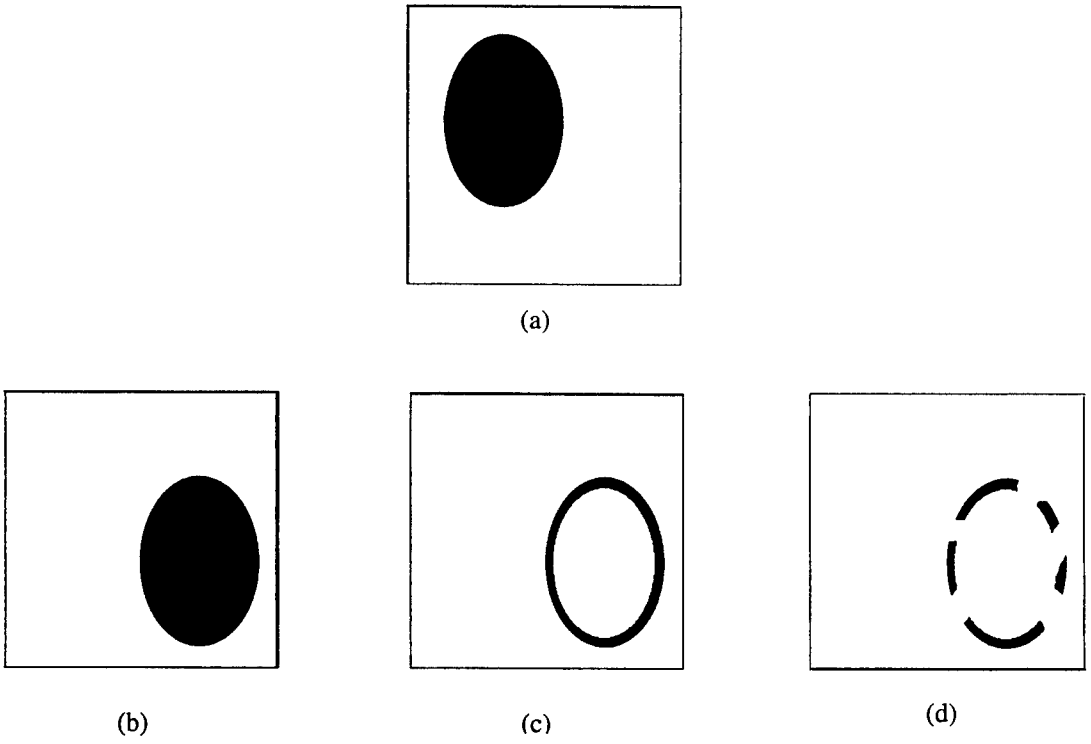


Figure 2.4 Registration of Extracted Image Boundaries

Based on this idea, we developed the algorithm to register IR and MMW images for CWD application as shown in Figure 2.6. First the scale factor is calculated based on available information regarding distances and sensor parameters. The IR image is scaled prior to body shape extraction. After body shapes have been extracted, binary correlation is used to determine x and y displacements denoted by d_x and d_y. Finally, image cropping is used to yield the two registered images. The details of the individual steps are described next.

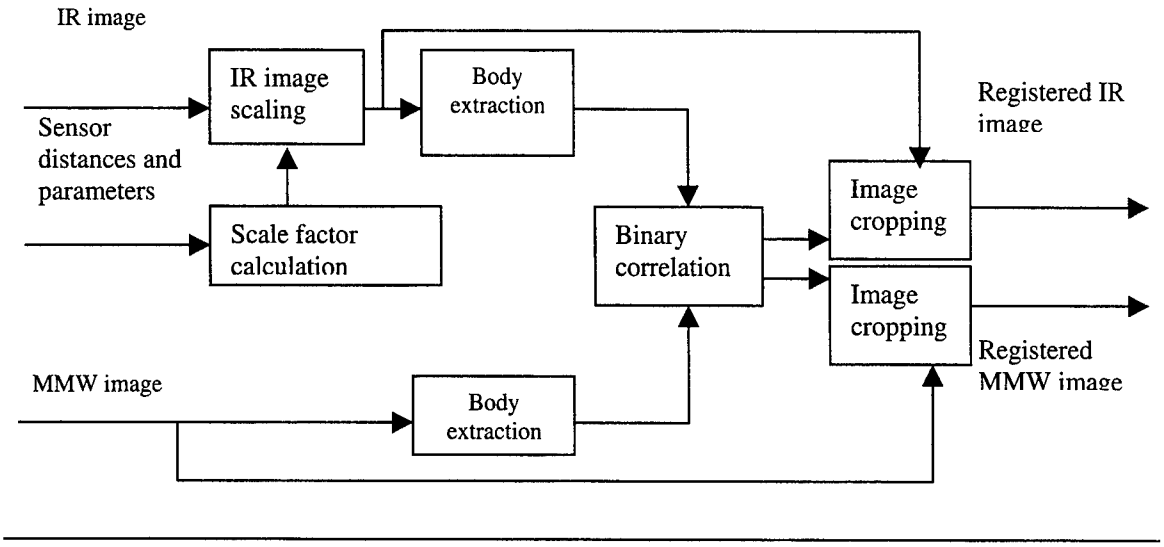


Figure.2.6. Block diagram for the registration algorithm

2.3.2 Body extraction algorithm for IR images

The goal of this step is to extract the silhouette of the human body from the given IR image. Generally speaking, the human body temperature is higher than the temperature of the background. Therefore, in IR images the body portion is darker than the background. This suggests that we can use a thresholding operation to separate the body shape from the background.

Based on the available IR images, we found that the histograms of IR images were bimodal in most cases. Therefore, in these cases, choosing the local minimum of the histogram as the threshold results in very good segmentation. In cases in which the histogram has more than two major modes, we found that one of the local minima results in very good segmentation. We determine the set of all local minima, which are all potential threshold candidates and select a threshold from this set. The criterion we use to select the threshold to extract the body shape is "shape connectivity". The concept of shape connectivity was first presented by Lie [Lie95] and defined through the co-occurrence matrix

[CM88]. It is also mentioned in [Lie95] that shape connectivity essentially represents the area/perimeter ratio of the minor class region, or the compactness of the minor class region if shape compactness [GW87] is generalized to any pattern. Here we use the area/perimeter ratio of the minor class region to calculate the shape connectivity. Based on the set of available IR images, we observed that the compactness of the minor region of the thresholded image using the desired threshold is larger than that of the thresholded images using the remaining thresholds. Therefore, we select the threshold that yields the largest value of compactness.

Due to its very nature, it is necessary to smooth the histogram before we find the set of local minima. Here we modified the method suggested by Glasbey [Gla93] to smooth histograms. Let y_0, y_1, . . ., y_{n-1}, denote the histogram of an IR image, where n is the number of gray levels used for the given IR image. We smooth the histogram by replacing y_i by $(y_{i-1} + y_i + y_{i+1}) / 3$ for i = 1,2,... n-2, with y_0 and y_{n-1} fixed. This procedure is repeated until there are M remaining modes in the histogram. In our case, M=3 was found to be adequate. Figure 2.7 summarizes the steps. An example of the algorithm is given in Figure 2.8. Note that the IR image is scaled prior to body extraction. Figures 2.8 (a) and 2.8 (b) show the scaled IR image and its corresponding histogram. The smoothed histogram with M=3 along with threshold candidates is shown in Figure 2.8 (c). Four thresholded images corresponding to the set of four potential threshold candidates are shown in Figure 2.8 (d) –

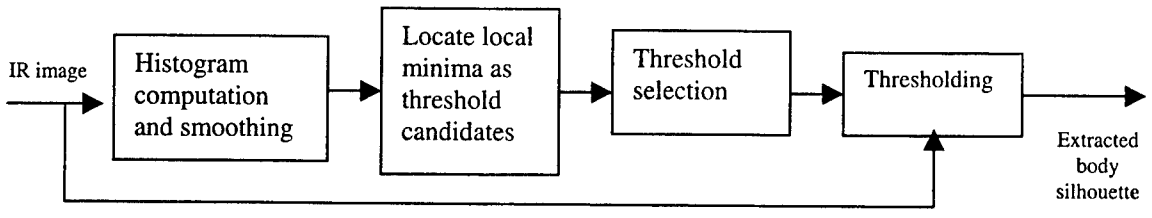


Figure 2.7. Body extraction algorithm for IR images

Figure 2.8 (g). The threshold used and the compactness of each thresholded image are also shown. Comparing the compactness measure r in Figure 2.8 (d) – Figure 2.8 (g), we choose the one shown in Figure 2.8 (f) as the desired thresholded image.

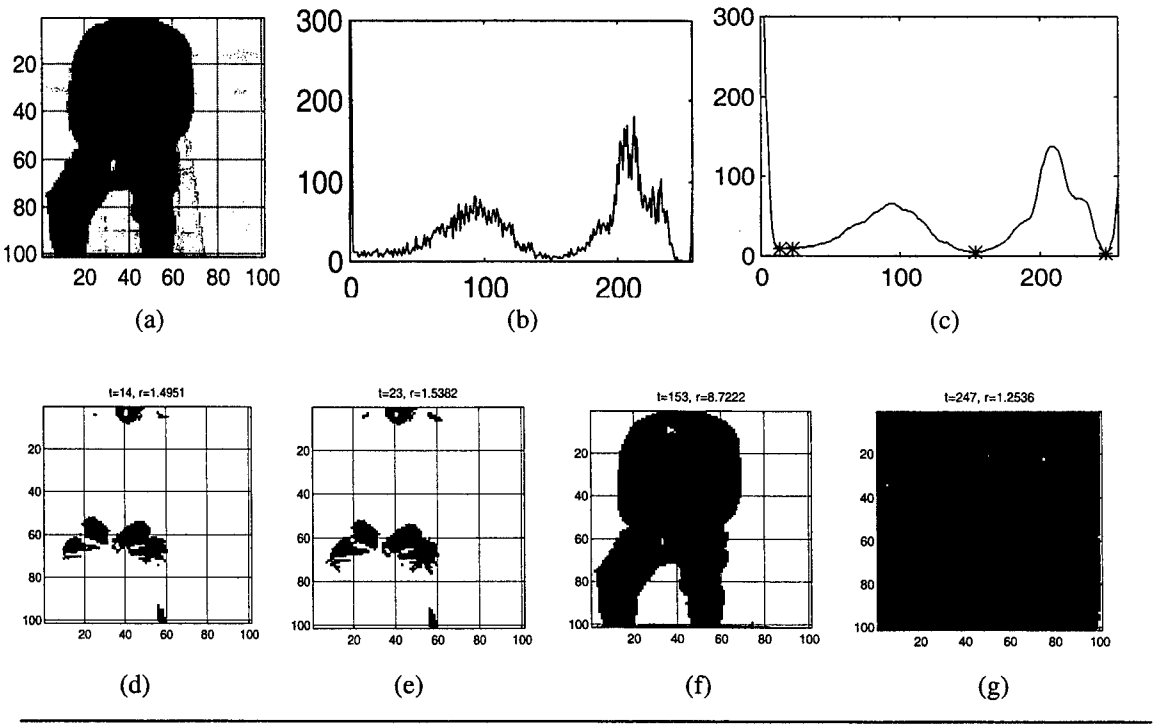


Figure 2.8: An example illustrating body extraction from IR images

2.3.3 Body extraction algorithm for MMW images

As we mentioned before, MMW images are formed due to both the temperature difference and light reflections. In the available MMW images, we note that the portion of the body without light reflections is darker than the background because its temperature is higher. But the parts of the image with light reflections are lighter. The boundary of the body is also light due to reflections. These observations suggest that we use two thresholds to extract the body shape. The resulting silhouette will be approximate due to the phenomenology involved. By examining the available

MMW images, it can be observed that there is only one major mode in the histogram and the histogram of the background falls within the main lobe of the histogram of the entire image. Therefore, if we can take out the part of the image that corresponds to the main lobe of the histogram, we can extract the object (body in our case) from the background. This requires determination of two suitable thresholds to carry out this operation. We use the same smoothing procedure as for IR images with M equal to one. After smoothing the histogram, we locate the mode and the two points of inflection. They are denoted as P, f_1, and f_2 respectively. Then we determine the two thresholds t_1 and t_2 as follows:

$$t_1=P-L_1*(P-f_1)$$

$$t_2=P+L_2*(f_2-P)$$

where L_1 and L_2 are two constants that need to be found empirically.

Once we have the two thresholds, we threshold the image with each threshold separately, invert one of the thresholded images, and then add them up to obtain the approximate silhouette of the body. Figure 2.9 summarizes the steps and an example is shown in Figure 2.10. Figures 2.10 (a) and 2.10 (b) show the original MMW image and its corresponding histogram. Smoothed histogram with M=1 along with points of inflection and thresholds is shown in Figure 2.10 (c). The two thresholded

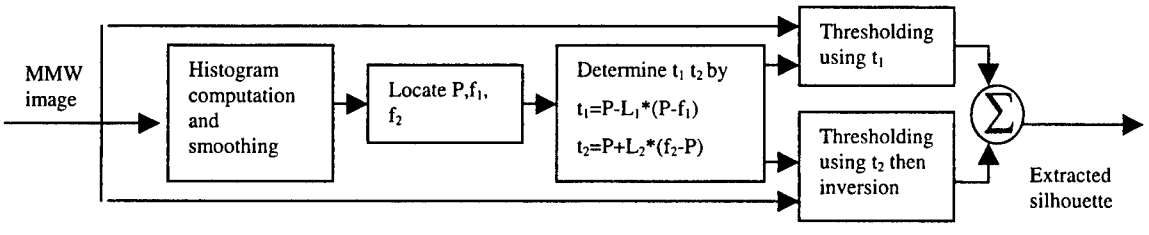


Figure 2.9: Body extraction steps for MMW images

images are shown in Figures 2.10 (d) and 2.10 (e). Figure 2.10 (f) is the composite of the two

thresholded images and represents the body shape determined by our algorithm.

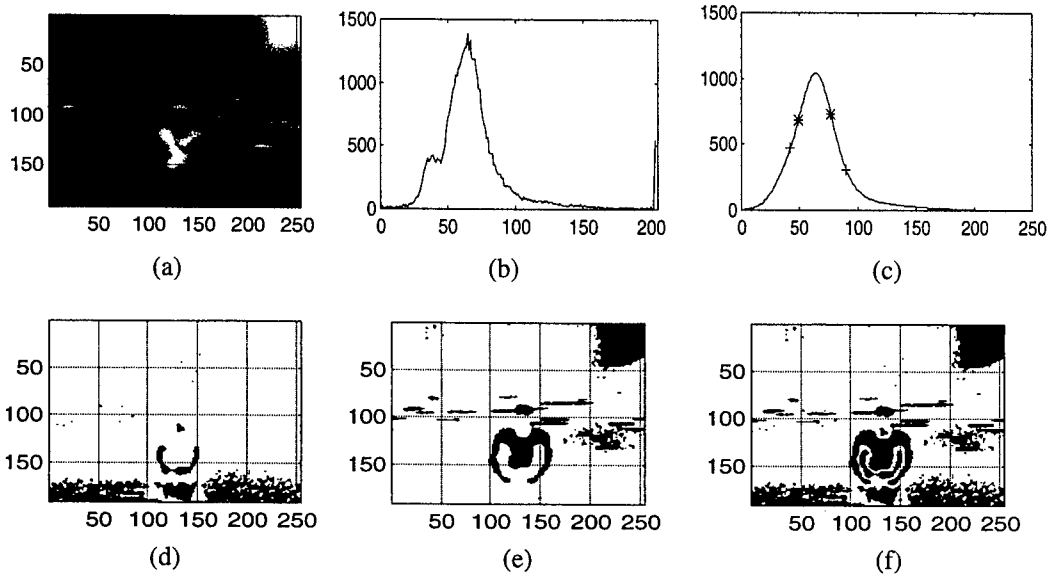
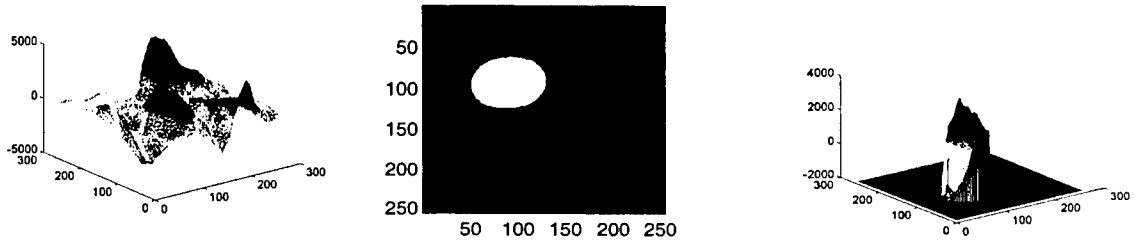


Fig.2.10 An example illustrating body extraction from MMW images

2.3.4 Binary correlation algorithm

The goal of this step is to determine the x and y displacements by correlating the binarized IR and MMW images. We use 1 and -1 to represent the two levels of the thresholded images and use 256 × 256 FFT and IFFT to compute the 2D correlation function. The result of correlating the extracted IR silhouette shown in Figure 2.8(f) with the extracted MMW silhouette shown in Figure 2.10 (f) is shown in Figure 2.11(4-8) (a). The correlation value at the point (x, y) is equal to the correlation of the binarized images when the top left point of the IR image, i.e. the point (1,1), corresponds to the point (x, y) in the MMW image. Note that there are multiple peaks. They result from the imperfect silhouette extracted from the MMW image due to noise and unwanted light reflections. There are two ways to overcome this problem. First, we can construct a mask for the

MMW image that removes the effects of noise and light reflections. The other approach is to construct a mask on the 2D correlation function to select the right peak. In practice, we found that the latter is easier to implement and was used in our algorithm. Figure 2.11 (b) shows the mask that will be used to select the right peak. Figure 2.11 (c) is the correlation function after applying the mask. The steps used to construct the mask shown in Figure 2.11 (b) are described in the next section.



(a) (b) (c)

Figure 2.11: 2D correlation function before and after applying the mask

2.3.5 Mask construction algorithm

Light reflections hamper the extraction of body silhouettes from MMW images. They exist not only in the boundary of the body but also in the image background. Therefore, the thresholded MMW image contains a lot of noise that results in multiple peaks in the 2D correlation function. Sometimes the strengths of the incorrect peaks are stronger than the strength of the correct one. Therefore, it is necessary to develop a procedure to identify the right peak that yields correct x and y displacements. We take advantage of the observation that the background of the MMW image is

smoother than the body portion. Using this property, we developed the following algorithm to construct a mask to select the right peak.

We first convolve the original MMW image with [1 –1] and then take the absolute value so that the differences between each pixel and the pixel to its right can be visualized easily. The result, denoted by I_d, provides a measure of image smoothness. After that, we threshold I_d and pass it through a low pass filter and threshold it again to obtain a mask on the MMW image. The first threshold is determined from the histogram of I_d. This histogram is normalized such that the maximum value of the histogram equals one. Then the threshold is chosen to be $2*q$, where q is the smallest value on the x-axis that yields a slope larger than -0.1. The second threshold is chosen to be half of the maximum value of the low pass filtered image. The mask thus obtained reduces the region where the body can be found. From mask1 we can construct another mask, mask2, to identify the right peak in the correlation function. This is accomplished by shifting mask1 by an amount [$-Xg$, $-Yg$], where [Xg, Yg] is the center of gravity of the silhouette extracted from the IR image.

Figure 2.12 summarizes the steps of the algorithm, and one example is given in Figure 2.13. Figs. 2.13 (a) and 2.13 (b) show I_d and its histogram. The threshold t and point q are also shown.

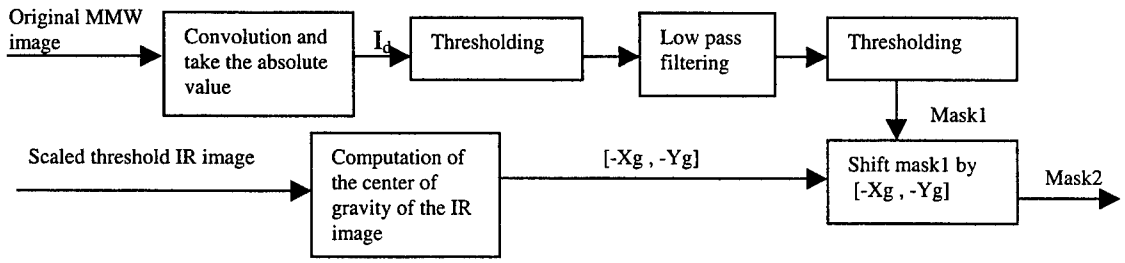


Figure 2.12: Block diagram for mask construction algorithm

Fig. 2.13 (c) is the thresholded image of I_d using threshold t. Fig. 2.13 is the thresholded image after passing it through a low pass filter, which is mask1. Mask2 is shown in Fig. 2.13 (e), which is the

mask used in the previous section to select the right peak. Note that the dimensions of mask1 are the same as the dimensions of the MMW image, while the dimensions of mask2 are the same as the dimensions of the 2D correlation function. The complete registration algorithm including the mask is shown in Figure 2.14.

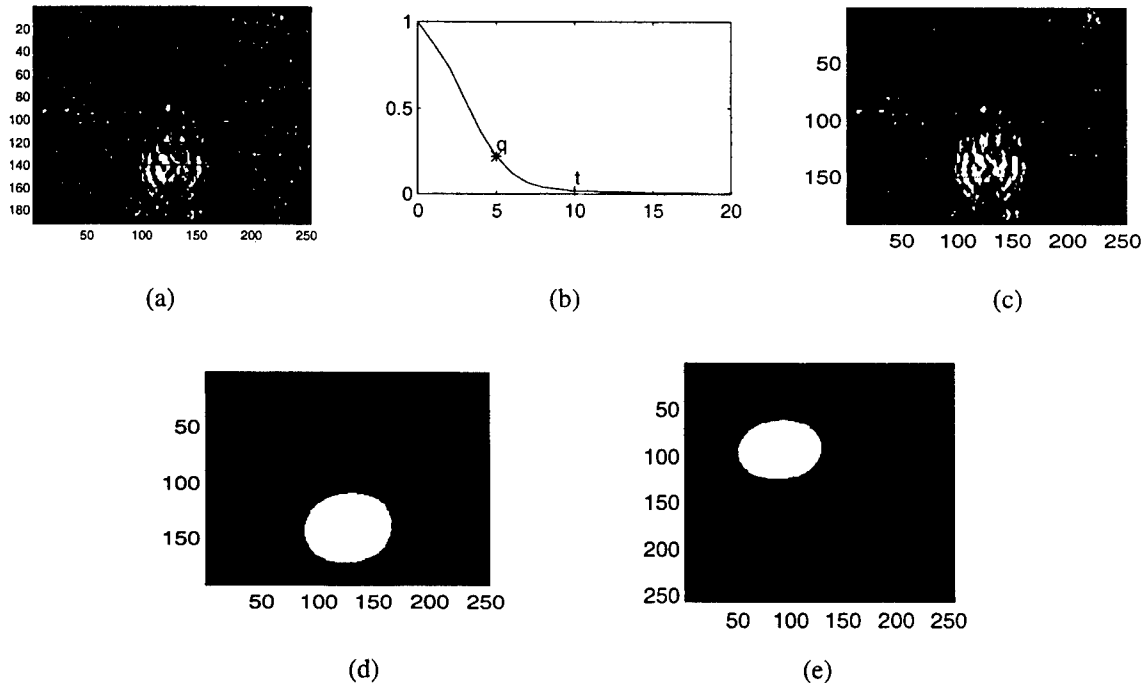


Fig.2.13 An example showing mask construction steps

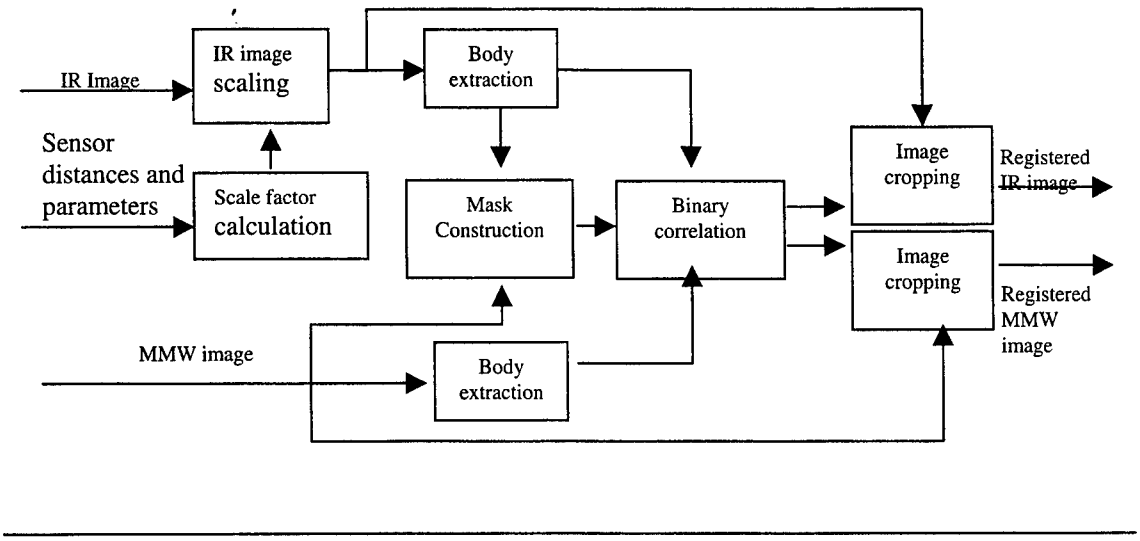


Figure 2.14. Complete registration algorithm

2.4. Experimental Results

Five IR and MMW image pairs are used in our experiments. The size of each IR image is 257×258 and the size of each MMW image is 192×254. Each pair of images is assumed to be taken simultaneously and the transformation between IR and MMW images is assumed to be a rigid body transformation without rotation as stated previously. Figure 2.15 shows typical IR and MMW images with distances from the sensor to the person.

First, we assume that the scale factor can be calculated from the sensor distances and parameters. However the formula for calculating the scale factor can not be easily derived because the formation of IR and MMW images is quite different from the formation of optical images. Instead of calculating the scale factor directly, we try to get it empirically. By manually registering the IR and MMW image pairs we collected, we found the scale factor between the IR and MMW image is about 0.32 if they are taken from the same distance. The scale factor between any IR and MMW image pair is than approximated by $S = 0.32*d_{IR}/d_{MMW}$ where d_{IR} is the distance from the

object to IR sensor and d_{MMW} is the distance from the object to MMW sensor. The parameters used in all of the five experiments are: number of modes ,$M = 3$, and constants $L1 = 1$, and $L2 = 1.5$.

Figures 2.16-2.20 show the experimental results from the five image pairs shown in Fig 2.15. The original image pairs, the extracted binary images, and the registered image are given in each figure. The registered MMW images are shown with the edges of the registered IR image superimposed. The scale factors and the x-y displacements are listed in Table 2.2.

By comparing the registered MMW image and the boundary of the registered IR image for the five examples, we see that the g59/14*t*9 and g62/14*t*8 pairs (Figures 2.19 and 2.20) are not registered satisfactorily. One reason the algorithm did not work well for these examples is that during data collection the sensors were neither co-located nor exactly parallel to each other. The transformation between the IR image and MMW image in Figure 2.20 includes rotation in addition to simple x and y displacements. So the assumption that only x and y displacements are needed to register the image pairs is violated. Since we are taking advantage of body silhouettes to register the IR and MMW image pairs, it is essential to obtain the complete shape of the body. In the pairs of Figures 2.16-2.18, the body portion was completely visible in the IR image, while in the g59/14*t*9 pair of Figure 2.19, the upper portion of the shoulder was not included in the IR image. Therefore, misregistration in the vertical direction is quite large.

Table 2.2: Registration parameters

	dx	*dy*	*s*
g35/13t6 pair	26	-15	0.42
g39/13t7 pair	29	10	0.40
g32/13t5 pair	20	9	0.45
g59/14t9 pair	8	-22	0.36
g62/14t8 pair	22	-1	0.36

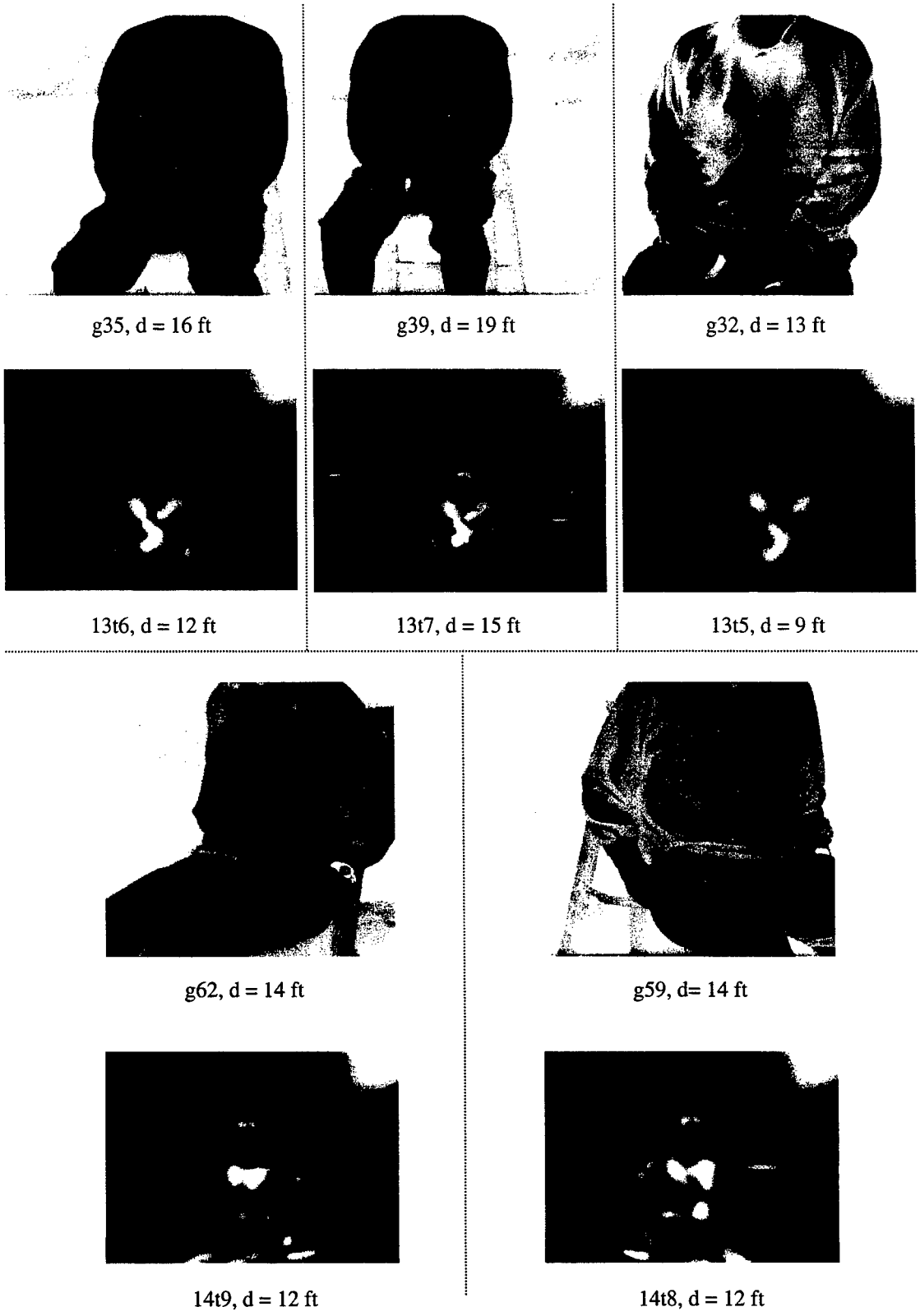


Figure 2.15. Image pairs used in our experiments. d is the distance between object and imager.

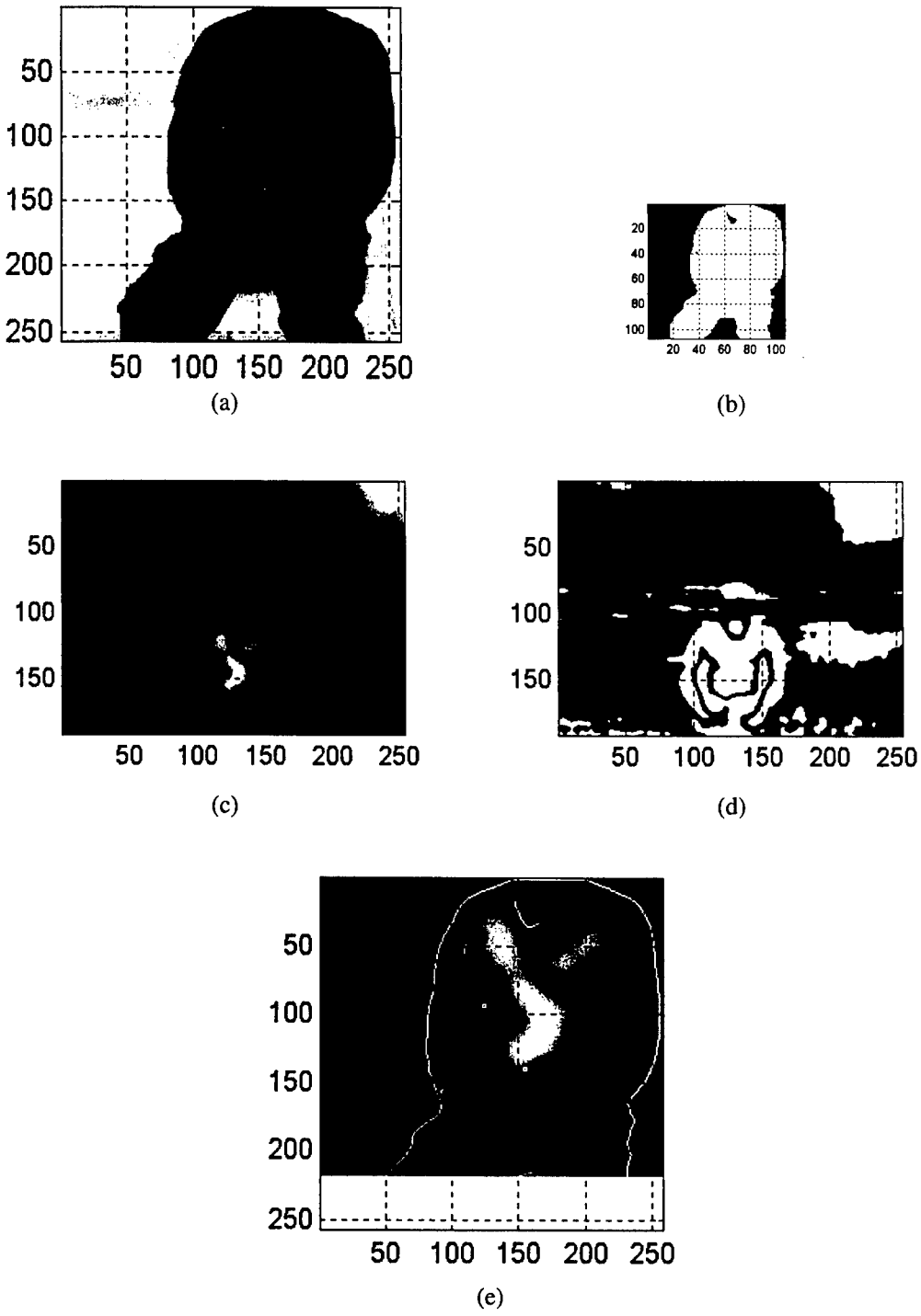


Figure 2.16: Registration results for the g35 pair.

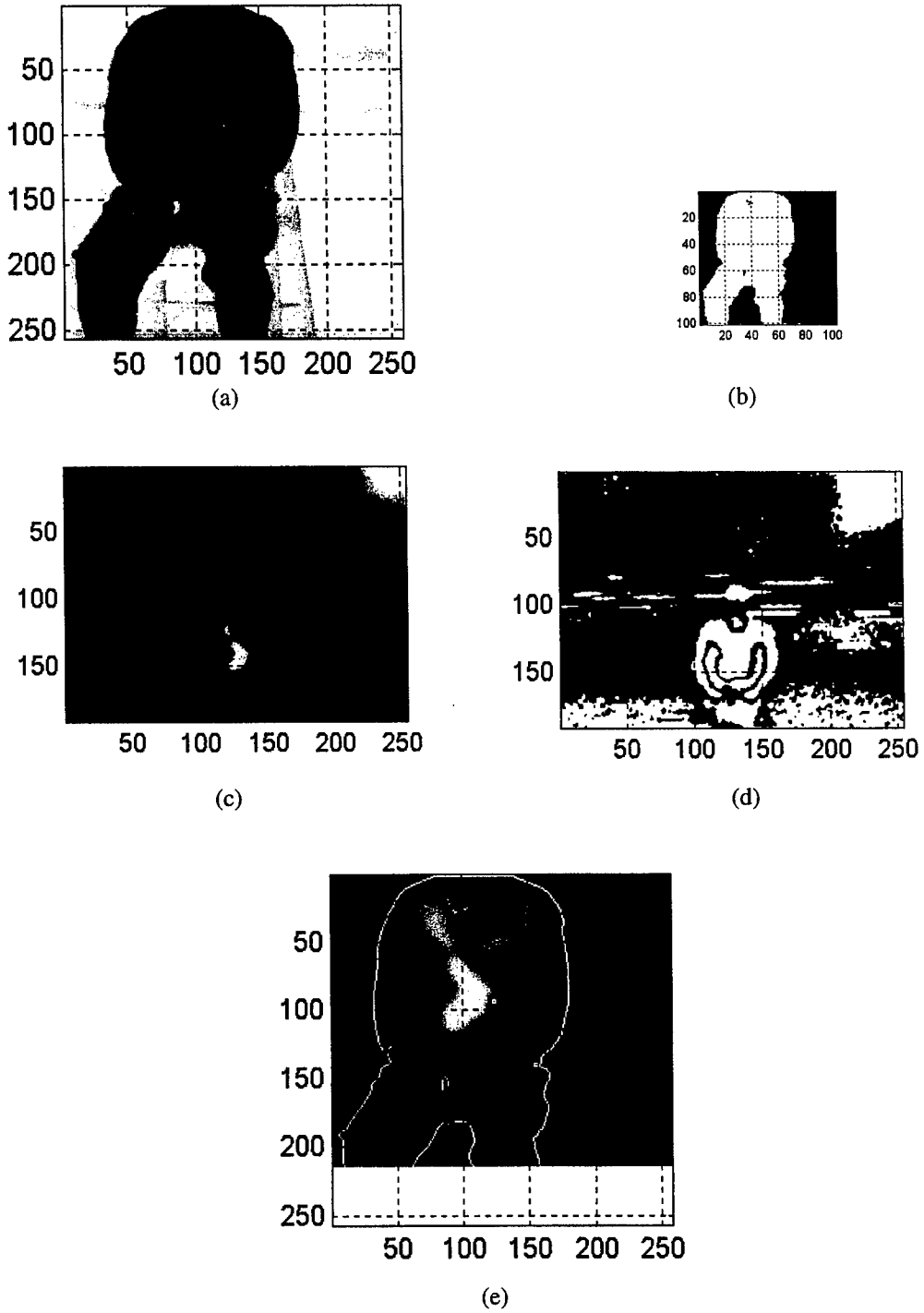


Figure 2.17: Registration results of the g39 pair.

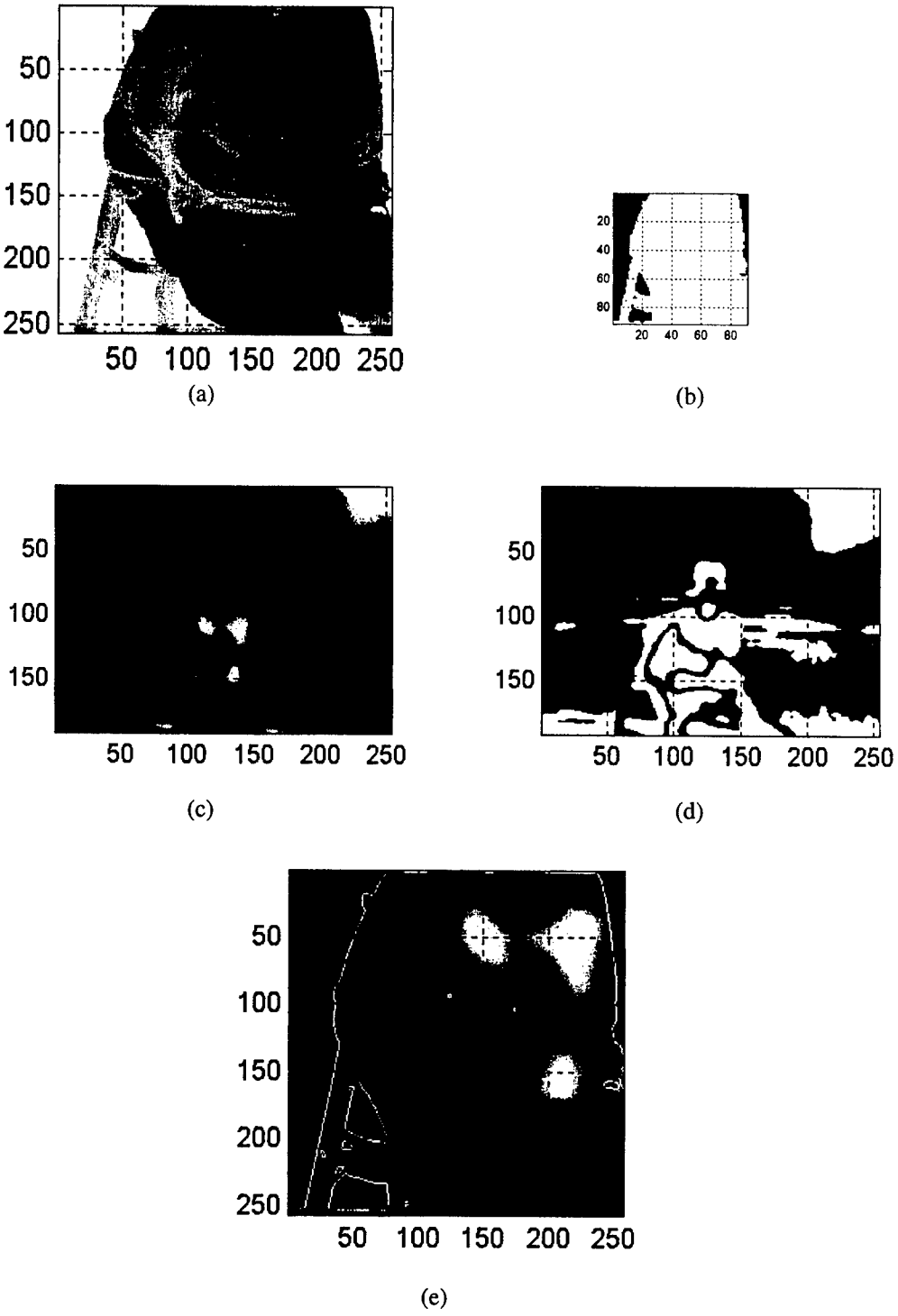


Figure 2.19: Registration results of the g59 pair.

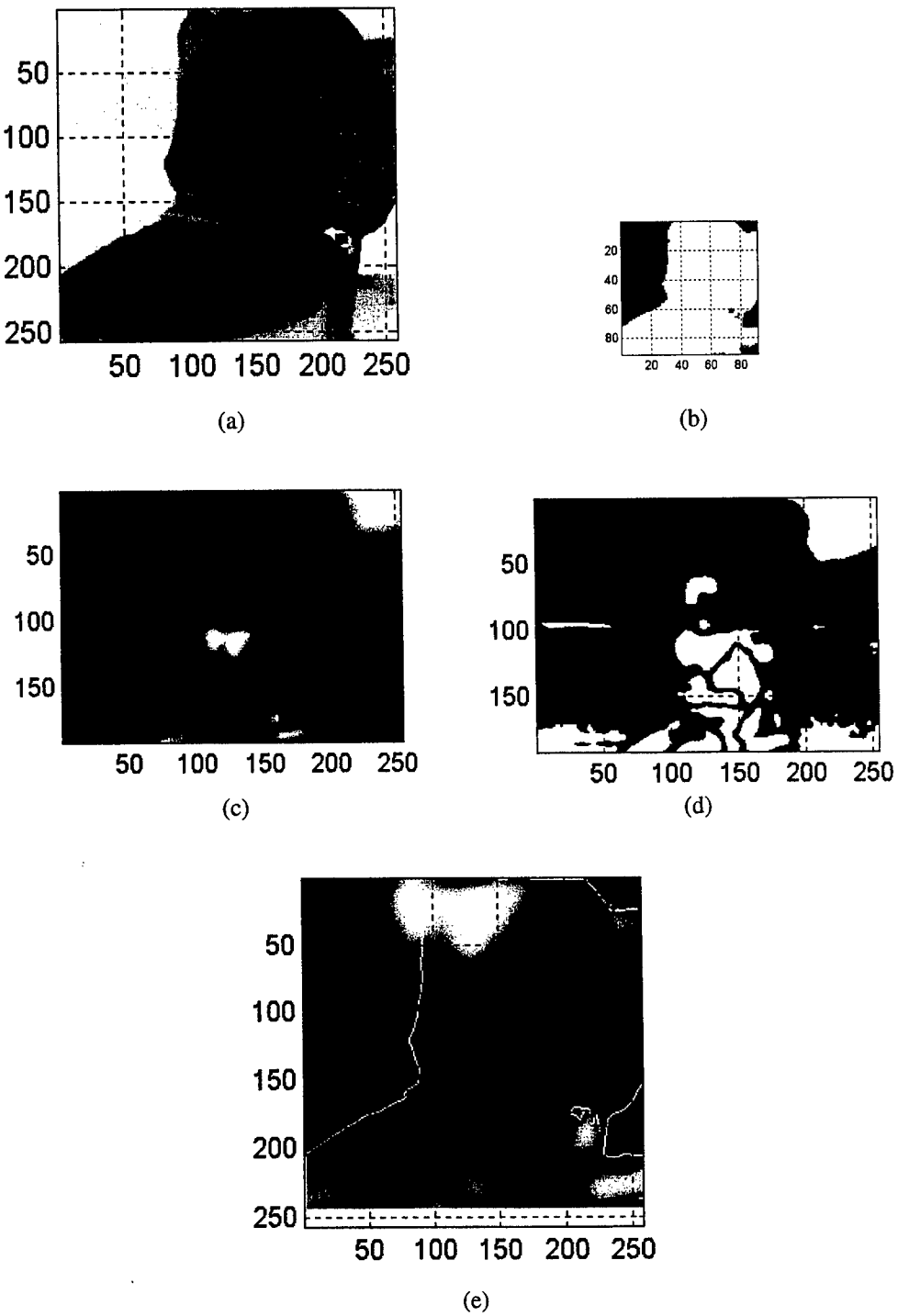


Figure 2.20: Registration results of the g62 pair.

3. MORPHOLOGICAL FILTERING

Original sensor images may often contain details such as shadows, wrinkles, imaging artifacts, etc., that are not needed in the final fused image and can adversely affect the performance of the recognition stage of CWD. Therefore, before fusing the images, it is desirable that they go through a preprocessing in which these unwanted details from an image are removed. This would help improve the recognition performance after fusion. Here, we use morphological filters in the preprocessing stage to clean out the unwanted details.

Mathematical morphology [GW87] is an image processing tool for extracting image features which are useful for representing or describing shapes. It can also be used for pre-processing such as filtering or removing objects within a given size range. The basis for mathematical morphology is set theory, where the sets may represent shapes of objects in an image. First, we will explain morphology (or morphological filtering) for binary images and later extend it to gray-scale images.

3.1 Binary Morphology

Most morphological operations are built upon two basic operations, dilation and erosion. Before defining these operations, some notation will be explained. Let A, B be sets in Z^2 with coordinates $a = (a_1,a_2)$ and $b = (b_1,b_2)$ in two dimensional space being elements of A and B, respectively:

$$A = \{a|a = (a_1, a_2)\} \text{ and } B = \{b|b = (b_1, b_2)\}.$$

The *translation* of set A by $x = (x_1,x_2)$ is defined as

$$A_x = \{c | c = a + x, a \in A\}.$$

The *reflection* of set A is defined as

$$\hat{A} = \{x | x = -a, a \in A\}.$$

The *complement* of set A is

$$A^c = \{x | x \notin A\}.$$

So the complement of set A, A^c, consists of all elements in Z^2 which are not in A. The *difference* between sets A and B is defined as

$$A - B = \{x | x \in A, x \notin B\} = A \cap B^c.$$

Now we define the morphological operations dilation and erosion for binary images. The *binary dilation* of A by B, where A is the object and B is called the structuring element, is defined as

$$A \oplus B = \{x | \hat{B}_x \cap A \neq \varnothing\} = \{x | [\hat{B}_x \cap A] \subseteq A\}$$

This process can be explained as all the translations x of $\hat{B}$ where $\hat{B}_x$ and A overlap by at least one element. So, the operation of dilation expands the object in the image, or increases the number of elements in the dilated result. The *binary erosion* of A by B is defined as

$$A \overset{\ominus}{} B = \{x | B_x \subseteq A\}.$$

This process can be explained as all the translations x of B where B_x is completely contained in A (all elements in B_x must overlap with elements in A for x to be an element in the eroded result). So, the operation of erosion shrinks the object in the image, or decreases the number of elements in the eroded result.

Figure 3.1 demonstrates binary dilation and erosion for an image. Figure 3.1(a) shows the original image and the structuring element. Figure 3.1(b) shows the difference between the original image and the dilated portion and the dilated result. The image on the left shows the original object in gray with the additional dilated portion shown in black. We can observe that gap on the left was filled in by the dilation operation. The image on the right shows the dilated result in black. The difference between the original object and the eroded portion and the result of erosion are shown in Figure 3.1(c). The image on the left shows the original object (in black and gray) and the smaller eroded portion in black. The portion that was removed is in gray. We see that the narrow part in the center of the object and the two narrow parts on the right side of the object are removed. The image on the right shows the final eroded object.

Now that the basic morphological operations have been defined, we define two more commonly used operations that are built upon dilation and erosion. These two operations are called opening and closing. The *binary opening* of A by B is defined as

$$A \circ B = (A \ominus B) \oplus B.$$

In words, the opening of A by B is A eroded by B followed by a dilation by B. This operation will smooth the object contours and remove the narrow parts of the object which are smaller than the structuring element. The *binary closing* of A by B is defined as

$$A \bullet B = (A \oplus B) \ominus B.$$

In words, A closed by B is A dilated by B followed by an erosion by B. This operation will also smooth the contours but, instead of removing narrow parts, will fill in narrow gaps in the object which are smaller than the structuring element.

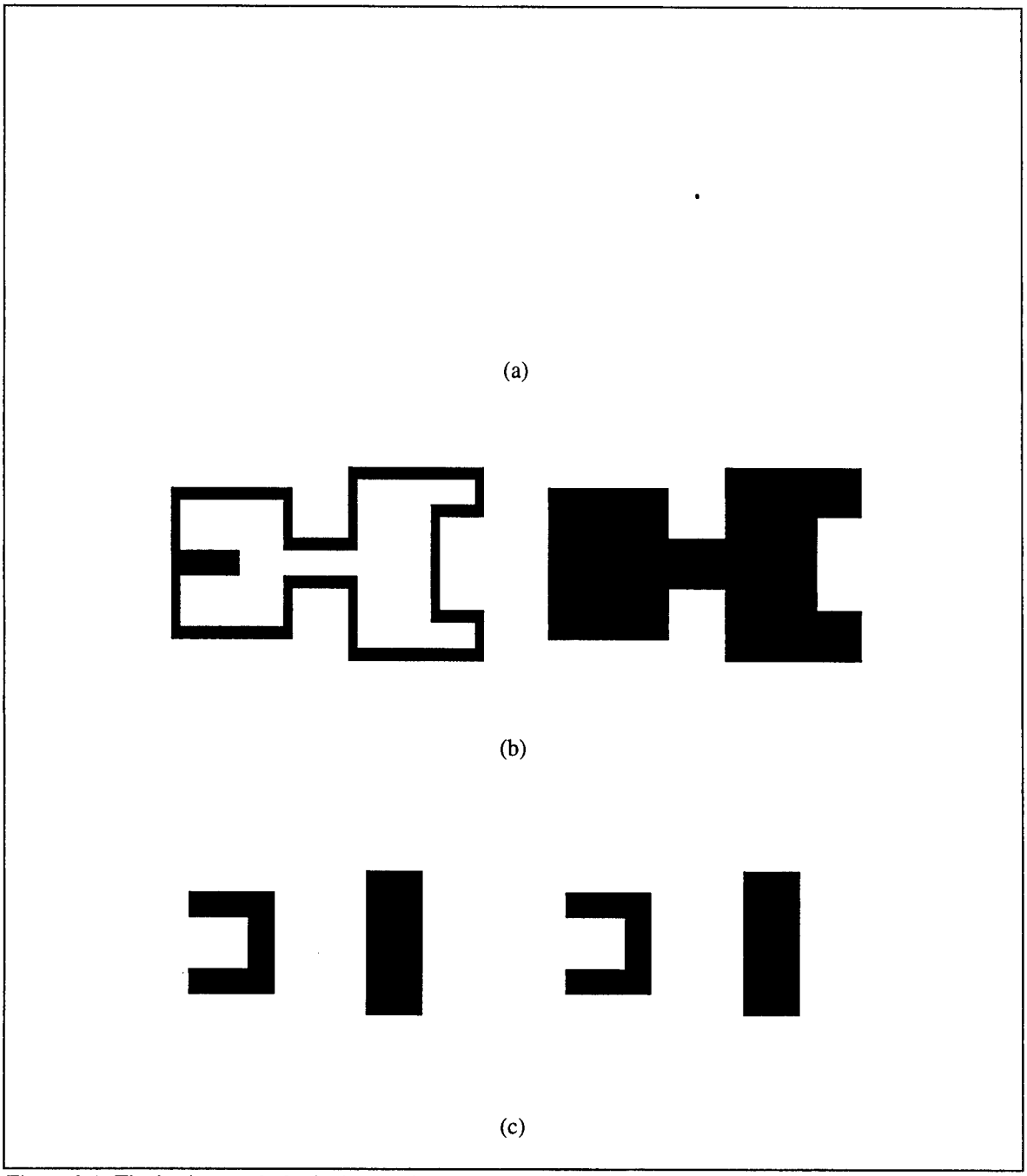


Figure 3.1: The basic morphological operations: (a) original image and structuring element, (b) the dilated image and (c) the eroded image.

An example of binary opening and closing is given in Figure 3.2. The same object and structuring element as in Figure 3.1 are used for this example. Figure 3.2(a) shows the result of erosion followed by dilation that gives the opened result. The left image shows the object (black and gray) eroded first (black portion only). Then the right image shows the final opened result in black where the gray portion is what was removed from the original object. The result is the black part of the figure on the right. Figure 3.2(b) shows the result of dilation followed by erosion that gives the result of closing the image. The image on the left shows the dilation of the original object (shown in gray). On the right, the final closed result is shown in gray and black, where the portion that was added to the original object by this operation is black.

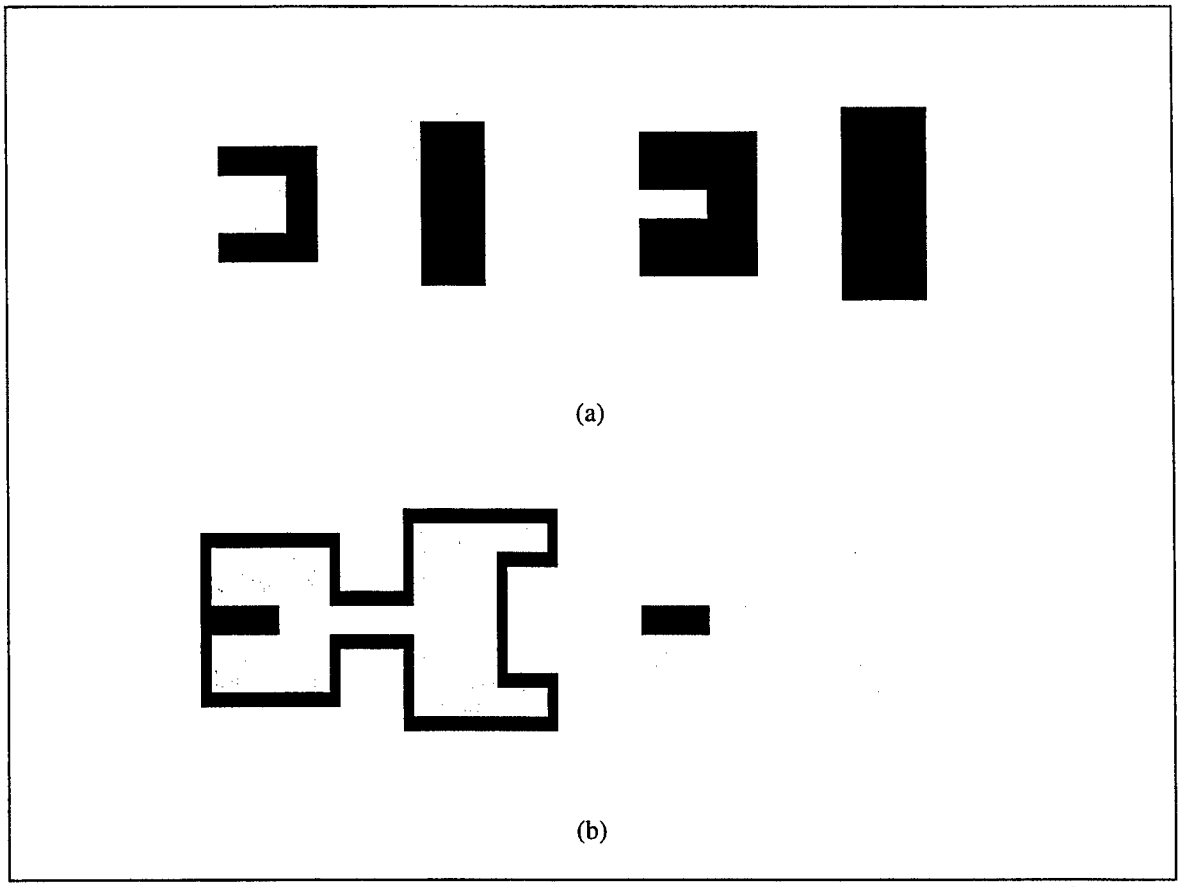


Figure 3.2: The (a) binary opening - erosion followed by dilation and (b) binary closing - dilation followed by erosion.

3.2 Gray-scale Morphology

Now the concepts of dilation, erosion, opening and closing are extended to gray-scale images. Instead of dealing with sets such as A and B in the previous discussion, we use functions of the form $f(x,y)$ and $b(x,y)$, where $f(x,y)$ is the input image and $b(x,y)$ is the structuring element. These operations are now in three-dimensional space instead of two-dimensional space and the value that changes is the graylevel not the spatial coordinate. Here we translate the image instead of the structuring element, but the operations can also be performed by translating the structuring element. *Gray-scale dilation* is defined as

$$(f \oplus b)(s,t) = \max\{f(s-x,t-y)+b(x,y) \mid (s-x),(t-y) \in D_f ; (x,y) \in D_b\}$$

where D_f and D_b are the domains of $f(x,y)$ and $b(x,y)$, respectively. This operation will brighten the entire image and fill in any dark valleys smaller than the structuring element, because it chooses the maximum value within a window determined by the size of the structuring element. *Gray-scale erosion* is defined as

$$(f \ominus b)(s,t) = \min\{f(s+x,t+y)-b(x,y) \mid (s+x),(t+y) \in D_f ; (x,y) \in D_b\}.$$

This operation tends to darken the overall image while removing any bright peaks that are smaller than the given structuring element. This is because this operation chooses the minimum graylevel within a window.

A one-dimensional example of gray-scale dilation and erosion is given in Figure 3.3. The original signal (this can be thought of as one row or column of an image) is shown in Figure 3.4(a). The structuring element used in this example was 1×5 with magnitude equal to one. Figure 3.4(b) shows the relation of the dilated signal (solid line) to the original signal (dotted line). We can observe

that the values are larger over the entire signal, and the valleys smaller than five pixels wide were removed. Figure 3.3(c) shows the result of gray-scale erosion in comparison to the original signal. This shows the eroded signal (solid line) in comparison to the original (dotted line). We see that the values over the whole signal are lower, and the peaks smaller than five pixels wide were removed.

Now the expressions for gray-scale opening and closing are defined, they have the same structure as the binary case. The *gray-scale opening* of an image $f(x,y)$ by structuring element $b(x,y)$ is defined as

$$f \circ b = (f \ominus b) \oplus b.$$

The *gray-scale closing* of $f(x,y)$ by $b(x,y)$ is defined as

$$f \bullet b = (f \oplus b) \ominus b.$$

Opening will remove the bright peaks of the image which are smaller than the structuring element but leave the dark valleys unaffected. Closing will remove the dark valleys of the image but leave the bright peaks unaffected. Neither operation will affect the overall brightness of the image. So, if an opening is followed by a closing with the same structuring element, bright peaks and dark valleys will be removed.

Some properties of opening and closing are now given which apply to both the binary and gray-scale cases. First, opening and closing are not commutative. Therefore, for most images a change in the order of operations will produce different results:

$$(f \circ b) \bullet b \neq (f \bullet b) \circ b.$$

In addition, opening and closing are both monotonically increasing and idempotent.

If $f_1 \subseteq f_2$ then $(f_1 \circ b) \subseteq (f_2 \circ b)$ and $(f_1 \bullet b) \subseteq (f_2 \bullet b)$

and $(f \circ b) \circ b = (f \circ b)$ and $(f \bullet b) \bullet b = f \bullet b$.

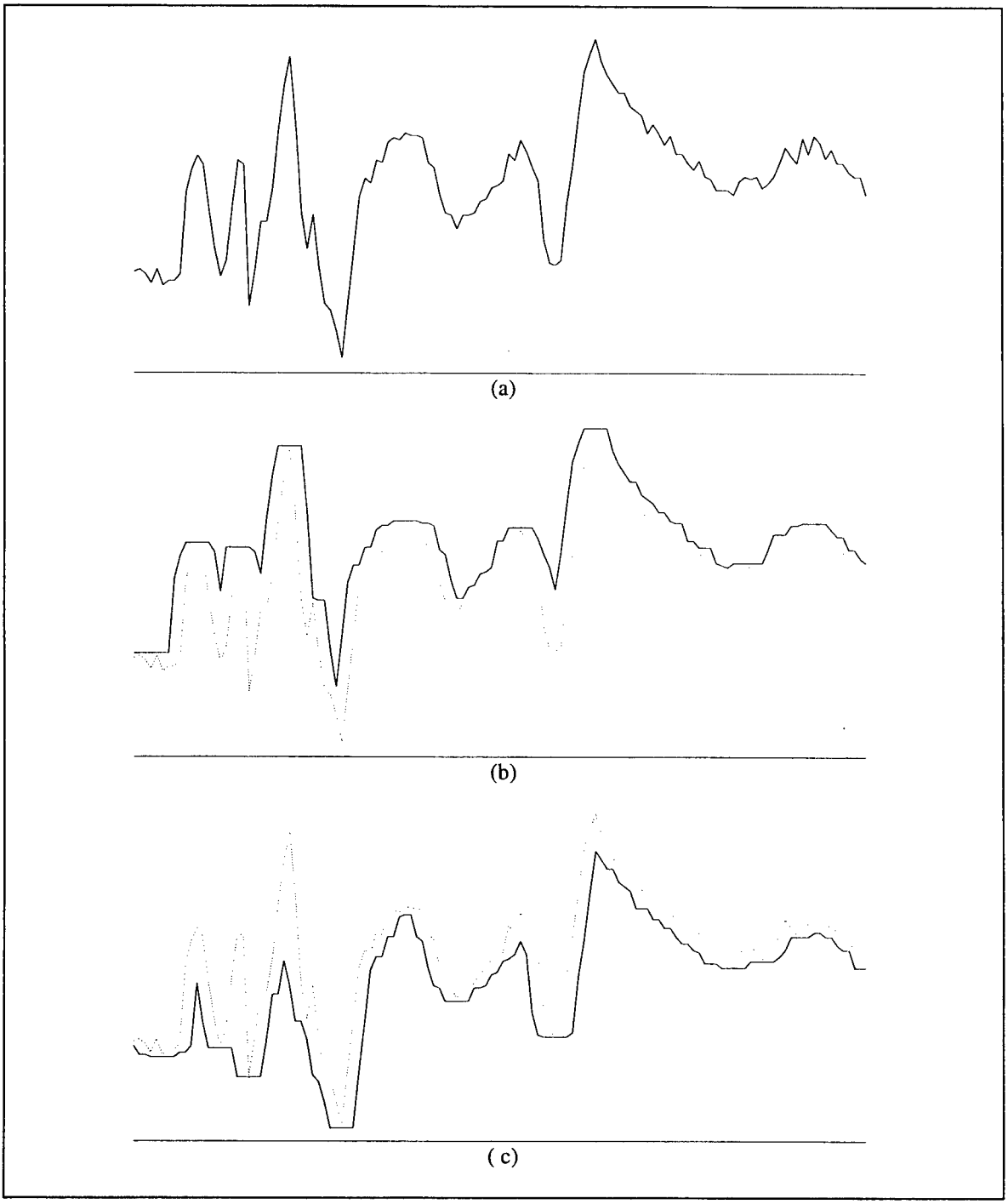


Figure 3.3: Basic gray-scale operations: (a) the original signal, (b) the dilated signal and (c) the eroded signal.

An example of gray-scale opening and closing are shown in Figure 3.4. The signal and structuring element of Figure 3.3 are used here. Figure 3.4(a) shows the result of gray-scale opening (solid line) in relation to the original signal (dotted line). It shows that the bright peaks were removed from the signal while the other values were unchanged. Figure 3.4(b) shows the result of gray-scale closing (solid line) in relation to the original signal (dotted line). It shows that the dark valleys were removed from the signal while leaving the other gray values unchanged.

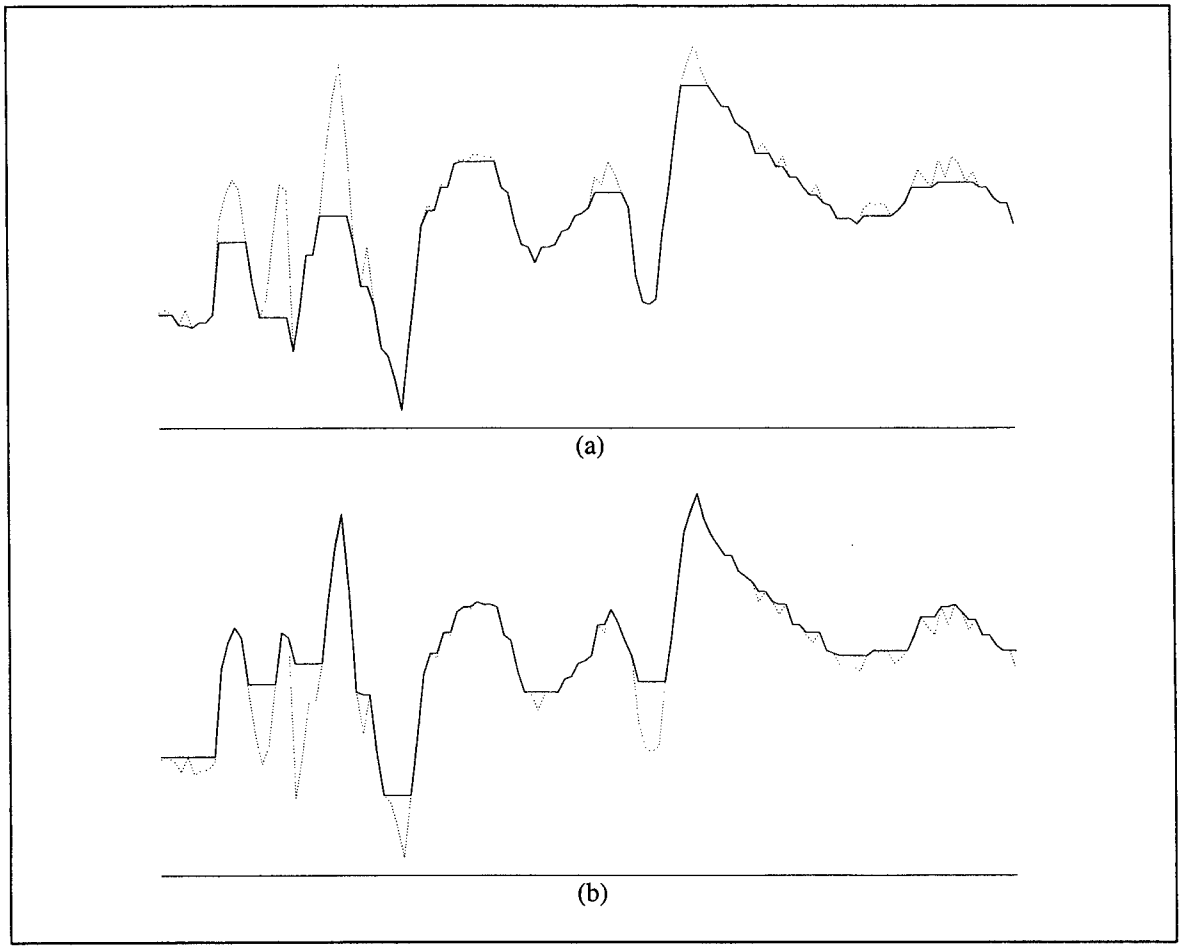


Figure 3.4: Relationships between (a) gray-scale opening and the original signal and (b) gray-scale closing and the original signal.

3.3 Mathematical Morphology for CWD

Since mathematical morphology is suited for describing or representing shape in an image, it is a useful tool for "cleaning" gray-scale images. If we choose the structuring element to be larger than the unwanted details in the image, we can remove both the dark and the bright details by filtering with an opening followed by a closing (or closing followed by an opening). Another useful aspect of morphological filtering is that we can extract objects of a certain size range from the image.

Examples of using morphological filtering for noise removal are shown in Figure 3.6. The images shown are from IR and MMW sensors. The original images were not registered, so the examples here use images (Figure 3.5) which were manually registered. Figure 3.6(a) shows the morphologically filtered versions of the IR and MMW images using a 3 by 3 filter (opening followed by a closing). Figure 3.6(b) shows the morphologically filtered versions of the IR and MMW images using a 5 by 5 filter. Figure 3.6(c) shows the morphologically filtered versions of the IR and MMW images using a 9 by 9 filter. Visually, the MMW images do not seem to change much after morphological filtering but the IR images appear to contain less detail as the filter size increases. The filtered IR images appear smoother around the arm and neck areas, but the contrast around the gun remains about the same.

Figure 3.5: The original IR and MMW images (manually registered).

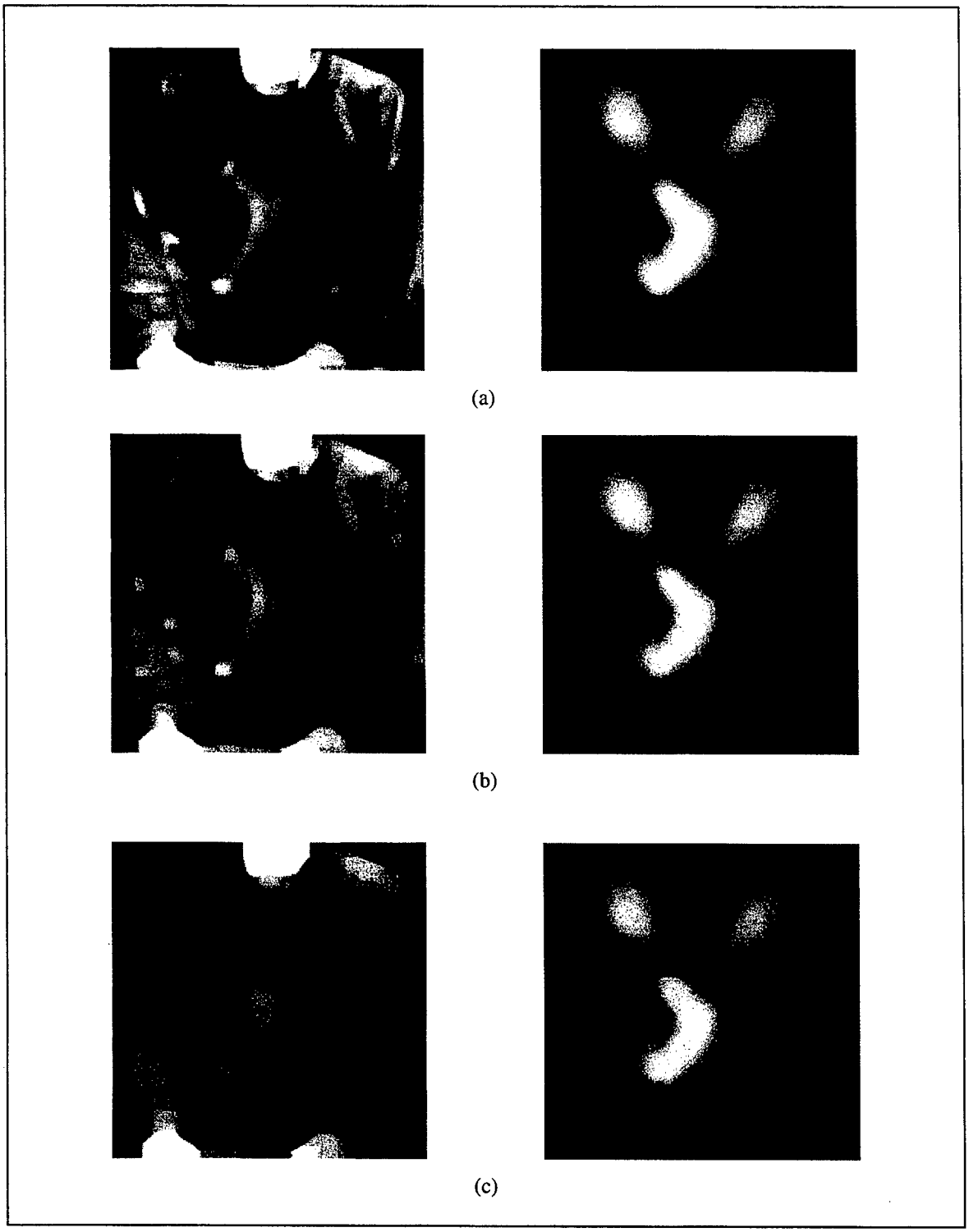


Figure 3.6: The morphologically filtered versions using (a) a 3 by 3 filter (b) a 5 by 5 filter and (c) a 9 by 9 filter.

4. IMAGE FUSION

A single sensor, depending on its features and/or viewing position, may not receive all the information necessary for detecting an object by human or computer vision. If several sensors with different features and/or viewing positions are used then the image information obtained from each one can be combined with the others. The use of multiple sensors in fields such as remote sensing, medical imaging, and automated machine vision has increased in the past decade. As a result of this, techniques for fusing different sensor images to form a composite image have emerged. This final composite image has more complete and detailed information content than the individual source images. Therefore, the composite or fused image is more useful for human perception as well as for automatic computer analysis tasks such as segmentation, feature extraction, and object recognition. In the case of CWD, one imaging sensor may not be able to detect a weapon depending upon the type of clothing a person may be wearing or the distance between the imaging sensor and the person. The goal of image fusion for CWD is to provide a composite image that improves the capability of detecting weapons on people.

The straightforward approach to image fusion is to take the average of the source images, but this can produce undesired results such as a decrease in contrast. Our fusion method involves multiresolution image decomposition based on the wavelet transform. This transform allows us to fuse details at different resolutions. Burt [Bur84] first proposed the approach as a model for binocular fusion in human stereo vision. His implementation used a Laplacian pyramid and a "maximum" selection rule. Others have used similar pyramid methods, including the wavelet transform, and

different selection rules to fuse images [Toe90, PLA91, BL93, LMM95]. First, an image pyramid is constructed for each source image by applying the wavelet transform to the source images. This transform domain representation will emphasize important details of the source images that will be useful for choosing the best fusion rules. Then, using a feature selection rule, a fused pyramid is formed for the composite image from the pyramid coefficients of the source images. Finally, the composite image is obtained by taking an inverse pyramid transform of the composite wavelet representation. Figure 4.1 demonstrates the general image fusion process. This figure illustrates the fusion process for two source images, but the process can be implemented for combining multiple source images.

Multiresolution decomposition and reconstruction based on the two-dimensional wavelet transform will be described in Section 4.1. The image fusion rules will be described in Section 4.2 and examples shown in Section 4.3.

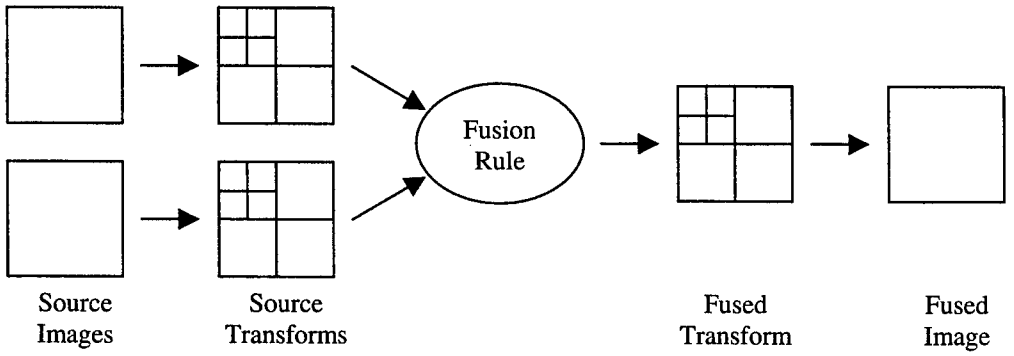


Figure 4.1: General image fusion process.

4.1 Wavelet Decomposition and Reconstruction

Multiresolution image decomposition provides a unique and very useful analysis tool for image processing applications. This type of transform is able to separate details of an image by scale

and orientation. At different resolutions, the details of an image generally characterize different physical structures of the scene. For example, at low resolution, the details correspond to the larger structures in the image. As the resolution gets finer, the details represented in the image also get finer. Depending on the features being searched, multiresolution analysis such as the one based on the wavelet transform provides a very useful mathematical tool for image representation. The idea behind multiresolution analysis is to represent an n-dimensional signal by lower resolution approximation (A) and detail signals (D). These detail signals provide the difference between two approximation signals at successive resolutions. The representations discussed here are for two-dimensional signals (images) with approximations at resolutions 2^j, for $j \leq 0$.

The wavelet transform represents an arbitrary function, f, as a superposition of wavelets. These wavelets are a group of functions that are generated from translations and dilations of one function, ψ, known as a mother wavelet. The wavelet transform is capable of providing uncorrelated data at different resolutions. It is also able to separate details by scale and orientation. Therefore, the wavelet transform provides a complete and orthogonal multiresolution decomposition. Multiresolution analysis using the wavelet transform actually involves two functions, a scaling function $\phi(x)$ and a mother wavelet $\psi(x)$. The scaling function is used to obtain the approximation signals and the wavelets are for the detail signals. Wavelet decomposition is computed by using pyramidal algorithms that are based on convolutions with quadrature mirror filters. Reconstruction of the original image can be accomplished with similar pyramid based algorithms. The following sections will describe decomposition and reconstruction for the two-dimensional case for image analysis. Also, the use of biorthogonal filters will be discussed for image analysis.

The wavelet representation for two dimensions can be computed by using a separable pyramidal algorithm. At each level, the approximation signal $A^d_{2^{j+1}} f$ is decomposed into $A^d_{2^j} f$, $D^1_{2^j} f$, $D^2_{2^j} f$, and $D^3_{2^j} f$ where $A^d_1 f$ is the image at the original resolution. The four subimages

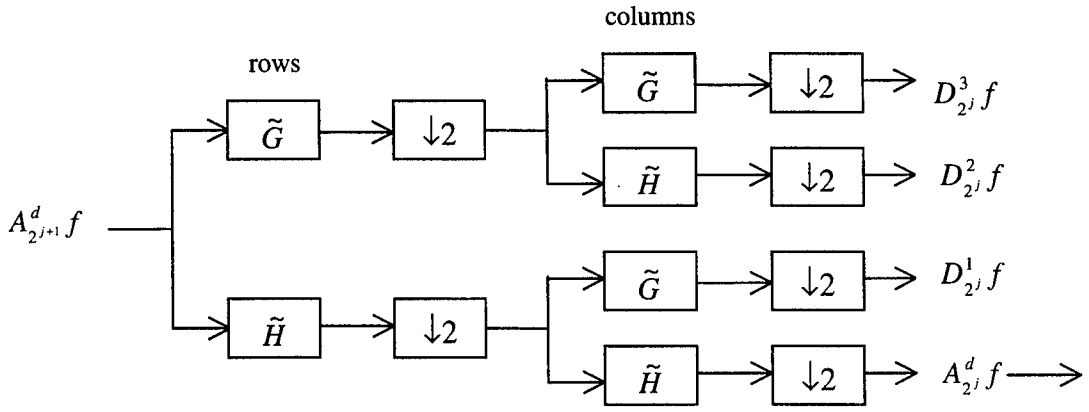


Figure 4.2: One level wavelet decomposition from resolution 2^{j+1} to 2^j for two-dimensional signal.

correspond to a low-resolution image and vertically, horizontally and diagonally oriented detail images. The filters used for this decomposition are the quadrature mirror filters $\tilde{H}$ and $\tilde{G}$ from the one-dimensional analysis. For one level of two-dimensional wavelet decomposition, the rows of $A^d_{2^{j+1}} f$ are convolved with one-dimensional filters $\tilde{H}$ and $\tilde{G}$ and sub-sampled by two. Then, the columns are convolved with one-dimensional filters $\tilde{H}$ and $\tilde{G}$ followed by sub-sampling the columns by two. The wavelet transform of image $A^d_1 f$ is computed by repeating this process for resolutions corresponding to $-1 \geq j \geq -J$, where J is the desired number of decomposition levels. Figure 4.2 shows a block diagram of one decomposition level from resolution 2^{j+1} to 2^j.

A pyramidal algorithm can also describe two-dimensional wavelet reconstruction. At each level, the approximation image $A^d_{2^{j+1}} f$ is reconstructed from subimages $A^d_{2^j} f$, $D^1_{2^j} f$, $D^2_{2^j} f$, and $D^3_{2^j} f$. First, zeros are inserted between neighboring row coefficients in each of the subimages, and the rows are convolved with a one-dimensional filter. Then, zeros are inserted between neighboring column coefficients and the columns are then convolved with another one-dimensional filter. The filters used for reconstruction are the quadrature mirror filters H and G. The image $A^d_1 f$ is

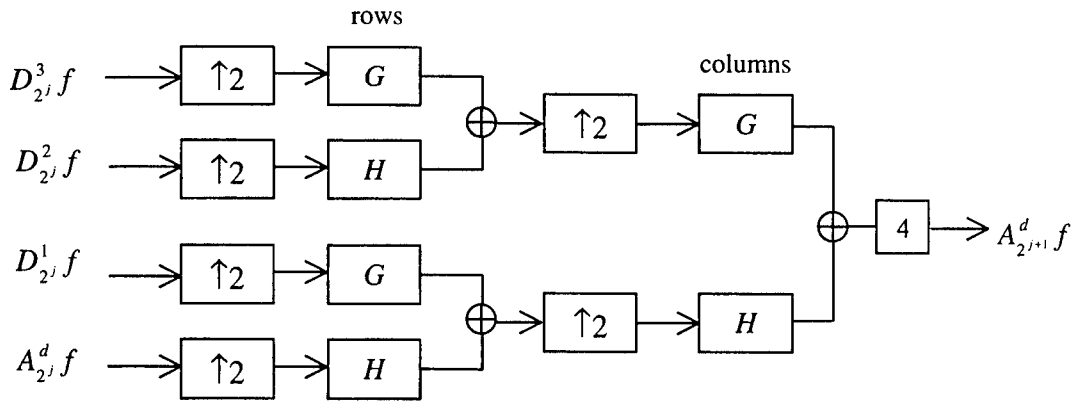


Figure 4.3: One level wavelet reconstruction from resolution 2^j to resolution 2^{j+1} for two-dimensional signal.

reconstructed by repeating this process for levels $-J \le j \le -1$. Figure 4.3 shows a block diagram of the reconstruction process from level 2^j to resolution level 2^{j+1}.

For purposes of fast computation, the filters for decomposition and reconstruction should be short. However, in order to cascade filters in pyramidal structures without needing phase compensation, the filters should also have linear phase. There are no nontrivial orthonormal linear phase filters for exact reconstruction. In order to preserve linear phase, the orthonormality requirement can be relaxed by using biorthogonal bases filters. The use of biorthogonal bases still keeps the same decomposition method as in the orthogonal case, but reconstruction uses filters h' and g' which may be different from h and g. For exact reconstruction, the following restrictions on the filters are imposed:

$$g'(n) = (-1)^n \cdot \tilde{h}(1-n)$$

$$\tilde{g}(n) = (-1)^n \cdot h'(1-n)$$

$$\sum_n \tilde{h}(n) \cdot h'(n+2k) = \delta_{k,0}.$$

4.2 Image Fusion Algorithm

First, the two-dimensional wavelet transform is used to construct a multiresolution representation of each source image. Once the source images are decomposed, the details are combined to form a composite decomposed image. This method allows details at different levels to be combined independently so that important information is maintained in the final composite image. We have developed image fusion algorithms for concealed weapon detection (CWD) applications. Fusion is useful in situations where the sensor types have different properties, e.g., IR and MMW sensors. Fusing these types of images results in composite images that contain more complete information for CWD applications such as detection of concealed weapons on a person.

We apply Burt's feature selection algorithm [Bur84] for image fusion. In this algorithm, salient features are identified in each source image, then are copied to the composite image. The salience of a feature is defined as a local energy in the neighborhood of a coefficient:

$$S(i,j,k) = \sum_m \sum_n c(i+m, j+n, k)^2$$

where (i, j) is the location of the current wavelet coefficient $c(i, j)$, k is the decomposition level, and (m, n) define a window of coefficients around the current coefficient. The size of the window is typically small, i.e. 3 by 3. Less salient features that may partially mask the more salient features are discarded. In this way, the features in the composite image are obtained at full contrast and double exposure artifacts are avoided.

At a given resolution level, the fusion algorithm uses two distinct modes of combination: selection and averaging. In addition to the salience measure, the image fusion algorithm uses a match measure that determines the mode of combination. The match measure at location (i, j) and decomposition level k is defined as

$$M_{AB}(i,j,k) = \frac{2\sum_m \sum_n c_A(i+m, j+n, k) \cdot c_B(i+m, j+n, k)}{S_A(i,j,k) S_B(i,j,k)}$$

where subscripts A and B refer to the corresponding source image for the coefficients and salience measures. To determine whether selection or averaging will be used, M_{AB} is compared to a threshold, α. At each coefficient position the salience measure determines which source coefficient is chosen in the selection mode. If M_{AB} is less than or equal to α, then the coefficient with the largest salience is placed in the composite transform while the less salient coefficient is discarded. The selection mode is implemented as:

$$c_C(i,j,k) = \begin{cases} c_A(i,j,k) & \text{if } S_A(i,j,k) \geq S_B(i,j,k) \\ c_B(i,j,k) & \text{if } S_A(i,j,k) < S_B(i,j,k) \end{cases}$$

where c_C are the coefficients in the composite wavelet transform.

If M_{AB} is greater than α, the source patterns are more similar and the weighted average is calculated from coefficients of both source transforms. The weights used for averaging are defined as follows:

$$w_{\min} = \frac{1}{2}\left(1 - {(1 - M_{AB})}\Big/{(1-\alpha)}\right),$$

$$w_{\max} = 1 - w_{\min},$$

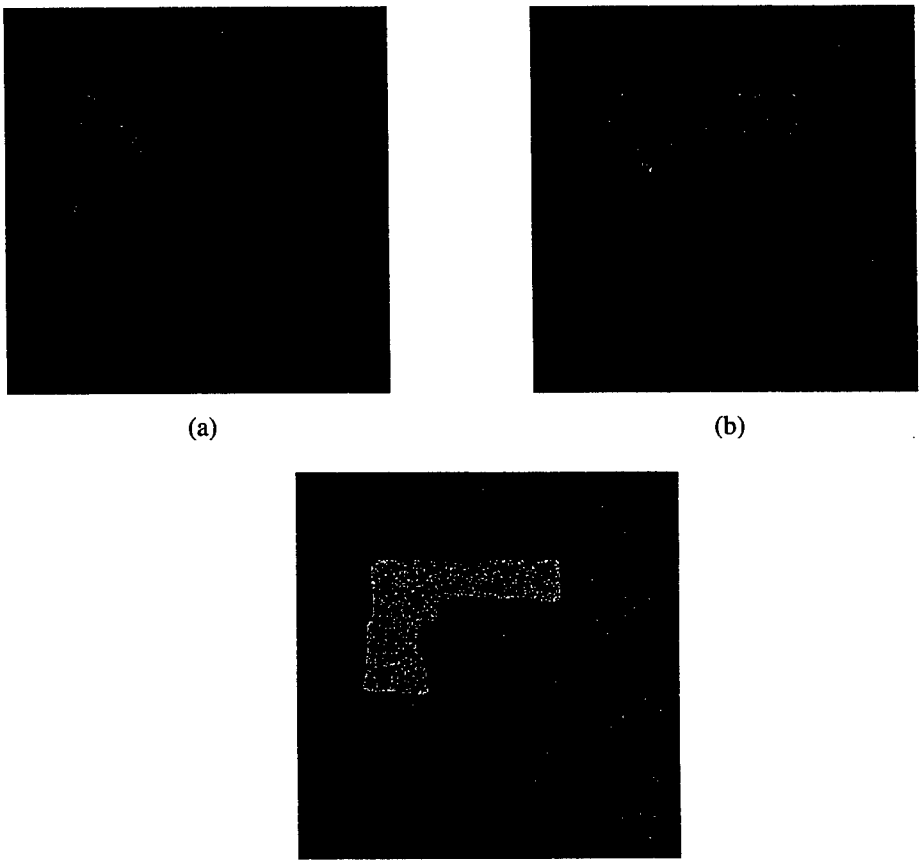
where w_{min} is applied to coefficient with lower salience and w_{max} is applied to the coefficient with higher salience. In the averaging mode, the composite transform coefficient is the weighted average of the source coefficients and is implemented as:

$$c(i,j,k) = \begin{cases} w_{\min}(i,j,k) \cdot c_A(i,j,k) + w_{\max}(i,j,k) \cdot c_B(i,j,k) & \text{if } S_A(i,j,k) \leq S_B(i,j,k) \\ w_{\max}(i,j,k) \cdot c_A(i,j,k) + w_{\min}(i,j,k) \cdot c_B(i,j,k) & \text{if } S_A(i,j,k) > S_B(i,j,k) \end{cases}$$

After all of the coefficients in the composite wavelet representation are obtained, the inverse wavelet transform is performed to get the final fused image.

4.3 Image Fusion Examples

For testing our image fusion algorithms, we were provided with images from IR and MMW sensors for the CWD problem. First, we tested the fusion algorithm on synthetically generated images where different portions of a gun were visible in the source images. Next, we fused images where both source images were obtained from IR sensors with different parts of the gun partially hidden in each image. The third type of image pairs we fused was from both IR and MMW sensors.



(a) (b)

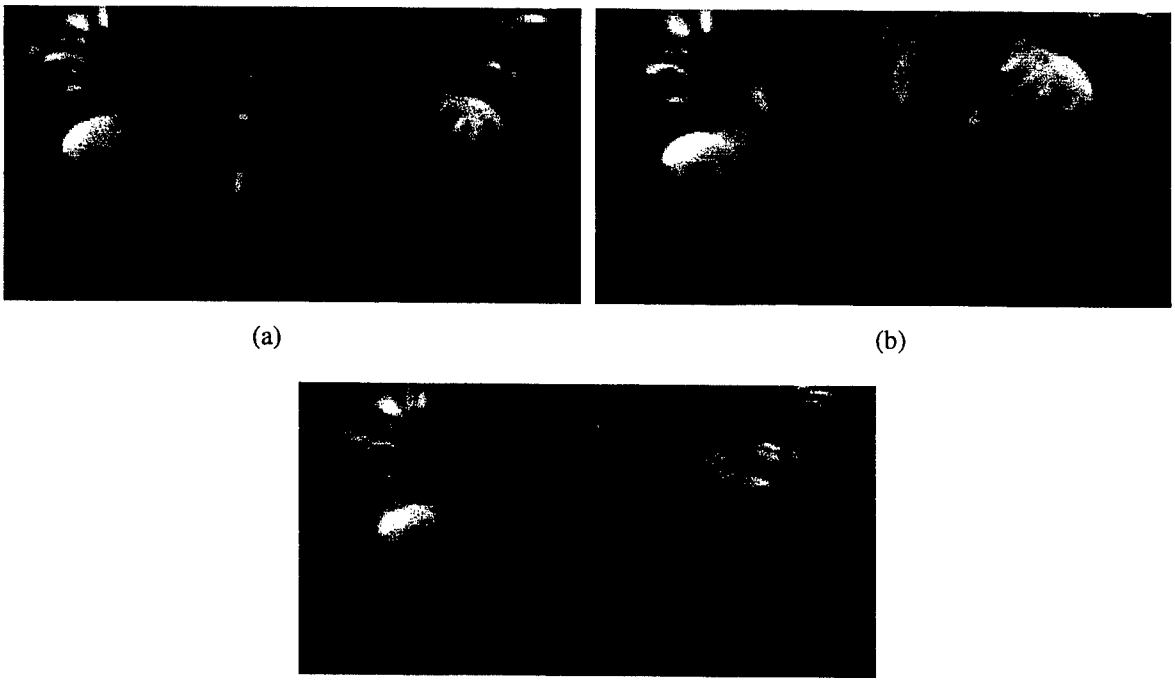
(c)

Figure 4.4: Synthetic image pairs: (a), (b) source images and (c) result.

Our first example deals with fusing the synthetic image pairs. Figure 4.4(a) and (b) show the source images used for fusion. The handle of the gun is visible in one image and the barrel is visible in the other image. These images were used using two levels of decomposition and a threshold of $\alpha = 0.5$. In Figure 4.4(c), the entire gun is visible and the overlap from the source images is evident. Two dark lines can be seen at the middle of the gun that show where the parts of the source images overlap.

To demonstrate the usefulness of fusion, the IR image pairs were produced by covering part of the concealed weapon so that a different part was visible in each image. For this example, the images used for fusion were obtained from actual IR imaging sensors. Figure 4.5(a) shows that the left side of the gun is visible and Figure 4.5(b) shows that the right side of the gun is visible. The person in the images is holding markers to show the ends of the gun. These markers actually show that the images are not correctly registered. This happened because the images were taken at different times to simulate the situation where complementary parts of an object are visible. In Figure 4.5(c), the whole gun can still be seen although no registration was used before fusion. For the example we also used two decomposition levels and $\alpha = 0.5$.

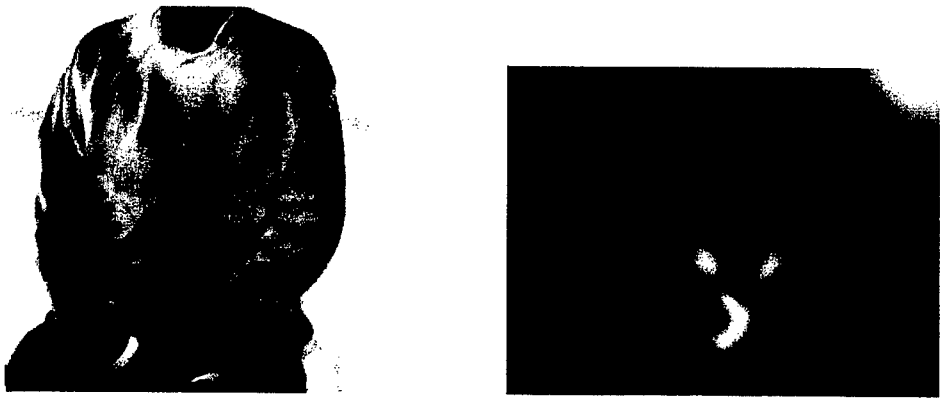
Finally, we fused IR and MMW source images to form a composite image. In these examples, the gun is visible in both images, but at different resolutions and degrees of visibility. Here we show that our fusion algorithm also works well for images from different sensors. For this case, the IR sensor provides better spatial resolution while the MMW sensor provides better clothing penetration. Figure 4.6 shows the original unregistered IR and MMW images. For this example, the images were manually registered before applying the image fusion algorithm. The outline of the person's body is more visible in the IR image and the gun is apparent in both images. The results of fusing the original IR and MMW images and the morphologically pre-processed images are shown in Figure 4.7. Morphological filters of increasing size are used for each set of fused images in Figures 4.7(b)-(d).



(a) (b)

(c)

Figure 4.5: IR image pair: (a), (b) source images and (c) fused result.



(a) (b)

Figure 4.6: Original (a) IR and (b) MMW images.

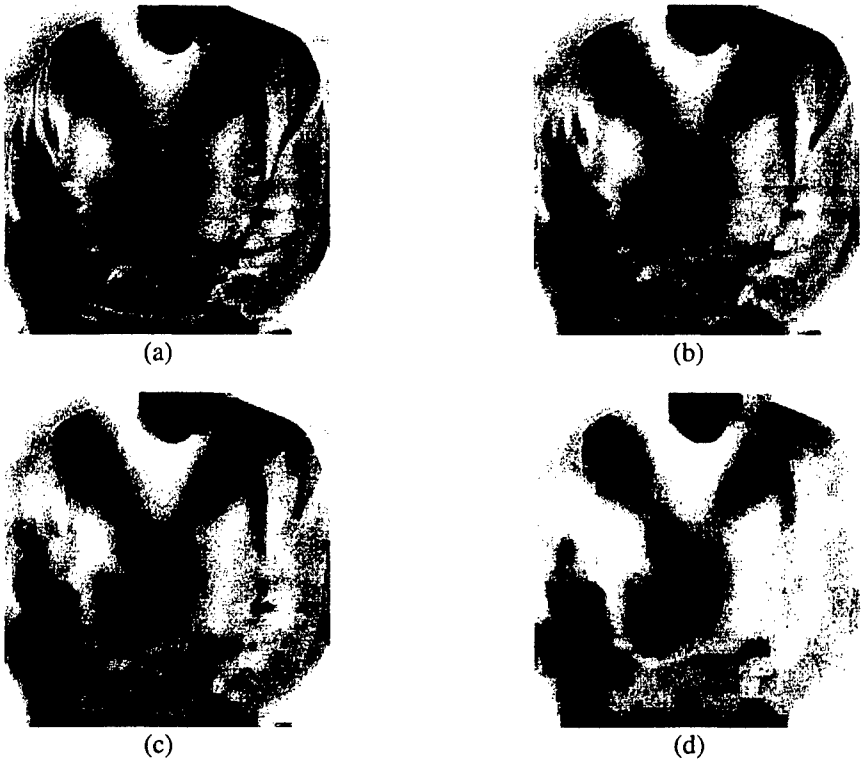


Figure 4.7: Fused images of the (a) original IR and MMW images and (b) filtered IR and MMW images using a 3 by 3 filter, (c) a 5 by 5 filter and (d) a 9 by 9 filter.

Thresholding (using Otsu's method [Ots79]) was used to compare the fused results to each other and to the original IR and MMW images. Figure 4.8 shows the thresholded results for the original unfiltered IR and MMW images without fusion. The IR image shows the gun, but it is also connected to wrinkles on the lower part of the image. In addition, several other wrinkles are visible for the IR image. No shape resembling a gun is apparent in the MMW image.

Figure 4.8: Thresholded results for original (a) IR and (b) MMW images.

Figure 4.9 shows the thresholded results for the fused images with and without preprocessing. The result without pre-processing in Figure 4.9(a) shows the gun shape is separated from the rest of the objects. In Figure 4.9(b), the source images were pre-processed with a 3 by 3 filter. The result shows that fusion improves the ability to segment out the gun shape over the individual filtered images. It also shows a slight improvement in comparison to the result in Figure 4.9(a). Notice that some of the wrinkles and other artifacts have disappeared. Figure 4.9(c) also shows improvement in comparison to the result shown in Figure 4.9(a). More of the wrinkles and other artifacts have disappeared, but the gun shape remains the same. In Figure 4.9(d), the fused image is much improved over the other fused results. We notice that all of the wrinkles appear to be gone and the smaller artifacts have disappeared, but the gun shape remains the same. These results show that fusion, based on these thresholded images, is expected to improve object recognition. In addition, using morphological filtering before fusion helped remove artifacts while leaving the basic gun shape intact.

(a)

(b)

(c) (d)

Figure 4.9: Fused results using the (a) original and (b) filtered images with a 3×3, (c) a 5×5 and (d) 9×9 filter.

5. SHAPE RECOGNITION

In this stage of the CWD system, we assume that the images are acquired from the sensors, registered and fused based on the methods described in the previous sections. Then through a segmentation and extraction algorithm the shapes to be recognized are obtained.

In general, it is not known how the shapes appear in the observed frame. They could be rotated with an angle and/or scaled in size and/or shifted in position. So, in the recognition process these unknown conditions must be handled properly in order to be successful. Therefore, the algorithms to be considered should be rotation, scale and translation invariant. Three algorithms are considered based on these criteria for our system. Below, these algorithms are described in detail, then the recognition process for the objects (as weapon and non-weapon objects) is explained.

5.1 Recognition Algorithms

Recognition algorithms may classify the same shape differently, so our object recognition method uses several different algorithms that determine whether a weapon is present or not present. The algorithms we are using include recognition metrics such as Fourier descriptors, moments and compactness.

5.1.1 Moments

First, regular moments are described for an image, $f(x, y)$, of size N_x by N_y. These moments are calculated using all of the pixels in the image. The $(p+q)$-order moment of the image is defined as

$$m_{pq} = \sum_{x=1}^{N_x} \sum_{y=1}^{N_y} x^p y^q f(x,y)$$

and the centroid of the image, $(\bar{x}, \bar{y})$, is given by

$$\bar{x} = \frac{m_{10}}{m_{00}} \qquad \bar{y} = \frac{m_{01}}{m_{00}}.$$

Moments of Region Boundaries:

In the next step, we define a different type of moment descriptor. Instead of using all the pixels of the shape of interest, we only consider the sequence of contour pixels (boundary pixels) of the shape. Let the coordinates of the N contour pixels of the object be described by an ordered set $(x(i), y(i))$, $i = 1,2,...,N$. Next we compute the Euclidean distance between the centroid, $(\bar{x}, \bar{y})$, and the ordered sequence of the contour pixels of the shape. Let us denote this set of Euclidean distances as $d(i)$, $i = 1,2,...,N$. This set forms a single-valued, one-dimensional, and unique representation of the contour. Based on the set $d(i)$, the p-th moment is defined as [GS87]

$$m_p = \frac{1}{N} \sum_{i=1}^{N} [d(i)]^p$$

and the p-th order central moment is defined as

$$M_p = \frac{1}{N} \sum_{i=1}^{N} [d(i) - m_1]^p .$$

Using these moment definitions, a feature set (F_1, F_2) is described by

$$F_1 = \frac{(M_2)^{1/2}}{m_1} \qquad \text{and} \qquad F_2 = \frac{(M_4)^{1/4}}{m_1}.$$

Note that these values are dimensionless, and rotation, scale, and translation invariant.

The first algorithm, which is based on moments techniques, uses the quantity (F_2-F_1) as the feature information of the shape which provides very good distinctive information for the roughness of the shape [SRD94].

5.1.2 Fourier Descriptors:

In this approach, again the boundary pixels are used to extract the feature information. Suppose that there are N points on the contour of a shape (region). The *x-y* coordinates of each point in the contour are represented as a complex number in a complex plane. The ordered contour sequence may then be written as a complex sequence Z_i

$$z_i = x_i + j\,y_i \qquad i = 0,1,2,\ldots,N-1.$$

Note that the complex number sequence is periodic with each transversal of the complete boundary.

The Fourier descriptors (FD) are defined as

$$B(k) = \frac{1}{N}\sum_{i=0}^{N-1} z_i \exp[-j2\pi ki / N] \qquad k = 0,1,\ldots,N-1$$

Before using the FDs for shape analysis, we need to eliminate their dependence on orientation, size, position, and starting point of the contour. Based on the following properties [GW87], the Fourier descriptors are modified accordingly ;

1) A change in the position of the contour (translation) changes the value of $B(0)$ only. So, by setting $B(0) = 0$, the FDs become translation invariant.

2) To change the size of the contour (scaling) the z_is need to be multiplied by a scalar. This is equivalent to multiplying $B(k)$s by a scalar. So if the FDs are divided by $B(1)$ then they become normalized and independent of scaling.

3) Rotating the contour in the complex plane is equivalent to multiplying each coordinate by $exp(j\theta)$ where θ is the angle of rotation. Due to linearity of the Fourier transform, it results in multiplying the frequency domain coefficients, $B(k)$, by $exp(j\theta)$.

Shifting the starting point of the contour in the complex plane corresponds to multiplying the k-th frequency coefficient, $B(k)$, in the frequency domain by $exp(jkT)$, where T is the fraction of a period through which the starting point is shifted. Note that as T goes from 0 to 2π, the starting point traverses the whole contour once.

So for the analysis, if only the magnitudes of FDs are used, then it becomes independent of orientation (rotation) and starting point of the contour. This is because the phase change in the coefficients does not affect the magnitude of FDs. Hence, the normalized FDs, (NFD), are obtained as

$$NFD(k) = \begin{cases} 0, & k = 0 \\ B(k) / B(1), & k = 1,2,...,N/2 \\ B(k+N) / B(1), & k = -1,-2,...,-(N/2)+1 \end{cases}$$

The second algorithm, which is based on Fourier Descriptors, uses FDM as a measure of the feature information of the shape that is defined as

$$FDM = \left[\sum_{k=-\frac{N}{2}+1}^{N/2} \| NFD(k) \| / |k| \right] \Bigg/ \sum_{k=-\frac{N}{2}+1}^{N/2} \| NFD(k) \|$$

where $\|\cdot\|$ is the norm. In the measure above, the division by $|k|$ reduces its sensitivity to high frequency noise (which appears in the shapes with rough boundaries) and the denominator normalizes the quantity and limits its range within 0 to 1.

5.1.3 Compactness:

A dimensionless measure of shape compactness, C, is defined as [GW87],

$$C = \frac{P^2}{A}$$

where P is the length of the region perimeter and A is the area of the region. Compactness provides a measure of contour complexity versus area enclosed. A shape with a rough contour including several incursions will have a high value of C, indicating low compactness. It is clear that this quantity is independent of rotation, scale and translation.

It should be noted that these shape factors measure the roughness of shapes with different concepts. Therefore, by using these three measures, the robustness of the shape recognition procedure will increase because the information content in each one supplements that of others.

5.2 Test Procedure

In the first stage, for the training purpose of the shape recognition system, several shapes are extracted from the known weapon and non-weapon images. Each shape is run through the three algorithms described above and three dimensionless numbers are obtained (one from each algorithm) from which a three dimensional vector is formed. So, each shape, whether it belongs to a weapon or a non-weapon object, is represented by a three dimensional vector.

The vectors that are obtained from the known weapon shapes are grouped in one reference library (Library1). The vectors from the known non-weapon shapes are grouped in another reference library (Library2). For the known weapon shapes, pictures of different kind of handguns are digitized (see Figure 5.1) and the corresponding reference Library1 data are obtained by running the shape characterization algorithms on these images. For the known non-weapon shapes, basic synthesized geometric shapes are considered. These are circles, squares and rectangles (with different side ratios)

in several sizes and rotated positions. Reference Library2 data are obtained from these synthetic shapes in the same way as described for Library1 data.

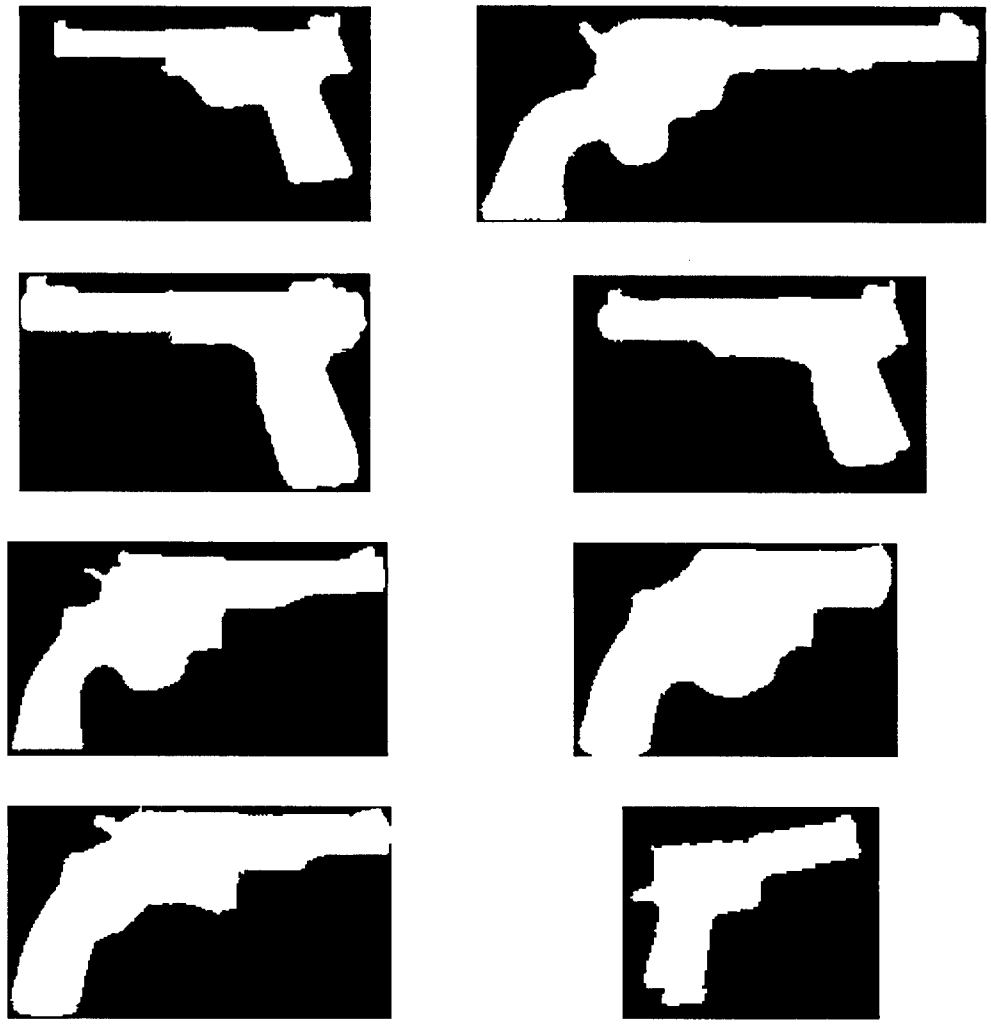


Figure 5.1. Typical shapes in the weapon library (Library1).

In the second stage, on a new extracted shape image with unknown origin (i.e., with no prior information whether it is a weapon or non-weapon shape) the three algorithms are run and its shape characterization values are obtained as a three dimensional vector. Note that since each shape was characterized by a three-dimensional vector, it can be represented as a point in a three dimensional space. Hence, the recognition of this new shape is achieved by applying the nearest-neighbor method

in this three-dimensional space. The normalized Euclidean distances between the new sample and each sample of the reference libraries are computed. The shortest distance determines the class of the new image. That is the new image is assigned to the same class (library) as of the one which is the closest to it among all reference library vector points.

The Normalized Euclidean distance (NED) between shape characterization vectors $\overline{X}_1$ and $\overline{X}_2$ is defined as

$$NED = \sum_{i=1}^{3} (x_{1i} - x_{2i})^2 / m_i^2$$

where $\overline{X}_j$ and m_i are defined by

$$X_j^t = (x_{j1}, x_{j2}, x_{j3})$$

where t is for transpose and

$$m_i = \frac{1}{N} \sum_{j=1}^{N} (x_{ji}).$$

The parameter N is the total number of vectors in the reference libraries.

As a test of system performance, leave-one-out [DK82] method was used to estimate the probability of error of the classifier. In the literature, it is recommended as the best method to evaluate the efficiency of a classifier using a given set of samples with known classes [DK82, Fuk90]. In leave-one-out algorithm one sample (vector point) at a time from the reference libraries is left out and used as a new unknown sample while the rest is used as reference library samples to classify this sample based on the nearest-neighbor method. This is repeated for all reference library samples and the ratio of the number of wrong classifications to the number of all performed classifications (equals the number of all samples in the reference libraries) gives the probability of error of the classifier.

In the leave-one-out test, the recognition system based on the three algorithms achieved 100% correct decision (classification). This result is obtained for the available library samples. Although it is a good result, testing the system with the sample images acquired from real sensor outputs will help us confirm that the system will perform well with real life data, too. If it is needed the procedure can be modified further by adding and/or removing algorithms.

6. CWD EXAMPLE

The entire CWD process of Figure 1 is demonstrated here by an example. The original IR and MMW images are shown in Figure 6.1. Figure 6.2 shows the images after registration, filtering and fusion. The images were registered using the algorithm of Chapter 2. After registration, each source image was morphologically filtered using a 5×5 structuring element to remove some of the unwanted details. The filtered images required additional cropping to allow for wavelet decomposition and reconstruction by powers of two. For fusion, two levels were used for the wavelet decomposition and reconstruction with α=0.5.

Figure 6.1. Original IR and MMW images.

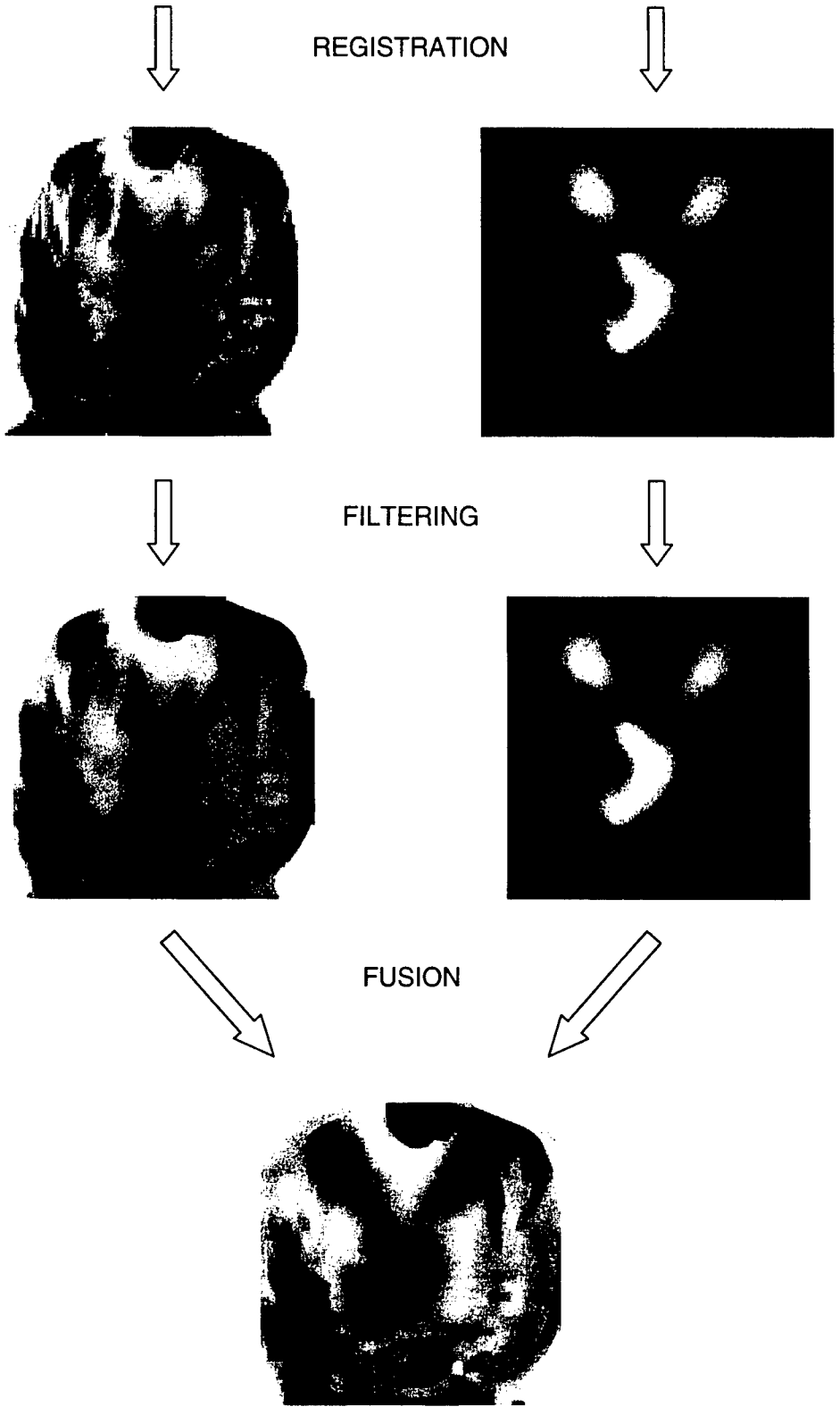


Figure 6.2. Registration, filtering and fusion of original image pair

The result of thresholding the fused image is shown in Figure 6.3 with the gun shape clearly visible. Also, the shape descriptors are calculated for the gun shape and compared to the other shapes in the weapon and non-weapon libraries. A three-dimensional plot is also shown in Figure 6.3 where the gun shape is denoted by an '×', the non-weapon library points are represented by 'o', and the weapon library points are represented by '+'. From this plot it is clear that the gun shape is closest to the points in the weapon library and that it will be classified as a weapon.

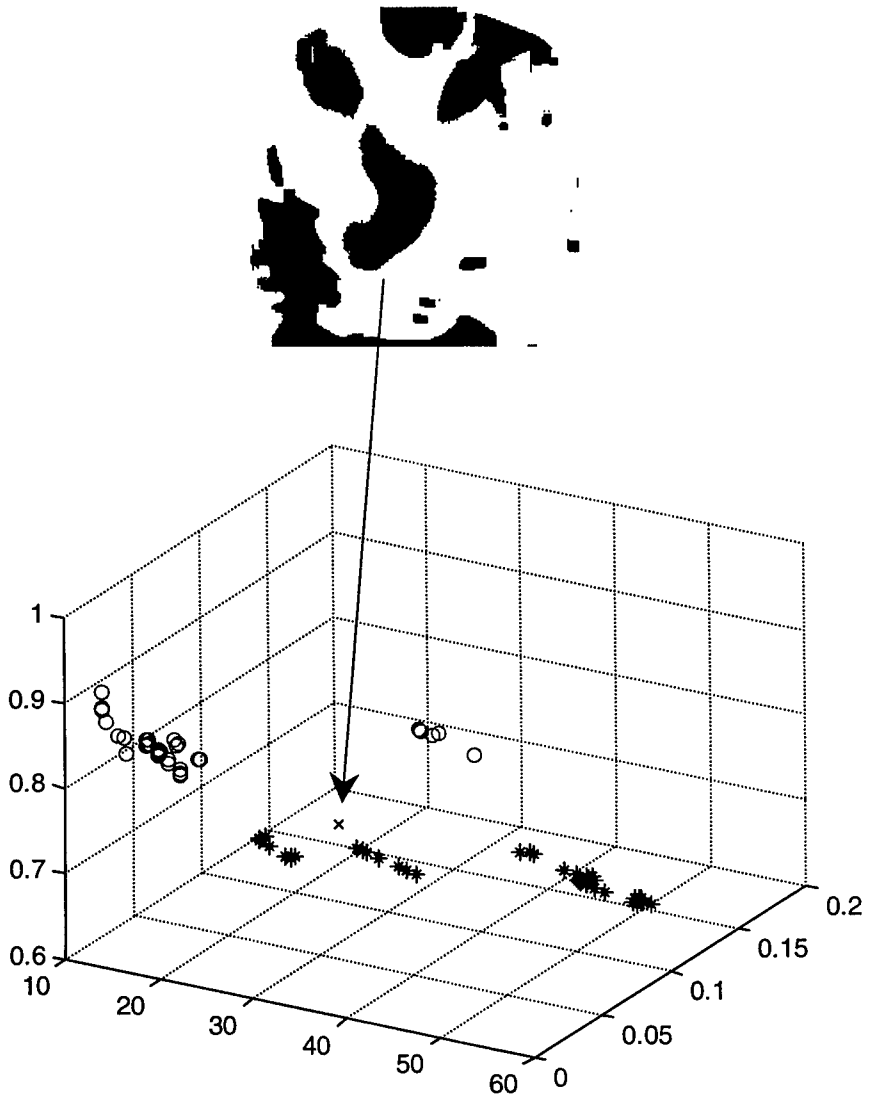


Figure 6.3. Thresholding and shape recognition.

7. IMAGE FUSION FOR WIRELESS IMAGE TRANSMISSION

Before an image is transmitted over a wireless channel, it is desirable to implement a method for representing the image that is resilient to channel errors. In addition, wireless transmission must also meet bandwidth constraints that require compressed image data for most practical situations. Here we will consider both uncompressed and compressed image data for transmission over bursty channels. For an error resilient representation, wavelet based decomposition will be implemented for transmitting the image in its uncompressed state. The compressed image will be represented using the SPIHT algorithm of Said and Pearlman [SP96] without arithmetic encoding. During transmission, the image will be subject to bursty channel errors. Therefore, a technique is needed at the receiver to correct or conceal any errors that may degrade the perceptual quality of an image beyond acceptable limits. The goal of this research is to introduce a novel image transmission method based on diversity combining that can produce an image of high perceptual quality at the receiver.

Several methods for image transmission over wireless channels have been proposed in the literature recently. Most of these methods have been proposed for transmission of compressed images based on the discrete cosine transform (DCT) or the wavelet transform. These image transmission methods can be divided into four categories: error correction/detection methods, error resilient image coding, error concealing techniques, and hybrid methods. The error correction/detection methods [HL98, KY95, KCR98, FK97, VH96] utilize automatic repeat request (ARQ) schemes or different types of error coding to combat the errors introduced during wireless transmission. The error resilient image coding techniques [SLCP96, JN95, MGTH95, Hem96, BA98]

make the transmitted data stream less sensitive to wireless channel errors. Therefore, when the data arrives at the receiver it is easier to recover from errors in the bit stream. Therefore, errors in these types of image data will not affect image quality as much as other image coding schemes such as JPEG. Error concealment schemes [WPE96, LR95, ZL95] attempt to hide errors that degrade the perceptual quality of the image after the image has been received. The proposed hybrid methods [CF98, KR97, CRSZ98, LLL97, WZ95, Iun98] use a combination of the three techniques described previously.

Diversity is a communication method used to improve wireless transmission that utilizes independent (or highly uncorrelated) communication signal paths to combat channel noise. The independent signal paths provide the receiver with multiple signals for appropriate diversity processing of the received signals. The types of diversity typically used for wireless communications include spatial, frequency and time diversity methods. Space or antenna diversity works by having spatially separated antennas at the receiver to obtain the independent or uncorrelated signals. Frequency diversity involves transmission of data on multiple carrier frequencies to get uncorrelated fading channels, but a disadvantage of this method is the need for extra bandwidth. Time diversity retransmits information at time intervals that allow for independent fading conditions. In all of the above diversity methods, multiple independent (or highly uncorrelated) signals are available at the receiver that need to be combined to generate the received information. Selection diversity is one simple example of diversity combining that takes the signal from the diversity branch with the highest SNR. Other common methods for diversity combining are equal gain combining and maximal ratio combining. All of these methods carry out diversity combining in the data domain in that they attempt to obtain the best estimate of the received digital data. For image transmission, a diversity technique has been employed in conjunction with ARQ [WZ95]. This approach involves switched antenna diversity that operates in the data domain.

Unlike the data domain diversity combining methods mentioned above, the diversity

combining method we propose here operates in the image domain by using the properties of the original image or its wavelet transform. Our novel approach to wireless image transmission combats the effects of fading and other channel impairments by employing a diversity combining method that attempts to directly improve image quality. This diversity combining method was inspired by the image fusion work of Burt [BL93] where he produced one composite image from multiple source images with different information content. Burt implemented his fusion method by taking a Laplacian pyramid transform of each source image, combining the transforms based on measures in the transform coefficient neighborhoods, and performing the inverse transform to obtain the composite image. Later, Li et al [LMM95] used this same image fusion methodology but with the wavelet transform. For image transmission over wireless channels, two or more diversity channels can be utilized to obtain multiple bit streams at the receiver, with each bit stream independently representing the image data. Then these bit streams can be combined in the image domain to improve the perceptual quality of the received image. Due to the random nature of radio propagation, we expect the errors on the individual channels to be independent or at least highly uncorrelated. The independent nature of the diversity channels allows for a combining method in the image domain that yields excellent quality images in the presence of wireless channel errors.

7.1. Channel Model

It is well known that a wireless transmission channel is corrupted by errors that are bursty in nature. Errors in a wireless channel occur because of physical phenomenon such as fading and multipath. Modeling of the physical channel is a complex problem that also depends upon the movement of the transmitter, receiver, and other objects in the signal path. While a number of models that characterize the physical phenomena have been proposed in the literature, here we employ an input-output channel model. This class of models attempts to represent the input-output relationships of the wireless channel [KS78]. They treat the channel as a black box and model the discrete error stochastic process. This class of models was used extensively for the evaluation of modems and

coding schemes. Here, we use such a model to generate error sequences that are typical of the wireless channels they represent.

One popular input-output error model is in terms of a finite state Markov chain. In this model, each state represents a different channel condition and the associated error behavior. These models are specified in terms of transition probabilities between the individual states and the corresponding error probability for each state. Two error statistics that can further describe these types of channels are average error rate and average burst error rate. The average error rate is the proportion of errors to the total number of transmitted bits and the average burst error rate is the time spent in the bad state. The model we use for our simulations is a two-state Gilbert-Elliott channel [Gil60, Ell63].

The two-state Gilbert-Elliott channel has one good state and one bad state, represented by 0 and 1 respectively as shown in Figure 7.1. This channel can also be described by its burst error length and error rate parameters, which are related to the transition probabilities between states and the error probabilities of the individual states. While in the good state the bits are transmitted incorrectly with probability $P_e(0)$, and while in the bad state the bits are transmitted incorrectly with probability $P_e(1)$. For this model it is assumed that $P_e(0) << P_e(1)$. The two-state channel model can

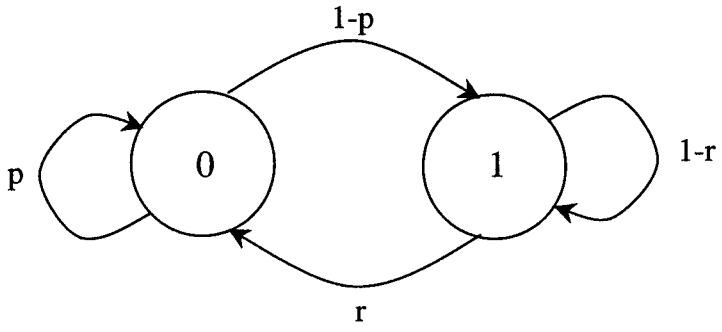


Figure 7.1. Two-state Gilbert-Elliott channel

be described by the binary Markov process γ_n with the following transition matrix:

$$\mathbf{P} = \begin{bmatrix} P(\gamma_n = 0|\gamma_{n-1} = 0) & P(\gamma_n = 1|\gamma_{n-1} = 0) \\ P(\gamma_n = 0|\gamma_{n-1} = 1) & P(\gamma_n = 1|\gamma_{n-1} = 1) \end{bmatrix} = \begin{bmatrix} p & 1-p \\ r & 1-r \end{bmatrix} \tag{7.1}$$

where γ_n=0 if the channel is in the good state at time n, and γ_n=1 if the channel is in the bad state at time n. The average burst length L is a geometric random variable with mean $1/r$, and the average time the channel is in the good state is also a geometric random variable with mean $1/(1-p)$. The steady state probability of the channel being in a bad state is $\pi_1 = (1-p)/(r+1-p)$. Also, the steady-state error rate is [YW95] given as $\varepsilon = (P_e(0) \cdot r + P_e(1) \cdot (1-p))/(r+1-p)$. This model will be used to generate errors to corrupt the images in our simulations to evaluate the performance of our diversity combining method.

7.2. Diversity Combining Method for Uncompressed Images

Our diversity combining method for uncompressed images involves computing the two-dimensional wavelet decomposition of the source image and quantizing the resulting wavelet coefficients. The coefficients are then transmitted as a bit stream over a wireless communications system employing diversity without any error control. Diversity is used to obtain multiple copies of the decomposed image data at the receiver. At the receiver, the individual decomposed images are combined to form a composite wavelet decomposition and then the final received image is reconstructed. This diversity combining method is depicted in Figure 7.2.

The first step in wireless image transmission is to consider how the image will be represented for transmission. The two-dimensional wavelet decomposition of an image is implemented with traditional subband filtering [WO86] using one-dimensional low-pass (H) and high-pass (G) quadrature mirror filters. First, the input image is convolved with H and G in the horizontal direction and then the output rows are down-sampled by two. Then the two resulting sub-images are further filtered along the vertical direction followed by down sampling of the columns. At the output, the

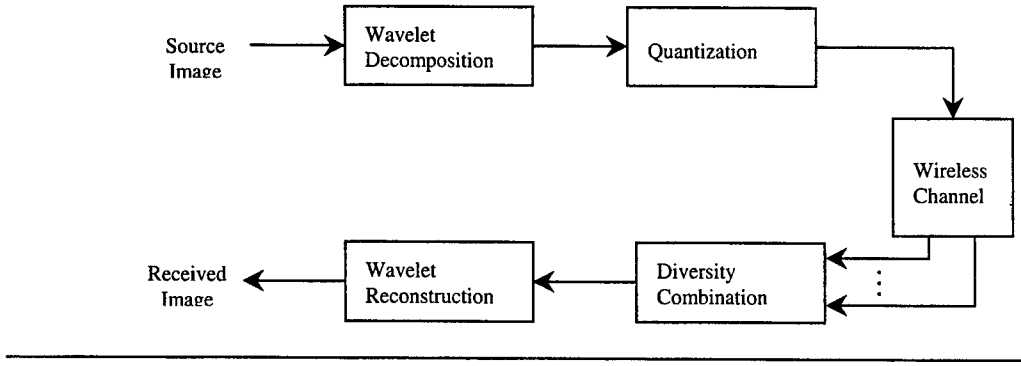


Figure 7.2. Diversity combining for uncompressed images.

source image at resolution k is decomposed into four sub-images: an image at lower resolution level k-1, a horizontally oriented detail image, a vertically oriented detail image, and a diagonally oriented detail image. The filtering can be repeated by using the low-resolution image as the source image until the desired decomposition level is reached. The image at resolution k is reconstructed from the four sub-images at resolution k-1 using reconstruction filters $\underline{H}$ and $\underline{G}$. The rows are up-sampled by two (one row of zeros is inserted between each row) and filtered in the vertical direction. Then the same procedure is followed in the horizontal direction. At the output, a reconstructed image at resolution k is obtained. Repeating the same procedure, the original level at which the decomposition was started can be reached.

In this section, we use images transformed in the wavelet domain with uniform scalar quantization of the coefficients. The results obtained will help demonstrate the usefulness of image domain diversity combining for image transmission over wireless channels. For images without compression, the wavelet representations are obtained from the bit streams received on the individual diversity channels. In general, the low-resolution subband is more important perceptually and a large error in pixel intensity can seriously affect image quality. An error in the high frequency subband is not as important to the overall image quality. Because the characteristics of the subbands are different, the diversity-combining rule for the low-resolution subband differs from the combination rule for the high frequency subbands. After obtaining the composite decomposed image from fusing the individual transformed images, the inverse wavelet transform is performed to obtain the final image.

The idea behind diversity combination is to significantly reduce visible errors in the received image without necessarily using techniques such as ARQ or error correction coding. The diversity combining method is demonstrated here using two independent channels, channel one and channel two, but the idea can easily be extended to more channels. When the bit streams containing the decomposed images are received, a decision is made as to whether to take the data from channel one, channel two, or from a combination of both. Depending upon the channel state the two received bit streams will contain the same values for many of the coefficients.

The low frequency subband and high frequency subbands have different sensitivities to bursty channel errors. Therefore, the rules for the two types of subbands are different. For both of the different subband types there are two combination modes: selection and coefficient combining. In the selection mode, one coefficient is selected from the two decomposed images and placed in the composite. In the coefficient-combining mode, groups of coefficients from neighborhoods of both decomposed images are examined and a value is placed in the composite decomposed image based on measures from both coefficient neighborhoods. The combination method is similar to using both image averaging and spatial filtering to remove channel noise.

Since the low-resolution subband is more perceptually important to the image, more care must be taken when dealing with detected channel errors in the low-resolution subband. First, the coefficients from the two diversity bit streams are compared as they arrive at the receiver. If the received wavelet coefficient values are the same, we assume that the value is correct and select the coefficient from either channel to place in the combined transform. If the coefficient values are different, the receiver waits until an m by n neighborhood of coefficients surrounding the coefficient of interest is available from both channels. Small neighborhoods (i.e. 3 by 3) of an image are generally smooth. Therefore, the intensity values usually do not vary significantly within these neighborhoods. When the two received coefficients at location (i, j) are different, the m by n neighborhoods of coefficients around them are grouped into a set of $2mn$ values. Then the median

value is chosen as the coefficient to place in the combined low-resolution sub-image at location (i, j). In general, this median-based method tends to be more robust to large channel errors than averaging the coefficients in order to obtain a combined coefficient value. Therefore, for each (i, j), the coefficient placed in the combined low resolution subband image is defined as follows (assuming m and n are odd):

$$c_{Lc}(i,j)=\begin{cases} c_{L1}(i,j) & \text{if } c_{L1}(i,j)=c_{L2}(i,j) \\ med\left[\{c_{L1}(k,l)\},\{c_{L2}(k,l)\}\right] & \text{if } c_{L1}(i,j)\neq c_{L2}(i,j) \end{cases} \tag{7.2}$$

for $(k,l)\in\left(\left\{i-\frac{m-1}{2},\ldots,i+\frac{m-1}{2}\right\},\left\{j-\frac{n-1}{2},\ldots,j+\frac{n-1}{2}\right\}\right)$ where c_{Lc} represents the wavelet coefficients in the low-resolution subband of the combined transform, and c_{L1} and c_{L2} are the low-resolution coefficients obtained from two diversity channels.

An error in the high frequency subbands does not affect the quality of the final reconstructed image as much as in the low frequency subbands. Also, most of the coefficients have magnitudes close to zero. Therefore, the errors in the detail subbands are processed differently when the received wavelet coefficients are not the same. Again, if the received wavelet coefficient values are the same, we assume that the value is correct and place this value in the combined transform. However, if the received coefficients are different, the coefficient with the minimum absolute value is chosen and placed in the final combined transform. The idea behind this selection method is that a coefficient that implies a strong edge where one does not exist will visually degrade the image more than a coefficient that implies no edge where one really exists. Since most of the coefficients in the high frequency subbands are near zero, there is a better chance that the coefficient with the minimum absolute value will be correct. Even if we set the coefficients to zero in the high frequency subbands, the quality of the final image will still be acceptable. The combined coefficient values for each location (i,j) in the high frequency subbands are given as follows:

$$c_{Hc}(i,j)=\begin{cases} c_{H1}(i,j) & \text{if } c_{H1}(i,j)=c_{H2}(i,j) \\ c_{H1}(i,j) & \text{if } |c_{H1}(i,j)|<|c_{H2}(i,j)| \\ c_{H2}(i,j) & \text{if } |c_{H2}(i,j)|<|c_{H1}(i,j)| \end{cases} \tag{7.3}$$

where c_{Hc} represents the wavelet coefficients in the detail subbands of the combined transform, and c_{H1} and c_{H2} are the detail subband coefficients obtained from two diversity channels.

In order to show the feasibility of using diversity combination for wireless image transmission, simulations were performed using uncompressed images. The results are compared to a system that uses error control coding for error protection. In our experiments, images were transmitted using a BCH(255, 179) code with error correction capability of 10 bits. For each simulation, two bit error patterns were generated using the two-state Markov model described in Section II. Both error patterns were applied to the image data bit streams for the diversity combination method and one of the error patterns was used for the error coding method. The parameters used for generating the bit error patterns were an average burst error length of 500 bits and various bit error rates (.0001, .0005, .001, .005, .01). The error probabilities within the individual states were set to $P_e(0) = 0.0$ and $P_e(1) = 0.5$. Performance is measured using peak signal to noise ratio (PSNR):

$$PSNR = 10\log_{10}\frac{255^2}{\frac{1}{N}\sum_i\sum_j (p(i,j)-\hat{p}(i,j))^2}, \tag{7.4}$$

where $p(i,j)$ are the pixel values of the original image and $\hat{p}(i,j)$ are the pixel values of the received image.

For our simulations we tested our diversity combining method on the two images shown in Figure 7.3. Both are 8-bit graylevel images with 256 by 256 pixels. First, the source images were decomposed to two levels using the wavelet transform. Then the wavelet coefficients were uniformly quantized to 8 bits per pixel in order to maintain the same number of bits as in the original image.

But for the bit stream with BCH coding, the total number of transmitted bits is greater than 8 bits per pixel. For uncompressed images we did not attempt to match bit rates for performance comparisons. The given PSNR results were averaged over twenty runs.

(a) (b)

Figure 7.3: Original test images for wireless image transmission: (a) Peppers and (b) Lenna.

Table 7.1 gives the PSNR results for the Peppers image using our diversity combination method versus BCH(255, 179). In this table, we see that the PSNR results for diversity combining are about 11 to 13 dB higher than error coding. Table 7.2 gives the PSNR results for the Lenna image using our diversity combination method versus BCH coding where the diversity method exceeds the error coding by about 15 to 16 dB. Examples of the received images are shown in Figures 7.4 and 7.5 for bit error rates of 0.0005, 0.005 and 0.01. These examples demonstrate that the diversity combination method provides images with significantly improved perceptual quality compared to using BCH error correction coding.

Table 7.1: PSNR (dB) for Peppers

BIT ERROR RATE	DIVERSITY COMBINING	BCH(255, 179)
.0001	31.5507	20.2975
.0005	31.8934	19.9742
.001	32.2498	20.8729
.005	32.8253	20.3403
.01	30.2323	17.0615

Table 7.2: PSNR (dB) for Lenna

BIT ERROR RATE	DIVERSITY COMBINING	BCH(255, 179)
.0001	34.1783	20.7307
.0005	35.5003	20.8015
.001	35.5044	19.6481
.005	35.0813	20.3599
.01	33.0693	17.5477

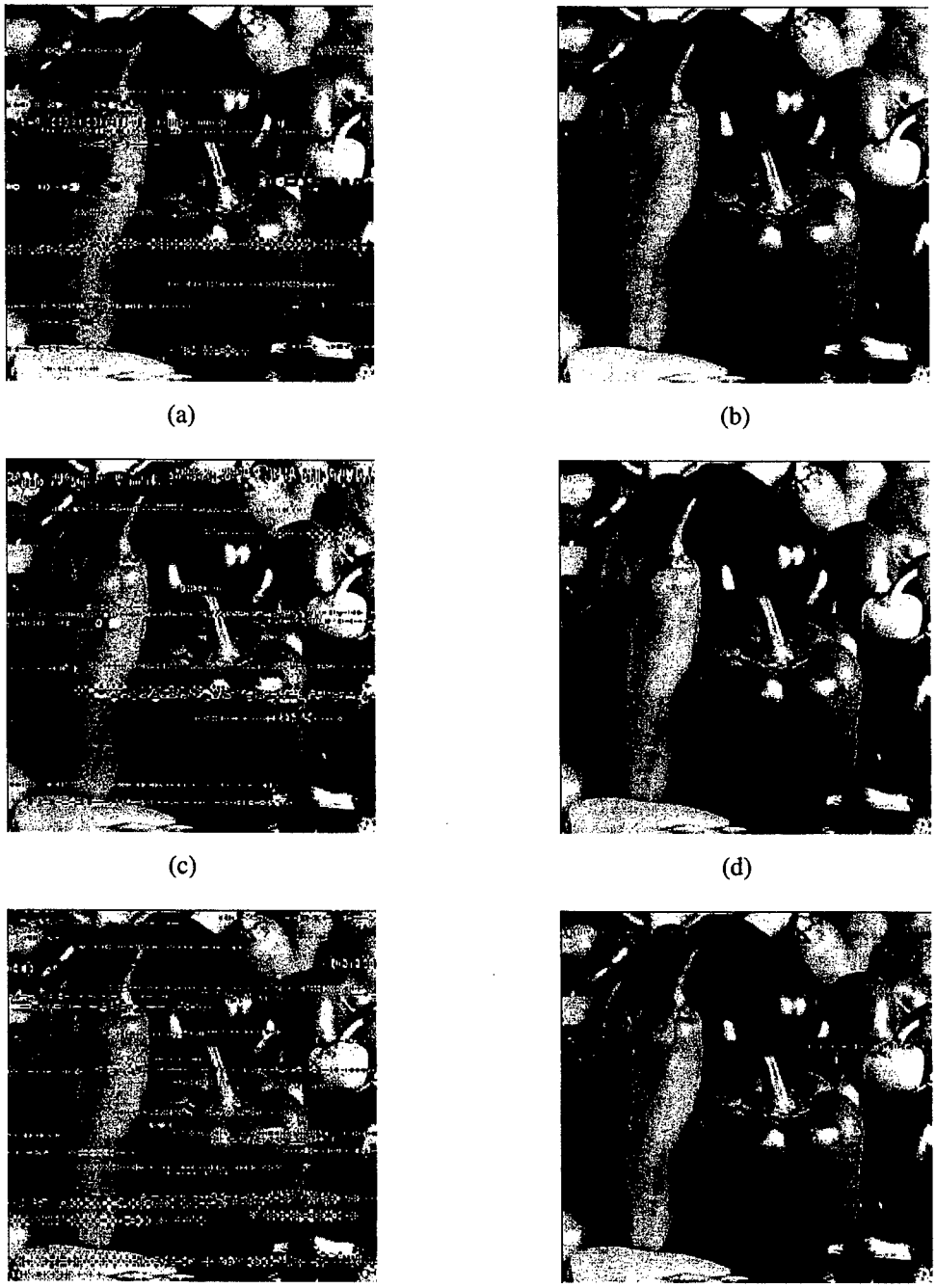


Figure 7.4. Results for Peppers image with BER = 0.0005 (a) BCH, (b) diversity combining; BER = 0.005 (c) BCH, (d) diversity combining; BER = 0.01 (e) BCH, (f) diversity combining.

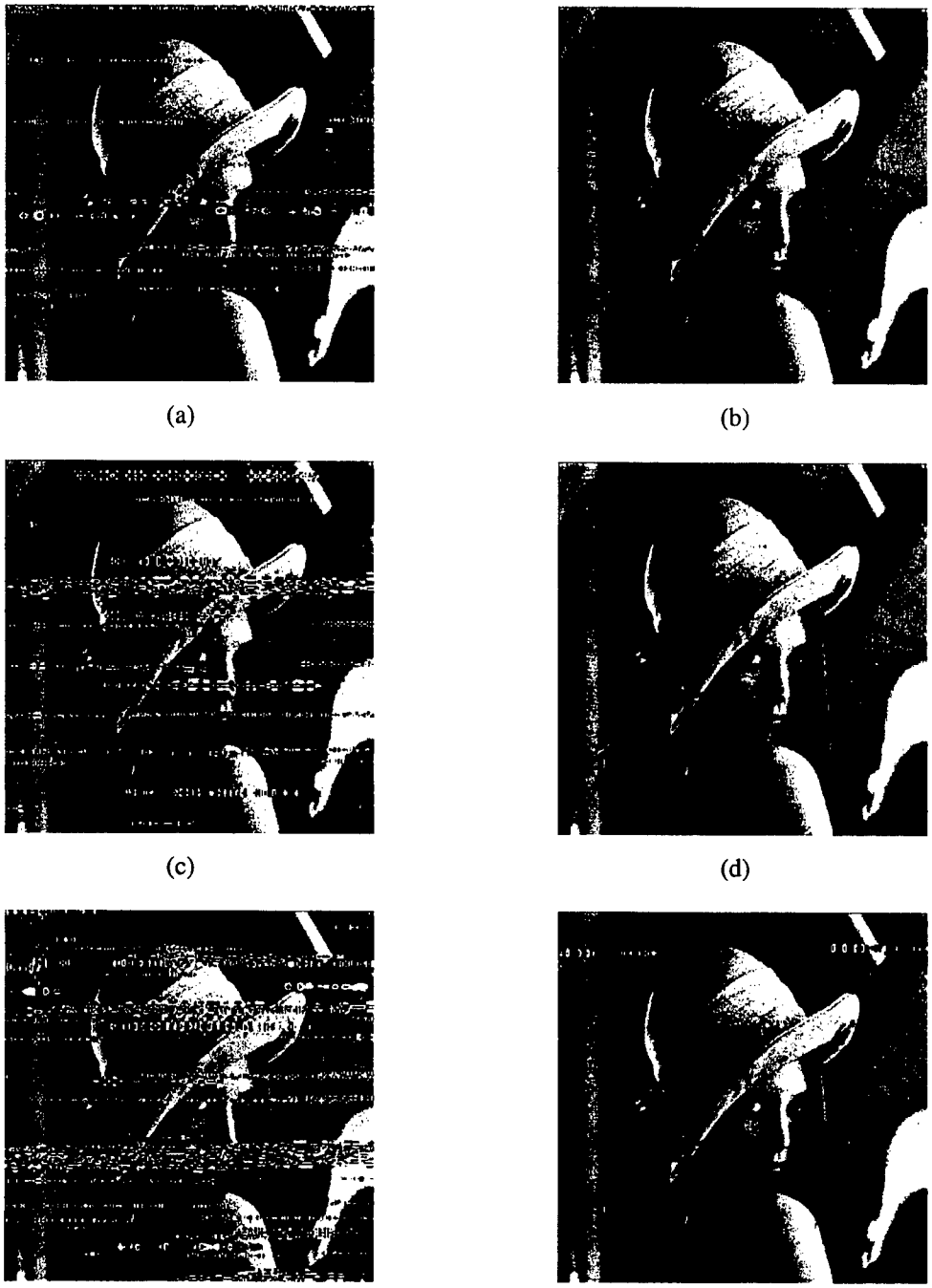


Figure 7.5. Results for Lenna image with BER = 0.0005 (a) BCH, (b) diversity combining; BER = 0.005 (c) BCH, (d) diversity combining; BER = 0.01 (e) BCH, (f) diversity combining.

7.3 Diversity Combining Method for Compressed Images

Our diversity combining method for compressed images is implemented using the SPIHT image compression algorithm without arithmetic coding. The compressed image data is first protected with interleaving and error control coding and then transmitted over a wireless communications system using diversity. Multiple bit streams representing the image data are obtained at the receiver. Then these multiple bit streams are decoded using appropriate channel decoding algorithms followed by image decompression based on the SPIHT method up to the point where the wavelet representations of the multiple images are obtained. Before computing the inverse wavelet transform, the individual wavelet representations are combined using rules based on wavelet transform characteristics. After diversity combining, a composite wavelet representation is obtained and the received image is reconstructed by performing the inverse wavelet transform on this composite representation. A block diagram of this diversity combining process is shown in Figure 7.6.

The output bit stream of the SPIHT image compression algorithm consists of three types of data: 1) the header information, 2) the sorting information, and 3) the refinement information.

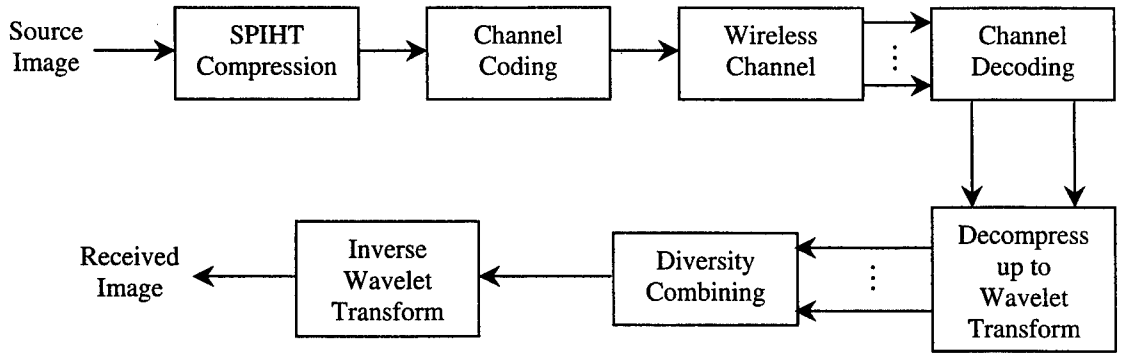


Figure 7.6. Diversity combining for compressed images.

Because the SPIHT algorithm uses variable length coding, one error can result in loss of synchronization and an unrecoverable image. The header information is a small portion of the data but most of the bits need to be received correctly so that the image can be recovered. Also, the sorting information bits must be received correctly (with the exception of the sign bits) in order to avoid a loss of synchronization and the resulting inability to reconstruct the image. The refinement bits do not contribute to any coefficient location information, so an error in these bits will not affect the ability to reconstruct the image. If a refinement bit is in error it may cause a perceptual error in the received image. However, these errors can be corrected or concealed in the final received image.

Another characteristic of the SPIHT bit stream is that the bits at the beginning are more important than the bits at the end. The majority of the bits at the beginning of the SPIHT bit stream tend to determine the coefficient placement and value of the most significant bit. These bits should be highly protected in order to insure that the received image can begin to be reconstructed. Toward the end of the SPIHT stream, there are still some bits that determine coefficient placement, but these bits contribute less to the perceptual quality of the image. Therefore, the bits at the beginning will be protected with a more powerful error correction code than the bits at the end of the bit stream. We employ interleaving and multiple channel codes for error protection of different portions of the SPIHT stream. Similar ideas for error protection have been used in [LR95, KR97, SZ98].

We will describe the diversity combination process assuming two diversity channels but these rules can be extended to more than two channels. In our simulation examples, the error correction scheme first interleaves the binary data stream by rows at a depth of 50 bits and groups them into blocks, then three different Reed-Solomon (RS) codes with different error correction capabilities are used to protect the data. The beginning of the image bit stream is protected with a more powerful code than the bits at the end. The first 7500 bits are grouped into two blocks of data and each block is protected with a RS(31, 15, 8) code down the columns. The next 11500 bits are also grouped into two blocks of data and each block is protected with a RS(31, 23, 4) code down the columns. The

final two blocks of data includes 35400 bits that are protected with a RS(63, 59, 2) code. This error protection scheme results in the transmission of 0.2075 source bits per pixel and 0.2625 bits per pixel overall (source plus channel coding), and therefore, the efficiency is about 80 percent. Figure 7.7 shows our coding method with interleaving and RS channel coding for a given data block. The blocks are transmitted by rows, with all of the data bits in the block sent before the error control bits. Therefore, if the two blocks received from the individual diversity channels have the same bits we can assume no errors occurred on either channel. In this case, no channel decoding is necessary and the error control bits can be discarded. This is similar to the diversity combining rule employed in the image domain and here we use it in the data domain also.

After the data streams are decoded and the associated individual wavelet transforms are obtained, the transforms are combined. Like the rules for the uncompressed images, the combination rules are dependent on which subband is being processed. The actual combination rules for the SPIHT bit streams are slightly different from those rules in equations (7.2) and (7.3). An error in one of the sorting bits will contribute to random errors in the image, unlike the uncompressed images

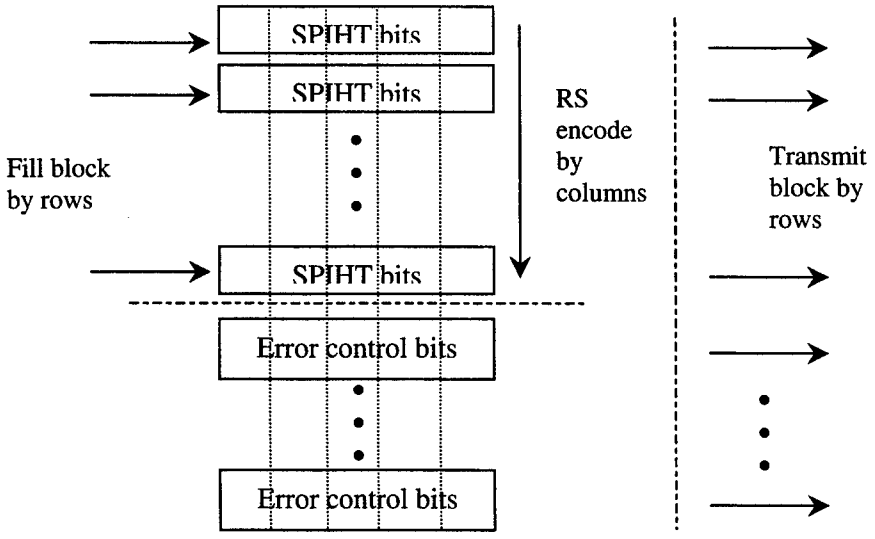


Figure 7.7. Interleaving and encoding scheme for a given data block.

where the errors were generally isolated to certain areas in the images. The diversity combination rule for the low-resolution subband keeps the coefficient value at position (i, j) if the values are the same in both transforms. If the coefficient values are not equal, one of two types of error occurred at that location. Either 1) there was an error in the sorting pass of the SPIHT algorithm where one coefficient was assigned the wrong number of bits, or 2) there was an error in the refinement pass where bits other than the most significant bit were in error. A threshold value, α_1, is set to determine which type of error occurred. If the coefficient values are not the same and the difference is greater than the threshold, it is assumed the first type of error occurred, so one of the values is chosen based on the values of neighboring coefficients. But if the difference is less than α_1, it is assumed the second type of error occurred and the average of both coefficients at the same location is placed in the composite transform. The diversity combination rule for the low-resolution subband is as follows:

$$c_{Lc}(i,j)=\begin{cases} c_{L1}(i,j) & \text{if } c_{L1}(i,j)=c_{L2}(i,j) \\ \left(c_{L1}(i,j)+c_{L2}(i,j)\right)/2 & \text{if } \left|c_{L1}(i,j)-c_{L2}(i,j)\right|<\alpha_1 \\ c_{L1}(i,j) & \text{if } \left|c_{L1}(i,j)-c_{L2}(i,j)\right|\geq\alpha_1 \text{ and } d_1(i,j)<d_2(i,j) \\ c_{L2}(i,j) & \text{if } \left|c_{L1}(i,j)-c_{L2}(i,j)\right|\geq\alpha_1 \text{ and } d_2(i,j)<d_1(i,j) \end{cases} \tag{7.5}$$

where $d_k(i,j)=\left|c_{Lk}(i,j)-c_{Lk}(i,j+1)\right|$ for k = 1, 2 and c_{Lc}, c_{L1}, c_{L2} represent the wavelet coefficients in the low resolution subband for the combined and the individual transforms.

The combination rule for the high frequency subbands also keeps the coefficient value at position (i, j) if it is the same in both transforms. If the difference between the received coefficients is less than a threshold α_2, then the average of the received coefficients is placed in the composite transform. If the difference is greater than α_2 then one of the coefficient values is chosen based on the descendant coefficient values (instead of the neighboring coefficient values as used for the low-resolution subband). This combination rule takes into account the relationships among the

coefficients in the same hierarchical trees. The diversity combination rule for the high frequency subbands is as follows:

$$c_{Hc}(i,j)=\begin{cases} c_{H1}(i,j) & \text{if } c_{H1}(i,j)=c_{H2}(i,j) \\ \left(c_{H1}(i,j)+c_{H2}(i,j)\right)/2 & \text{if } \left|c_{H1}(i,j)-c_{H2}(i,j)\right|<\alpha_2 \\ c_{H1}(i,j) & \text{if } \left|c_{H1}(i,j)-c_{H2}(i,j)\right|\geq\alpha_2 \text{ and } t_1(i,j)<t_2(i,j) \\ c_{H2}(i,j) & \text{if } \left|c_{H1}(i,j)-c_{H2}(i,j)\right|\geq\alpha_2 \text{ and } t_2(i,j)<t_1(i,j) \end{cases} \quad (7.6)$$

where $t_k(i,j)=\left|c_{Hk}(i,j)-\left(\sum_{m=0}^{1}\sum_{n=0}^{1}c_{Hk}(2i+m,2j+n)\right)/4\right|$ for k = 1, 2 and c_{Hc}, c_{H1}, c_{H2} represent the wavelet coefficients in the high frequency subbands for the combined and the individual transforms.

For our simulations we assumed the channel model had an average bit error rate of 0.01 and an average burst error length of 500 bits, with the error probabilities of each state being $P_e(0)$ = 0 and $P_e(1)$ = 0.5. The results presented here are PSNR distributions (a performance measure proposed in [SZ98]) plotted for 1000 runs. In these plots, smaller probability means better performance. We tested our diversity method on the Peppers and Lenna images, both 512 by 512 pixels and coded to 0.25 bits per pixel.

Figure 7.8 shows the results for the Lenna image for no error control, unequal error protection RS coding only, and our image domain diversity combining method including unequal error protection. The plot shows that the performance for our diversity combining method significantly improves Prob(PSNR < x) for noisy channel conditions. For example, the probability that the UEP decoded image will have a PSNR less than 26 dB is 0.55 while for the diversity combining image the probability that the PSNR will be less than 26 dB is 0.15. The mean PSNRs for no error control, UEP and diversity combining are 23.49 dB, 27.21 dB and 29.04 dB, respectively. Thus, image domain diversity combining provides an improvement over the UEP method of about 1.8 dB in mean PSNR. Also, the PSNR variance for the UEP method is 18.72 while the variance for the

diversity combining method is 6.47. The smaller variance implies that the quality of the received image will be more uniform.

Figure 7.9 shows a similar plot for the Peppers image results. The plot also shows that the performance for our diversity combining method improves Prob(PSNR < x) for noisy channel conditions. For example, the probability that the UEP decoded image will have a PSNR less than 26 dB is 0.54 while for the diversity combining image the probability is only 0.22. The mean PSNRs for no error control, UEP and diversity combining are 22.27 dB, 27.01 dB and 28.02 dB, respectively. So our image domain diversity combining provides an improvement over the UEP method of about 1.0 dB in mean PSNR. Also, the PSNR variance for the UEP method is 18.85 while the variance for the diversity combining method is 5.75. The results again show that the quality of the received image for our diversity combining method will be more uniform.

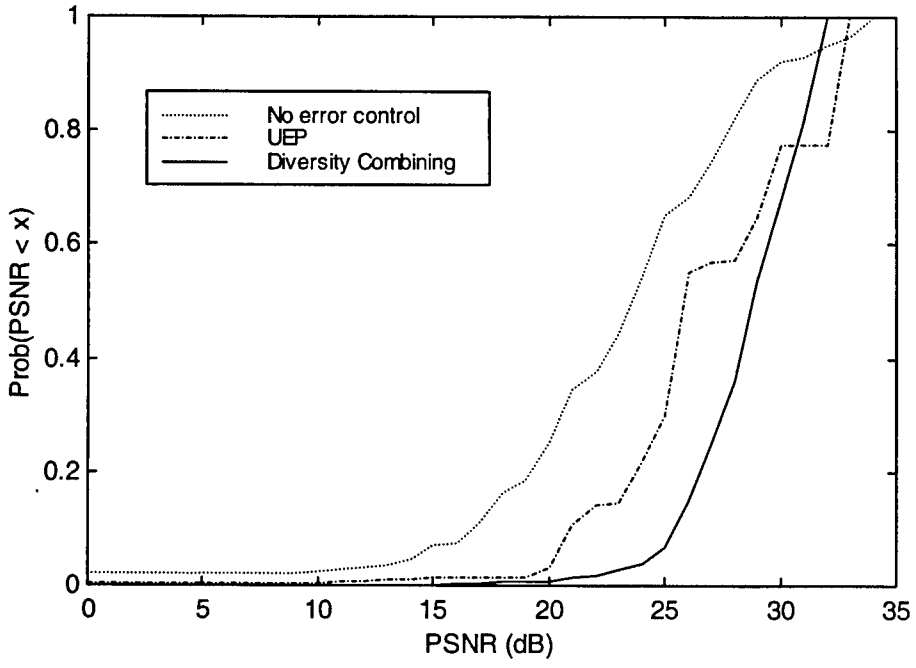


Figure 7.8: PSNR distribution for Lenna image source coded to 0.25 bpp.

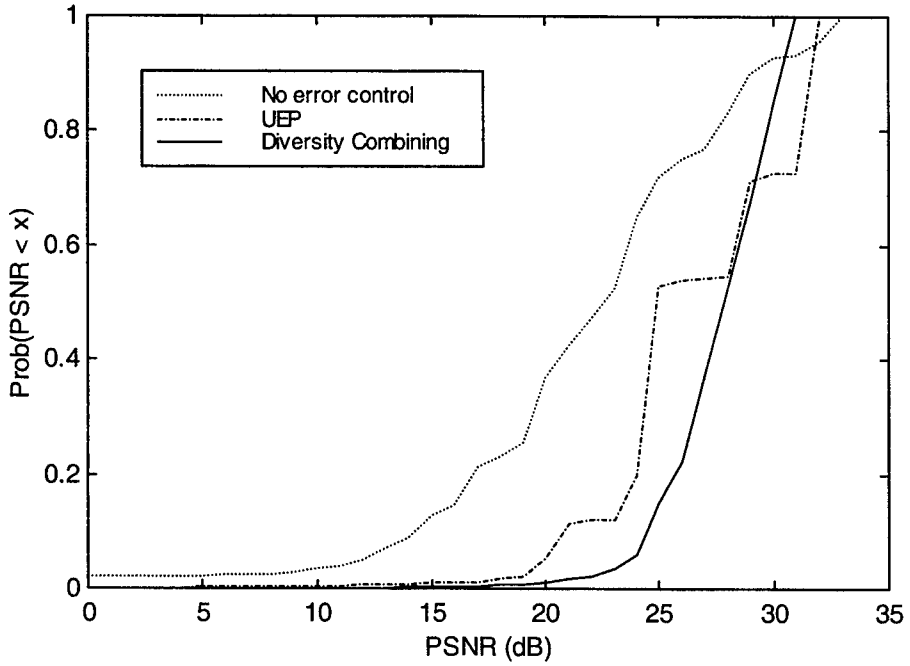


Figure 7.9: PSNR distribution for Peppers image source coded to 0.25 bpp.

For PSNR greater than 30 dB, both Figures 7.8 and 7.9 show that no error control and UEP perform better than diversity combining. But for PSNR values above 30 dB, the difference in the perceptual quality of the received image is negligible.

An image domain diversity method has been presented for the transmission of images over wireless channels. For images represented in the wavelet domain, diversity is used to obtain multiple data streams of the image at the receiver where these data streams are combined to obtain a composite image. The methods proposed here use some of the properties of the wavelet transform to significantly improve the perceptual quality of the received image. We first implemented diversity combining for uncompressed images to demonstrate the usefulness of this method. The diversity combining rules for compressed images were implemented with unequal error protection utilizing three different Reed Solomon error correction codes. However, image domain diversity combining can be used with any other channel coding scheme or wavelet based compression method to improve performance. Our results showed that image domain diversity could be used to improve performance

for images transmitted over wireless channels, possibly in conjunction with other error protection schemes.

8. CONCLUSIONS

Concealed weapons detection, an important problem for law enforcement, requires several processing stages that were covered in this report. The image registration algorithm described in this report specifically dealt with methods that perform well for IR and MMW images. When multiple images with different information were combined, wavelet based image fusion improved the ability to recognize weapons in the composite image. Preliminary results for the shape recognition algorithms show that this method is capable of recognizing weapon vs. non-weapon shapes. In addition, a fusion-based image transmission method was presented for sending images over wireless communication channels.

As indicated in Chapter of this report, there are certain conditions where our registration algorithm does not provide accurate results. Other registration methods and rules can be added to the current algorithm to deal with different imaging conditions. For the shape recognition stage, the libraries require additional data in order to build the libraries with more typical weapon and non-weapon shapes. When more data is available, further testing of the shape recognition stage will be performed. We may also add more shape descriptors to the three descriptors that are currently being utilized in our recognition algorithm. For this report, we only used data from imaging sensors for the CWD problem. In future work, we may investigate multisensor fusion of data from both imaging and non-imaging sensors.

REFERENCES

[BA98] P. Burlina and F. Alajaji, "An Error Resilient Scheme for Image Transmission over Noisy Channels with Memory," IEEE Trans. on Image Processing, Vol. 7, no. 4, pp. 593-600, April 1998.

[Ber76] R. Berstein, "Digital image processing of Earth observation sensor data," IBM J. Res. Devel., vol. 20, pp. 40-67, Jan, 1976.

[BL93] P.J. Burt and R.J. Lolczynski, "Enhanced image capture through fusion," in Proceedings of Fourth International Conference on Computer Vision, (Berlin, Germany), pp. 173-182, May 1993.

[Bro92] L.G. Brown, "A Survey of Image Registration Techniques," ACM Computing Surveys, vol. 24, no. 4, pp. 325-376, Dec. 1992.

[BS72] D.I. Barnea and H.F. Silverman, " A class of algorithms for fast digital registration," IEEE Trans. On Comput., C-21, pp. 179-186, 1972.

[Bur84] P.J. Burt, "The pyramid as a structure for efficient computation," in Multiresolution Image Processing and Analysis (A. Rosenfeld, ed.),pp 6-35, New York/Berlin: Springer-Verlag, 1984.

[CDF96] N.C. Currie, F.J. Demma, D.D. Ferris, Jr., B.R. Kwasowsky, R.W. McMillan, and M.C. Wicks, "Infrared and Millimeter Wave Sensors for Military Special Operations and Law Enforcement Applications," International Journal of Infrared and Millimeter Waves, vol. 17, no. 7, 1996.

[CF98] C.H. Chen and T.E. Fuja, " Transmission of compressed images over noisy channels using incremental precision and incremental redundancy," Proc. Of the Conf. on Information Sciences and Systems, Princeton, March 1998.

[CM88] B. Chanda and D.D. Majumder, "A note on the use of the gray-level co-occurrence matrix in threshold selection," Signal Processing, vol. 15, no. 2, pp. 149-167, 1988.

[CRSZ98] P.C. Cosman, J.K. Rogers, P.G. Sherwood, and K. Zeger, "Combined Forward Error Control and Packetized Zerotree Wavelet Encoding for Transmission of Images over Varying Channels," submitted to IEEE Transactions on Image Processing, July 1998.

[DG93] R. Deriche and G Giraudon, " A Computational Approach for Corner and Vertex Detection," International Journal of Computer Vision, 10:2, pp. 101-124, 1993.

[DK82] P.A. Devijver and J. Kittler, Pattern Recognition: A Statistical Approach, London, U.K.: Prentice-Hall, 1982, pp. 353-357.

[Ell63] E.O. Elliott, "Estimates of error rates for codes on burst error channels," Bell Systems Technical Journal, vol. 42, p. 1977, Sept. 1963.

[FK97] X. Fei and T. Ko, "Turbo-codes for compressed image transmission over frequency selective fading channel," IEEE 6th Int. Conf. On Universal Personal Communications, vol. 2, pp. 505-509, 1997.

[Fuk90] K. Fukunaga, Introduction to Statistical Pattern Recognition, Second ed., San Diego, CA: Academic Press, Inc., 1990, pp. 220-221.

[GB85] A.W. Gruen and E.P. Baltsavias, " Adaptive Least Squares correlation with geometrical constraints," Computer Vision for Robots, SPIE, vol. 595, pp. 72-82, 1985.

[Gil60] E.N. Gilbert, "Capacity of a Burst-Noise Channel," The Bell Systems Technical Journal, pp. 1253-1265, September 1960.

[Gla93] C.A. Glasbey, "An Analysis of Histogram-based Thresholding Algorithms," Graphical Models and Image processing, vol. 55, no. 6, pp. 532-537, Nov. 1993.

[Gos86] A. Goshtasby, "Piecewise Linear Mapping Functions for Image Registration," Pattern Recognition, vol. 19, no. 6, pp. 459-466, 1986.

[GS87] L. Gupta and M.D. Srinath, "Contour sequence moments for the classification of closed planar shapes," Pattern Recognition, vol. 20, no. 3, pp. 267-272, 1987.

[GW87] R.C. Gonzalez and P.Wintz, Digital Image Processing, Second ed., Reading, MA: Addison-Wesley, 1987.

[Hei92] C. Heipke, " A global approach for least squares image matching and surface reconstruction in object space," Photogrammetric Engineering and Remote Sensing, vol. 58, no. 3, pp. 317-323, 1992.

[Hem96] S.S. Hemami, "Reconstruction-Optimized Lapped Orthogonal Transforms for Robust Image Transmission," IEEE Trans. on Circuits and Systems for Video Technology, vol. 6, no. 2, pp. 168-181, April 1996.

[HL95] Y.H. Han, J.J. Leou, "Detection and Correction of Transmission Errors in JPEG Images," IEEE Trans. on Circuits and Systems for Video Technology, Vol. 8, no. 2, pp. 221-231, April 1998.

[Hor89] B.K.P. Horn, Robot vision, MIT Press, Cambridge, MA, 1989.

[HS88] C. Harris, M. Stephens, " A combined corner and edge detector," Proc. of 4th Alvey Vision Conference, pp. 147-151, 1998.

[Iun98] E.V.H. Iun, "Combined source-channel coding for the transmission of still images over a code division multiple access (CDMA) channel," IEEE Communications Letters, vol. 2, no. 6, pp. 168-170, June 1998.

[JN95] E.S. Jang and N.M. Nasrabadi, "Subband Coding with Multistage VQ for Wireless Image Communication," IEEE Transactions on Circuits and Systems for Video Technology, Vol. 5, no. 3, pp. 247-253, June 1995.

[KCR98] I. Kozintsev, J. Chou, and K. Ramchandran, "Image transmission using arithmetic coding based error detection," Proc. IEEE Data Compression Conference, pp. 339-348, 1998.

[KH75] C.D. Kuglin and D.C. Hines, " The Phase Correlation Image Alignment Method," Proc. IEEE Int. Conf. on Cybernetics and Society, pp. 163-165, 1975.

[KJ91] R. Katuri and R.C. Jain, Computer Vision: Principles, IEEE Computer Society Press, Los Alamitos, CA, 1991

[KR97] I. Kozintsev and K. Ramchandran, "Efficient Hybrid Framework for Wireless Image Transmission," Proc. IEEE, pp. 1037-1041, 1997.

[KRD80] D.J. Kahl, A. Rosenfeld, and A. Danker, " Some experiments in point pattern matching," IEEE Trans. Syst. Man Cybern., SMC-10, no. 2, pp. 105-116, 1980.

[KS78] L.N. Kanal and A.R.K. Sastry, "Models for Channels with Memory and Their Applications to Error Control," Proc. of the IEEE, vol. 66, no. 7, pp. 724-744, July 1978.

[KY95] Y. Koyama and S. Yoshida, "Error Control for Still Image Transmission over a Fading Channel," IEEE 45th Vehicular Technology Conference, Vol. 2, pp. 609-613, 1995.

[Lie95] Wen-Nung Lie, "Automatic Target Segmentation by Locally Adaptive Image Thresholding," IEEE Trans. Image Processing, vol. 4, no. 7, July 1995.

[LLL97] J. Lu, M.L. Liou, and K.B. Letaif, "Efficient image transmission over wireless channels," IEEE Int. Symp. on Circuits and Systems (Hong Kong), pp. 1097-1100, June 1997.

[LMM95] H. Li, B.S. Manjunath and S.K. Mitra, "Multisensor image fusion using the wavelet transform," Graphical Models and Image Processing, vol. 57, pp. 235-245, May 1995.

[LR95] W.M. Lam and A.R. Reibman, "An Error Concealment Algorithm for Images Subject to Channel Errors," IEEE Transactions on Image Processing, vol. 4, no. 5, pp. 533-541, May 1995.

[Mer] M. Merickel, "3D reconstruction: The registration problem," Computer Vision Graphics and Image Processing, vol. 47, pp. 203-226.

[MF93] C.R. Maurer and J.M. Fitzpatrick, "A review of medical image registration," Interactive Image-guided neurosurgery, R.J. Maciunas, ed. Park Ridge, IL: American Assoc. of Neurological Surgeons, 1993, pp. 17-44.

[MGTH95] T.H. Meng, B.M. Gordon, E.K. Tsern and A.C. Hung, "Portable Video-on-Demand in Wireless Communication," Proceedings of the IEEE, Vol. 83, no. 4, pp. 659-680, April 1995.

[MNR90] G.Q. Maguire, M.E. Noz, and H. Rusinek, "Software tools to standardize and automate the correlation of images with and between diagnostic modalities," IEEE Comput. Graph. Appl., 1990.

[Mor81] H. Moravec, "Rover visual obstacle avoidance," Proc. of the 7th Int. Conf. on Artificial Intelligence, Vancouver, B.C., Canada, pp. 785-790, 1981.

[Ots79] N. Otsu, "A Threshold Selection Method from Gray-level Histograms", IEEE Trans. SMC, vol. 9, no. 2, 62-66, 1979.

[PLA91] M. Pavel, J. Larimer and A. Ahumada, "Sensor fusion for synthetic vision," in Proceedings of AIAA Conference on Computing in Aerospace,(Baltimore, MD), October 1991.

[RSC96] B. Reddy, Strinivasa, B.N. Chatterji, " An FFT Based technique for Translation, Rotation and Scale-Invariant Image Registration," IEEE Trans. On Image Processing, vol. 5, no. 8, August 1996.

[RR78] W.S. Rutkowski, A. Rosenfeld, "A comparison of corner-detection techniques for chain-coded curves," TR-623 (MCS-76-23763), Computer Science Center, University of Maryland, College Park, MD, 1978.

[RR80] S. Ranade and A Rosenfeld, "Point pattern matching by relaxation," Pattern Recognition, vol. 12, pp. 269-275, 1980.

[Sch96] C. Schmid, "Appariement d'images par invariants locaux de niveaux de gris," unpublished Doctoral dissertation, l'Institut National Polytechnique de Grenoble, 1996.

[SFF] M.R. Stytz, G. Frieder, and O. Frieder, "Three-dimensional Medical Imaging: Algorithms and Computer Systems," ACM Comput. Survey, 23, 4 (Dec), pp. 421-424.

[SKB82] G.C. Stockman, S. Kopstein, and S. Benett, " Matching images t models for registration and object detection via clustering," IEEE Trans. Pattern Analysis and Machine Intelligence, 4, pp. 229-241, 1982.

[SLCP96] Z. Sun, J, Luo, C.W. Chen and K.J. Parker, "Analysis of a Wavelet-based Compression Scheme for Wireless Image Communication," Proc. of SPIE, Vol. 2762, pp. 454-465, 1996.

[SM95] C. Schmid and R. Mohr, "Matching by local invariants," Rapport de Recherche, N 2644, INRIA, 1995.

[SP96] A. Said and W.A. Pearlman, "A new, fast, and efficient image codec based on set partitioning in hierarchical trees," IEEE Transactions on Circuits and Systems for Video Technology, vol. 6, no. 3, pp. 243-250, June 1996.

[SRD94] L. Shen, R.M. Rangayyan and J.E. Leo Desautels, "Application of shape analysis to mammographic calcifications," IEEE Trans. Medical Imaging, vol. 13, no. 2, pp. 263-274, June 1994.

[SZ98] P.G. Sherwood and K. Zeger, "Error Protection for Progressive Image Transmission Over Memoryless and Fading Channels," IEEE Trans. on Communications, vol. 46, no. 12, pp. 1555-1559, December 1998.

[TBC86] I.L. Thomas, V.M. Benning, and N.P. Ching, Classification of Remotely Sensed Images, Adam Hilger, Bristol, England, 1986

[TH94] L. Tang and C. Heipke, "An approach for automatic relative orientation," Proc. Optical 3D Measurement Techniques II: Applications in inspection, quality control and robotics, A. Gruen and H. Kahmen, eds., SPIE, vol. 2252, pp. 347-354, 1994.

[Toe90] A. Toet, "Hierarchical image fusion," Machine Vision Applications, pp. 1-11, March 1990.

[URVA96] M.K. Uner, L.C. Ramac, P.K. Varshney, and M. Alford, "Concealed Weapon Detection: An Image Fusion Approach," Proc. of the Enabling Technologies for Law Enforcement and Security Symposium, Conf. 2942, SPIE's Photonics East, Boston, Nov. 1996.

[VH96] R.E. Van Dyck and G.M. Horiban, "Wavelet video transmission over tactical wireless channels," First Int. Conf. on Wireless Image/Video Communications, pp. 74-79, Sept. 1996.

[WO86] J.W. Woods and S.D. O'Neil, "Subband coding of images," IEEE Trans. on Acoust., Speech, Signal Processing, vol. ASSP-34, Oct. 1986.

[WPE96] V. Weerackody, C. Podilchuk, and A. Estrella, "Transmission of JPEG coded images over wireless channels," Bell Labs Technical Journal, pp. 111-125, Autumn 1996.

[WZ95] V. Weerackody and W. Zeng, "ARQ schemes with switched antenna diversity and their applications in JPEG image transmission," IEEE GLOBECOM, vol. 3, pp. 1915-1919, 1995.

[YW95] J.R. Yee and E.J. Weldon, "Evaluation of the performance of error-correcting codes on a Gilbert channel," IEEE Trans. on Communications, vol. 43, pp. 2316-2323, Aug. 1995.

[ZL95] W. Zeng and B. Liu, "Geometric-structure-based directional filtering for error concealment in image/video transmission," Proc. of the SPIE: Wireless Data Transmission, vol.2601, pp. 145-156, October 1995.

*U.S. GOVERNMENT PRINTING OFFICE: 2001-610-055-10107

MISSION
OF
AFRL/INFORMATION DIRECTORATE (IF)

The advancement and application of Information Systems Science and Technology to meet Air Force unique requirements for Information Dominance and its transition to aerospace systems to meet Air Force needs.