Available online at www.sciencedirect.com
ScienceDirect
Pattern Recognition 40 (2007) 1771–1784
PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY
www.elsevier.com/locate/pr

Fusion of color and infrared video for moving human detection

Ju Han, Bir Bhanu*

Center for Research in Intelligent Systems, University of California, Riverside, CA 92521, USA

Received 8 November 2005; received in revised form 17 October 2006; accepted 9 November 2006

Abstract

We approach the task of human silhouette extraction from color and thermal image sequences using automatic image registration. Image registration between color and thermal images is a challenging problem due to the difficulties associated with finding correspondence. However, the moving people in a static scene provide cues to address this problem. In this paper, we propose a hierarchical scheme to automatically find the correspondence between the preliminary human silhouettes extracted from synchronous color and thermal image sequences for image registration. Next, we discuss strategies for probabilistically combining cues from registered color and thermal images for improved human silhouette detection. It is shown that the proposed approach achieves good results for image registration and human silhouette extraction. Experimental results also show a comparison of various sensor fusion strategies and demonstrate the improvement in performance over non-fused cases for human silhouette extraction.

Keywords: Sensor fusion; Image registration; Human silhouette extraction; Color image sequence; Thermal image sequence; Genetic algorithm

1. Introduction

Current human recognition methods, such as fingerprints, face or iris biometrics, generally require a cooperative subject, views from certain aspects and physical contact or close proximity. These methods cannot reliably recognize non-cooperating individuals at a distance in a real-world under changing environmental conditions. Moreover, in many practical applications of personnel identification, most of the established biometrics may be obscured. Gait, which concerns recognizing individuals by the way they walk, can be used as a biometric without the above-mentioned disadvantages.

The initial step of most of the gait recognition approaches is human silhouette extraction [1–6]. Many gait recognition approaches use electro-optical (EO) sensors such as CCD cameras. However, it is very likely that some part of the human body or clothing has colors similar to the background. In this case, human silhouette extraction usually fails on this part. Moreover, the existence of shadows is a problem for EO sensors [7]. In addition, EO sensors do not work under low lighting conditions such as night or indoor environment without lighting. The top rows in Fig. 1 show human silhouette extraction results from two color images.

To avoid the disadvantages of using EO sensors, infrared (IR) sensors are used for object detection [8,9]. We investigate the possibility of using an IR sensor for gait analysis [10]. Unlike a commonly used video camera that operates in the visible band of the spectrum and records reflected light, a long wave (8–12 μm) IR sensor records electromagnetic radiations emitted by objects in a scene as a thermal image whose pixel values represent temperature. In a thermal image that consists of humans in a scene, human silhouettes can be generally extracted from the background regardless of lighting conditions and colors of the human clothing and skin, and backgrounds, because the temperatures of the human body and background are different in most situations [11]. Although the human silhouette extraction results from IR sensors are generally better than that from EO sensors, human silhouette extraction is unreliable when some part of the human body or clothing has the temperature similar to the background temperature. In addition, human body casts obvious projection on smooth surfaces such

* Corresponding author. Tel.: +1 951 827 3954.
E-mail addresses: jhan@cris.ucr.edu (J. Han), bhanu@cris.ucr.edu (B. Bhanu).

0031-3203/$30.00 © 2006 Published by Elsevier Ltd on behalf of Pattern Recognition Society.
doi:10.1016/j.patcog.2006.11.010

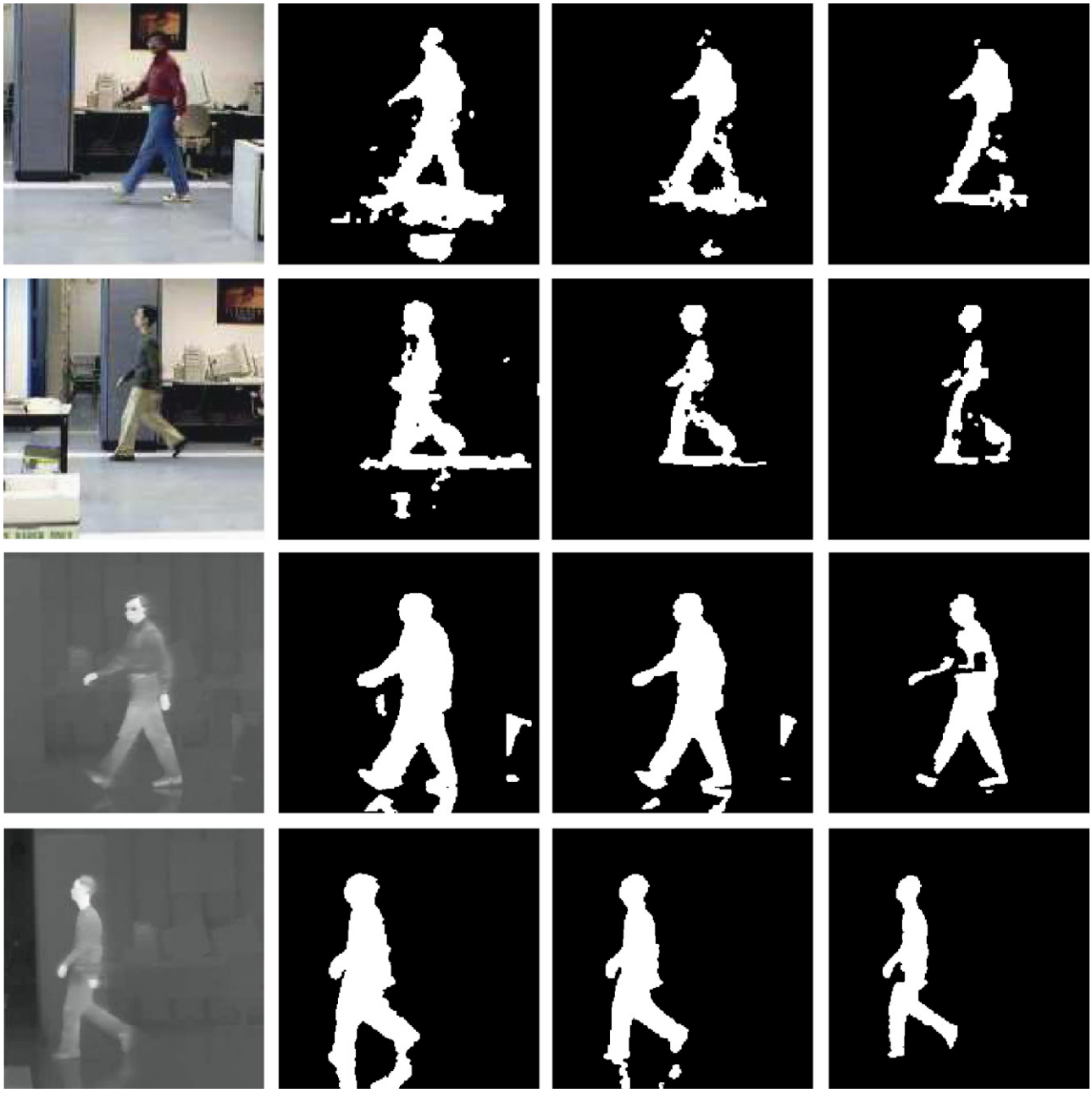


Fig. 1. Human silhouette extraction results from color images (first two rows) and thermal images (last two rows) using the background subtraction method with increasing thresholds from left to right. The leftmost image is the original image.

as a smooth floor. The last two rows in Fig. 1 show human silhouette extraction results from a thermal image.

In Fig. 1, notice that the unreliably extracted body parts from one sensor might be reliably extracted from the other sensor. This provides an opportunity for improving the human detection performance by the fusion of EO and IR sensors.

2. Related work and our contribution

Images from different kind of sensors generally have different pixel characteristics due to the phenomenological differences between the image formation processes of the sensors. In recent years, sensor fusion approaches have already been employed to improve the performance of object detection and recognition, especially in the field of automated target recognition [12,13] and remote sensing [14–17].

Wilder et al. [18] compare the effectiveness of EO and IR imagery for detecting and recognizing faces, and expect the improvement of face detection and recognition algorithms that fuse the information from the two sensors. Yoshitomi et al. [19] propose an integrated method to recognize the emotional expressions of a human using voice, and color and thermal images of face. The recognition results show that the integration method for recognizing emotional states gives better

performance than any of individual methods. In these approaches, images from different sensors are independently processed without an image registration procedure. In these applications, the fusion for object detection and recognition takes place at the decision level.

Image registration is essential for precise comparison or fusion of images from multiple sensors at the pixel level. Sometimes, manual image registration is employed in many sensor fusion approaches [12,13]. This needs lots of human interaction which is not desirable in processing large collections of image data taken under different field-of-views of the sensors.

Many approaches have been proposed for automatic registration between SAR and optical images. Li et al. [14] proposed an elastic contour matching scheme based on the active contour model for multisensor image registration between SAR (microwave) and SPOT (visible and near IR) images. Inglada and Adragna [15] proposed an approach for automatic image registration between SAR and SPOT images. They first extract edges in both images, and then use a genetic algorithm (GA) to estimate the geometric transformation which minimizes the matching error between corresponding edges. Similarly, Ali and Clausi [16] automatically register SAR and visible band remote sensing images using an edge-based pattern matching method. In order to locate reliable control points between SAR and SPOT images, Dare and Dowman [17] proposed an automatic image registration approach based on multiple feature extraction and matching methods, rather than just relying on one method of feature extraction.

Zheng and Chellappa [20] propose an automatic image registration approach to estimate 2-D translation, rotation and scale of two partially overlapping images obtained from the same sensor. They extract features from each image using a Gabor wavelet decomposition and a local scale interaction method to detect local curvature discontinuities. Hierarchical feature matching is performed to obtain the estimate of translation, rotation and scale. Li and Zhou [21,22] extend this single sensor image registration approach to the work of automatic EO/IR and SAR/IR image registration. Their approach is based on the assumption that some strong contours are presented in both the EO and IR images. Consistent checking is required to remove inconsistent features between images from different sensors.

Due to the difficulty in finding a correspondence between images with different physical characteristics, image registration between imagery from different sensors is still a challenging problem. In our task, objects in color and thermal images appear different due to different phenomenology of EO and IR sensors. Also, there are differences in the field-of-view and resolution of the sensors. Therefore, it is generally difficult to precisely determine the corresponding points between color and thermal images. However, in a human walking sequence, human motion provides enough cues for image registration between color and thermal images. In this paper, we first propose a GA-based hierarchical correspondence search approach for automatic image registration between synchronous color and thermal image sequences. The proposed approach reduces the overall computational load of the GA without decreasing the final estimation accuracy. The registered thermal and color images are then combined by probabilistic strategies at the pixel level to obtain better body silhouette extraction results.

Mandava et al. [23] proposed an adaptive GA-based approach for medical image registration with manually selected region-of-interest. Compared with their approach, our approach employs the similar concept of hierarchical search space scaling in GA. However, the two approaches are different in strategies, implementation and applications.

In comparison with state-of-the-art, the contribution of this paper are:

- *Automatic image registration-based preliminary silhouette matching*: Due to the phenomenological differences of objects in color and thermal images, it is difficult to automatically find accurate correspondence between color and thermal images. However, human motion provides enough cues for automatic image registration between synchronized color and thermal images in our human silhouette extraction application. Compared with the correspondence of individual points, the preliminary extracted body silhouette regions provide a more reliable correspondence between color and thermal image pairs. In this paper, we propose a automatic image registration method to perform a match of the transformed color silhouette to the thermal silhouette.
- *Hierarchical genetic algorithm-based search scheme*: We use GA to solve the optimization problem in silhouette matching. However, the accurate subpixel corresponding point search requires longer bit length for each coordinate value of each point. As a result, the population size of GA needs to be large to reduce the probability of falling into a local maxima. Due to the costly fitness function, the large population size is not desirable. In this paper, we propose a hierarchical genetic algorithm- (HGA) based search scheme to estimate the model parameters within a series of windows with adaptively reduced size at different levels.
- *Sensor fusion*: To improve the accuracy of human silhouette extraction, we combine the information from the registered color and thermal images. Various fusion strategies are applied for human body silhouette detection by combining registered color and thermal images. Experimental results show that the sum rule achieves the best results.

3. Technical approach

In this paper, we propose a GA-based hierarchical correspondence search approach for automatic image registration between synchronous color and thermal image sequences as shown in Fig. 2. Input to the system are videos possibly with moving humans, recorded simultaneously by both EO and IR cameras. Next, a background subtraction method is applied to both color and thermal images to extract preliminary human body silhouettes from the background. Silhouette centroids are then computed from the color and thermal silhouettes as the initial corresponding points between color and thermal images. A HGA-based scheme is employed to estimate the exact

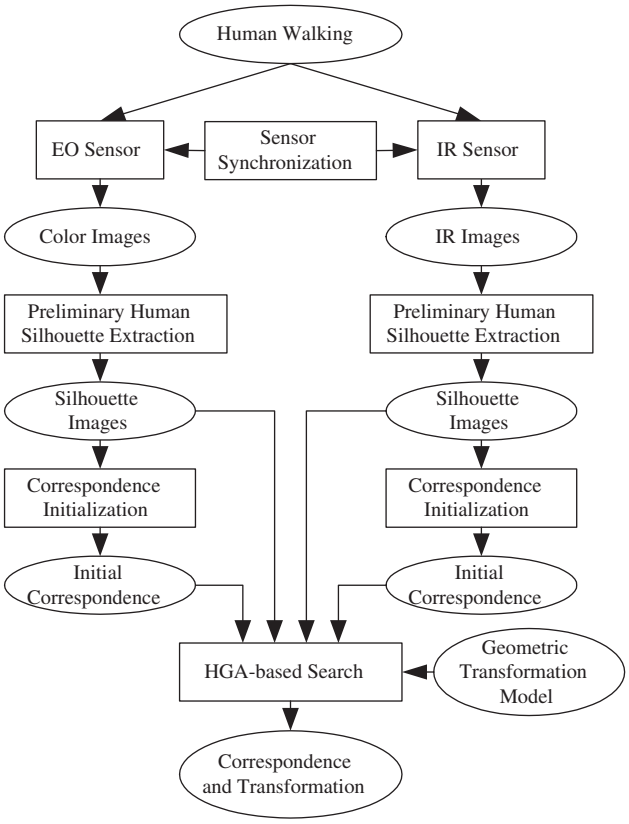


Fig. 2. Proposed hierarchical genetic algorithm-based multi-modal image registration approach.

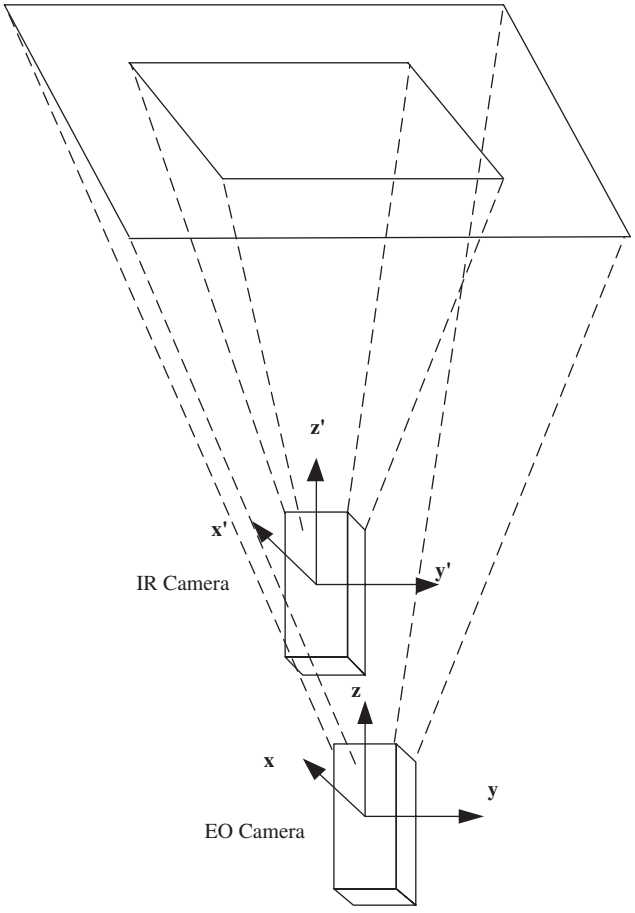


Fig. 3. IR and EO camera set-up.

correspondence so that the silhouettes form the synchronous color and thermal images are well matched. The transformation so obtained from this correspondence is used for the registration of images from EO and IR cameras. Finally, registered thermal and color images are combined using probabilistic strategies to obtain better body silhouette extraction results.

3.1. Image transformation model

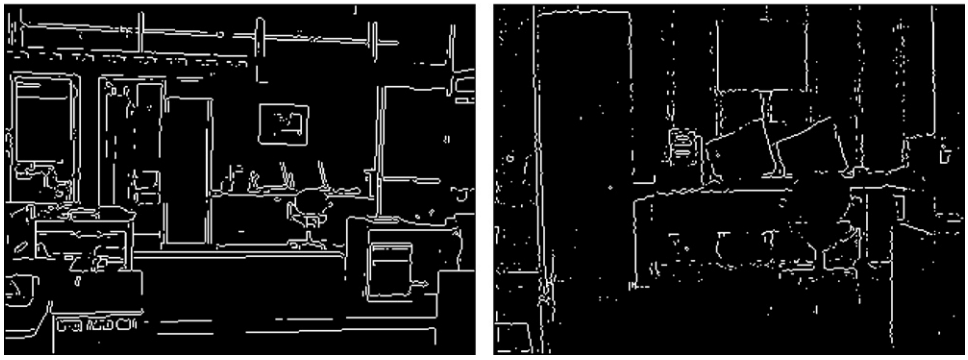
We use one EO camera and one IR camera for sensor fusion. We place the EO and IR camera as close as possible without interference as shown in Fig. 3, and adjust their camera parameters so that the field-of-views of both cameras contain the desired scene where human motion occurs. The geometric transformation between the cameras involved can be represented by a 3-D linear transformation and a 3-D translation. According to the degree of elasticity of the transformations, they can be rigid, affine, projective, or curved [24]. Assuming that there is a large distance between the camera and the walking people, the visible human surface from the camera view can be approximated as planar. The geometric transformation for planar objects can be strictly represented by a projective transformation [25]. Furthermore, assuming that the image planes of both EO and IR cameras are approximately parallel, the geometric transformation can be further simplified as the rigid model. A rigid transformation can be decomposed into 2-D translation, rotation and reflection. In the rigid transformation, the distance between any two points in the color image plane is preserved when these two points are mapped into the thermal image plane. The 2-D point (X, Y) in the color image plane is transformed into the 2-D point (X', Y') in the thermal image plane as follows:

$$\begin{pmatrix} X' \\ Y' \end{pmatrix} = s \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix} + \begin{pmatrix} \Delta X \\ \Delta Y \end{pmatrix}, \quad (1)$$

where θ is the rotation angle, s is the scaling factor and $(\Delta X, \Delta Y)^{\mathrm{T}}$ is the translation vector. Rigid transformations are used when shapes in the input image are unchanged, but the image is distorted by some combination of translation, rotation and scaling. Straight lines remain straight, and parallel lines are still parallel under the assumption as mentioned above. A minimum correspondence of two pairs of points is required in rigid transformation.

In this paper, the rigid transformation model is used for image registration between synchronous EO/IR sequence pairs. If the assumption of parallel image planes is not satisfied, the proposed approach can be easily extended by using more complex models such as projective transformation model. The only difference is the number of parameters to be estimated.

Color Background Image Thermal Background Image



Color Edge Image Thermal Edge Image

Fig. 4. Different object appearances in color and thermal images are due to the phenomenological differences between the image formation process of EO and IR cameras. The images are at different resolutions and the field-of-views (FOVs) of the two cameras overlap (EO camera FOV contains IR camera FOV).

3.2. Preliminary human silhouette extraction and correspondence initialization

Assume that both the EO and IR cameras are fixed and mounted on suitable stationary platforms. Further assume that human is the only moving object in the scene and there is only one person at any time in the scene. Under this situation, silhouettes of moving humans can be extracted by a background subtraction method. To model the color background, we choose all frames from a color image sequence that contains background only, and compute the mean and standard deviation values for each pixel in each color channel. Assuming that the background has a Gaussian distribution at each pixel, a pixel at (X, Y) in the input color image is classified as part of moving objects if

$$|r(X, Y) - \mu_r(X, Y)| > \alpha\sigma_r(X, Y), \tag{2}$$

$$|g(X, Y) - \mu_g(X, Y)| > \alpha\sigma_g(X, Y) \tag{3}$$

or

$$|b(X, Y) - \mu_b(X, Y)| > \alpha\sigma_b(X, Y), \tag{4}$$

where r, g and b represent pixel color values of the input image for red, green and blue channels, respectively; μ_r, μ_g and μ_b represent mean values of the background pixels; σ_r, σ_g and σ_b represent standard deviation values of the background pixels; α is the threshold.

Similarly, a pixel at (X, Y) in the input thermal image is classified as part of moving objects if

$$|t(X, Y) - \mu_t(X, Y)| > \beta\sigma_t(X, Y), \tag{5}$$

where t represents the pixel thermal value in the input thermal image; μ_t represents the mean value of the background pixel temperature; σ_t represents the standard deviation value of the background pixel temperature; β is the threshold.

After body silhouettes are extracted from each color image and its synchronous thermal image, the centroid of the silhouette region is computed as the initial correspondence between each pair of color and thermal images.

3.3. Automatic image registration

In applications with manual image registration, a set of corresponding points are manually selected from the two images to compute the parameters of the transformation model, and the registration performance is generally evaluated by manually comparing the registered image pairs. The same step is repeated several times until the registration performance is satisfied. If the background changes, the entire procedure needs to be repeated again. This makes manual image registration inapplicable when data are recorded at different locations with changing time or with different camera setup. The automatic image registration is desirable under this situation.

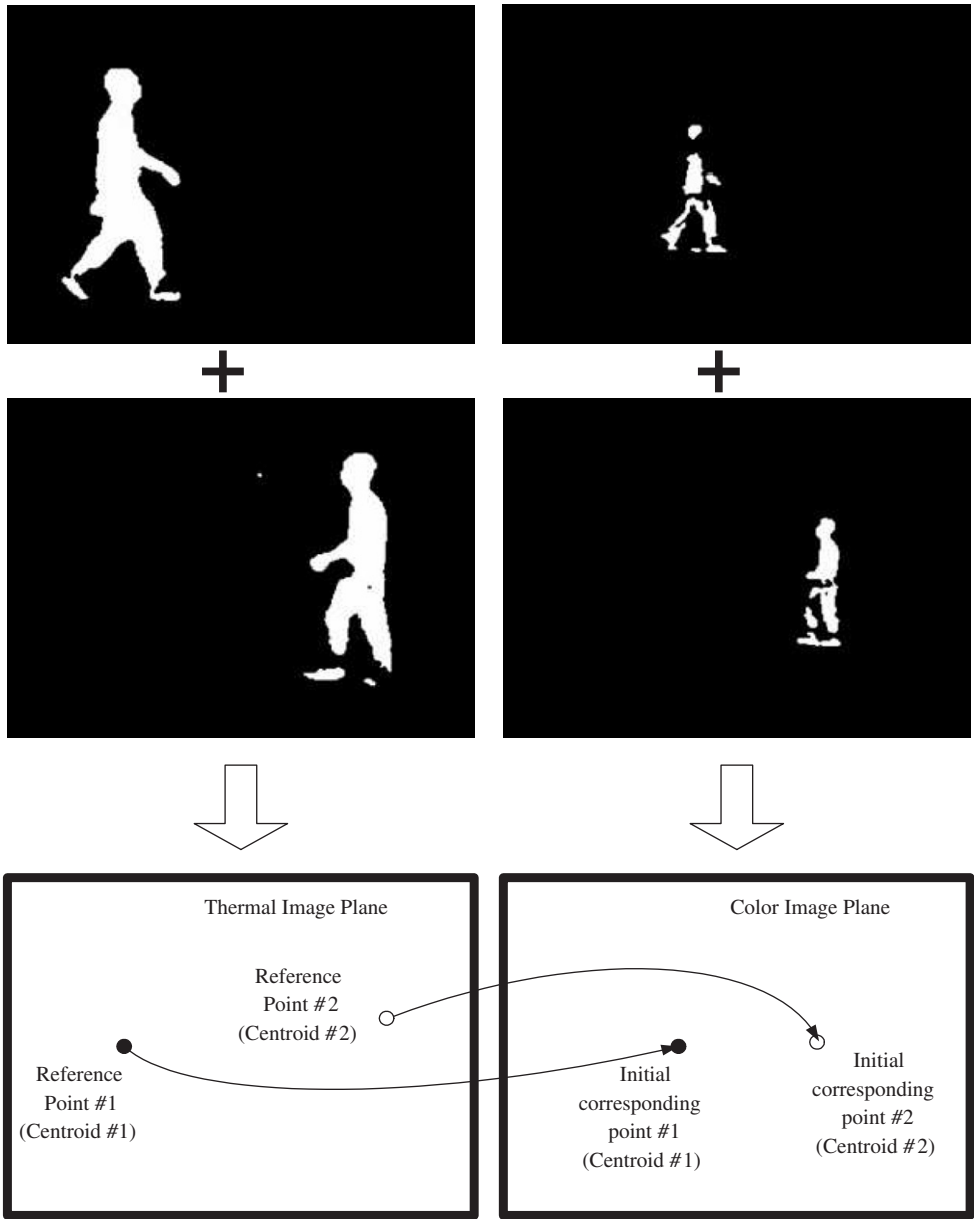


Fig. 5. Illustration of initial control point selection in both color and thermal image planes.

3.3.1. Model parameter selection

Due to the phenomenological differences of objects in color and thermal images, it is difficult to automatically find accurate correspondence between color and thermal images. Fig. 4 shows different object appearances in color and thermal images due to the phenomenological difference between the image formation process of EO and IR cameras. However, human motion provides enough cues for automatic image registration between synchronized color and thermal images in our human silhouette extraction application. Compared with the correspondence of individual points, the preliminary extracted body silhouette regions provide a more reliable correspondence between color and thermal image pairs. Therefore, we propose a method to perform a match of the transformed color silhouette to the thermal silhouette. That is, we estimate the set of model parameters $\mathbf{p}$ to maximize

$$\text{Similarity}(\mathbf{p}; I_i; C_i) = \prod_{i=1}^{N} \frac{\text{Num}(T_{C_i;\mathbf{p}} \cap I_i)}{\text{Num}(T_{C_i;\mathbf{p}} \cup I_i)}, \tag{6}$$

where I is the silhouette binary image obtained from the thermal image, C is the silhouette binary image obtained from color image, $T_{C;\mathbf{p}}$ is the transformed binary image of C by rigid transformation with parameter set $\mathbf{p}$, N is the number of color and thermal image pairs, and $\text{Num}(X)$ is the number of silhouette pixels in a silhouette image X. We use the product of similarity of image pairs instead of the sum to reduce the possibility of falling into local maxima on specific frame(s), i.e., to increase the possibility of the global maximum on all images pairs.

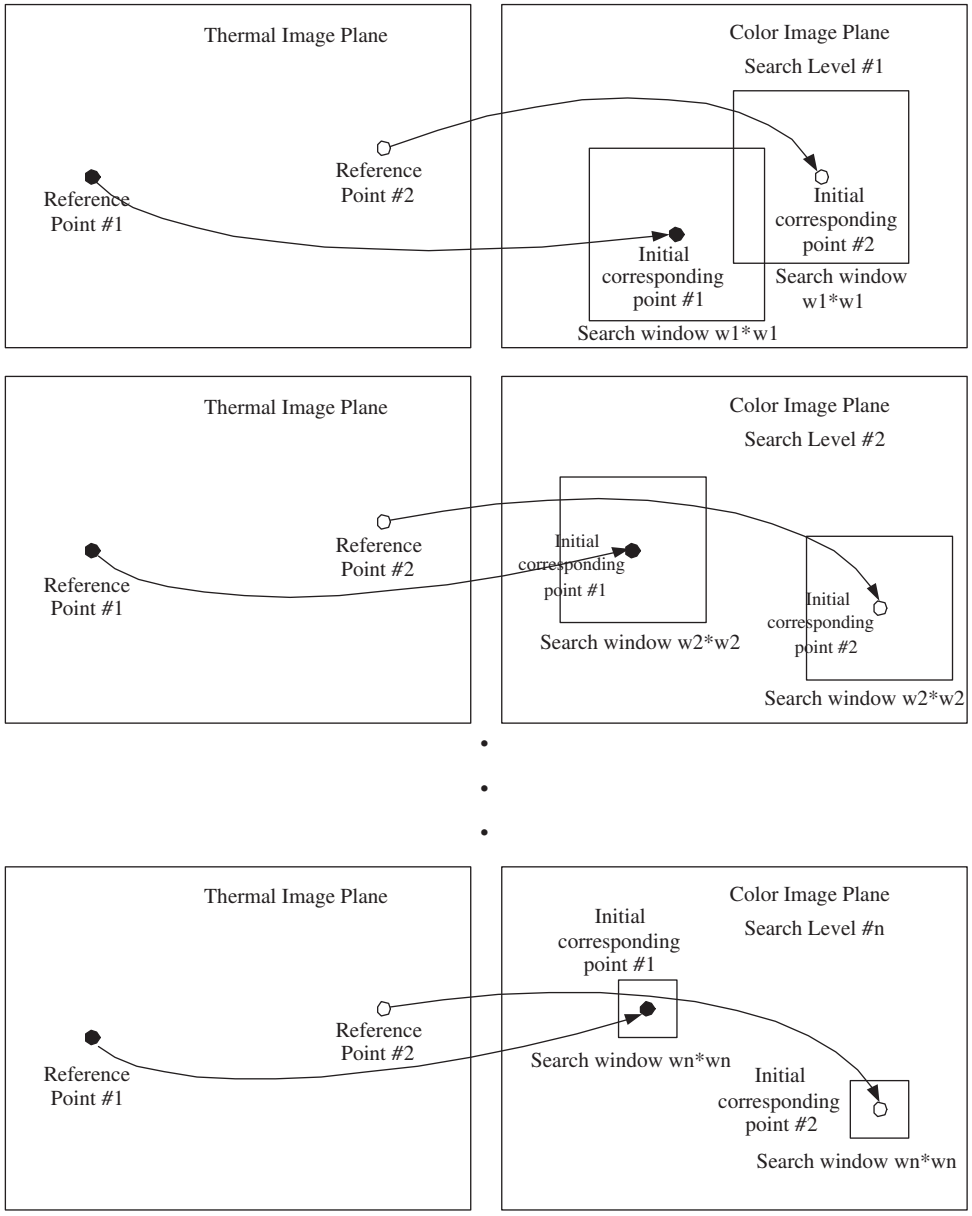


Fig. 6. Illustration of the HGA-based search scheme to estimate the model parameters—coordinate values of the two points (black and white) simultaneously in the color image plane. The search windows here are only for illustration whose sizes are much larger than the real window sizes.

In the rigid transformation model, the parameters are the elements of the 2-D linear transformation in Eq. (1). However, the ranges of these parameters are difficult to be determined. In the rigid transformation model, a maximum correspondence of two pairs of points is required. If we fix two points in the thermal image as the reference points, the 2-D coordinates of their corresponding points in the synchronous color image can be used to determine the rigid transformation model. Because the locations of the corresponding points should exist in limited local areas, the ranges of new parameters can be determined.

For each pair of color and thermal images, we obtain a pair of initial corresponding points, i.e., the centroids of the preliminary extracted silhouettes. Under the assumption of planar object surface (i.e., human walks along the same direction in the scene so that the human body surface from the camera view in each frame lies on the same plane over the whole sequence), we can choose initial correspondence from two image pairs in the given color and thermal image sequences. In this way, we have two pairs of initial correspondence: two points from the thermal images are reference points and two points from the color images are initial model parameters whose exact values need to be estimated. If the two pairs of points are chosen from a small local area, the resulting registration performance may not be globally satisfied in other areas. To avoid this problem, these points should be located as far away as possible in the images.

Hierarchical Genetic Algorithm for Model Parameter Estimation

Given initial corresponding points $\mathbf{x}_{1,0} = (x_{1,0}, y_{1,0})$ and $\mathbf{x}_{2,0} = (x_{2,0}, y_{2,0})$, and initial side length of the square search window w_1, pre-selected threshold w_l of minimum square window size, and N pairs of preliminarily extracted color and thermal silhouette images;

1. Apply GA in the search windows ($w_1 \times w_1$) centered at $\mathbf{x}_{1,0}$ and $\mathbf{x}_{2,0}$ at search level 1;
2. Obtain the estimated corresponding points $\mathbf{x}_{1,1}$ and $\mathbf{x}_{2,1}$ after the GA terminates;
3. Let $k = 2$;
4. Calculate the side length of search widows w_k by Equation (8);
5. Apply GA in the search windows ($w_k \times w_k$) centered at $\mathbf{x}_{1,k-1}$ and $\mathbf{x}_{2,k-1}$ at search level k;
6. Obtain the estimated corresponding points $\mathbf{x}_{1,1}$ and $\mathbf{x}_{2,1}$ after the GA terminates;
7. If $w_k < w_l$, go to step 5; otherwise, output $\mathbf{x}_{1,k}$ and $\mathbf{x}_{2,k}$.

Fig. 7. Pseudo-code for parameter estimation based on hierarchical genetic algorithm.

Fig. 8. Examples of registration results: first row—original color images, second row—original thermal images, third row—transformed color images.

3.3.2. Parameter estimation based on HGA

We use GA to solve the optimization problem in Eq. (6). GA provides a learning method motivated by an analogy to biological evolution. Rather than search from general-to-specific hypotheses, or from simple-to-complex hypotheses, GA generates successor hypotheses by repeatedly mutating and recombining parts of the best currently known hypotheses. At each step, a collection of hypotheses called the current population is updated by replacing some fraction of the population by offspring of the most fit current hypotheses. After a large number of steps, the hypotheses having the best fitness are considered as solutions. However, a single GA is not appropriate to estimate the subpixel location of corresponding points in given search windows. The accurate subpixel corresponding point search requires longer bit length for each coordinate value of each point. As a result, the population size of GA need to be large to reduce the probability of falling into a local maxima. Due to the costly fitness function (6), the large population size is not desirable. In this paper, we propose a HGA-based search scheme to estimate the model parameters within a series of windows

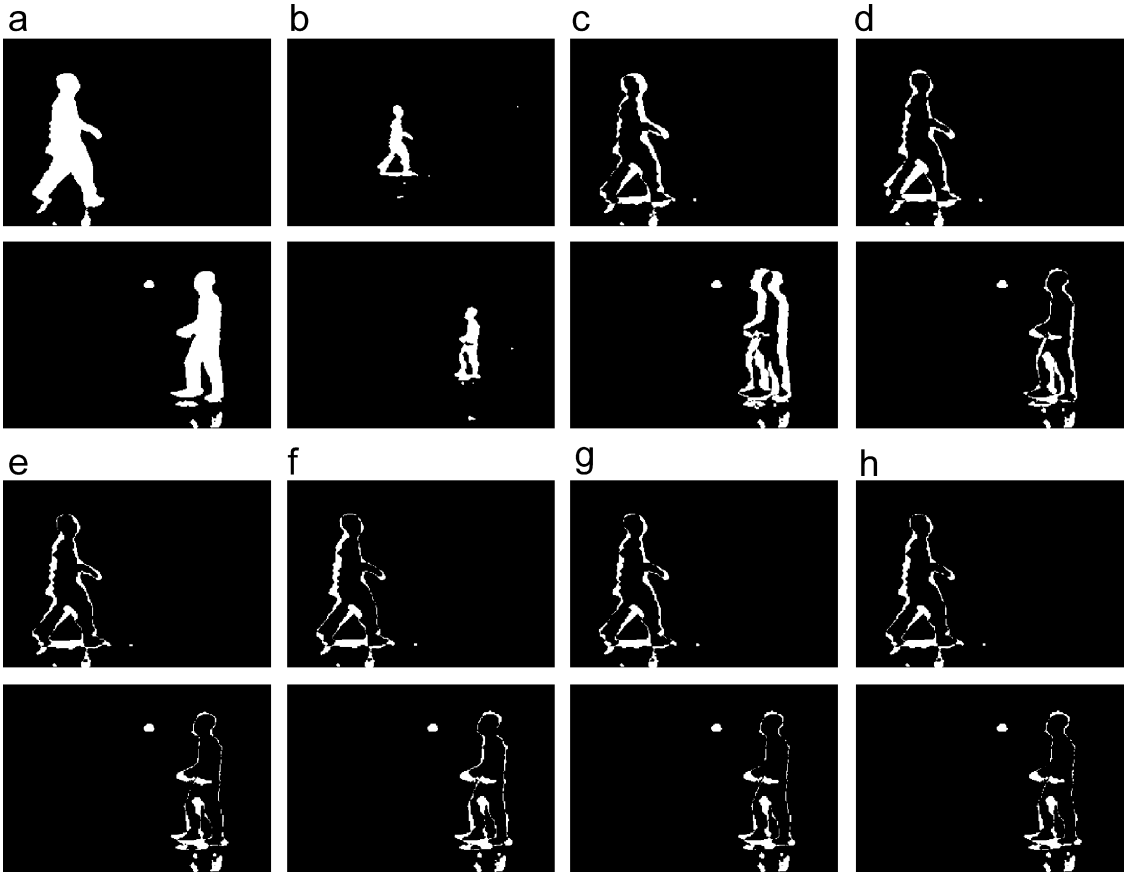


Fig. 9. Examples of estimated transformation results from good initial correspondence: (a) original silhouettes from thermal images, (b) original silhouettes from color images, (c) matching error of the initial transformation, (d) matching error after the first search level, (e) after the second search level, (f) after the third search level, (g) after the fourth search level, (h) after the 12th search level.

with adaptively reduced size as shown in Fig. 6. The model parameters are coordinate values of the two points in the color image plane, corresponding to the two reference points in the thermal image plane.

We choose the estimated human silhouette centroids (as mentioned in Section 3.2) from two thermal images in the same IR video as two reference points in the thermal image plane as shown in Fig. 5. Let $\mathbf{p} = [x_1, y_1, x_2, y_2]^{\mathrm{T}}$ be the model parameter to be estimated, where $\mathbf{x}_1 = (x_1, y_1)$ and $\mathbf{x}_2 = (x_2, y_2)$ are corresponding points to be searched in the color image plane (Fig. 6). The estimated two centroids from the two synchronized color images, $\mathbf{x}_{1,0} = (x_{1,0}, y_{1,0})$ and $\mathbf{x}_{2,0} = (x_{2,0}, y_{2,0})$, are chosen as the initial corresponding points in the color image plane. At each search level of the HGA-based search scheme, GA is applied to estimate the two corresponding coordinates according to Eq. (6). The center and size of search windows are both determined by the previous three estimates of corresponding points. In the kth search level, the centers of the two search windows for the two corresponding points are chosen as follows:

$$\mathbf{c}_{i,k} = \begin{cases} \mathbf{x}_{i,0} & \text{if } k = 1, \\ (\mathbf{x}_{i,0} + \mathbf{x}_{i,1})/2 & \text{if } k = 2, \\ (\mathbf{x}_{i,k-3} + \mathbf{x}_{i,k-2} + \mathbf{x}_{i,k-1})/3 & \text{if } k > 2, \end{cases} \tag{7}$$

where $i = 1, 2$, $\mathbf{x}_{i,j}$ is the new estimate of $\mathbf{x}_i$ after the jth search level. The length of the search windows (square in shape) is chosen as follows:

$$w_k = \begin{cases} w_1 & \text{if } k = 1, \\ \max_{j=1}^{4}\{|p_{j,0} - p_{j,1}|\} & \text{if } k = 2, \\ \max_{j=1}^{4}\{\max\{p_{j,k-3}, p_{j,k-2}, p_{j,k-1}\} & \\ \quad - \min\{p_{j,k-3}, p_{j,k-2}, p_{j,k-1}\}\} & \text{if } k > 2, \end{cases} \tag{8}$$

where w_1 is the preselected initial length of the search window. This iterative procedure is repeated (see pseudo code in Fig. 7) until the search w_k is lower than a pre-selected lower limit w_l.

In the proposed approach, the code length of parameters in each GA can be small without decreasing the final estimation accuracy. Considering the costly fitness function in our application, the population size cannot be large. Short code length is desired because a GA with high ratio of code length over population size has a high probability of falling into the local maximum. Generally, the window size will be adaptively reduced until reaching the lower limit. Even if the real correspondence exists outside of the initial search window, the approach still have the possibility to find a good estimate because the new window might cover areas outside of the initial window. After the correspondences in the color image plane are located,

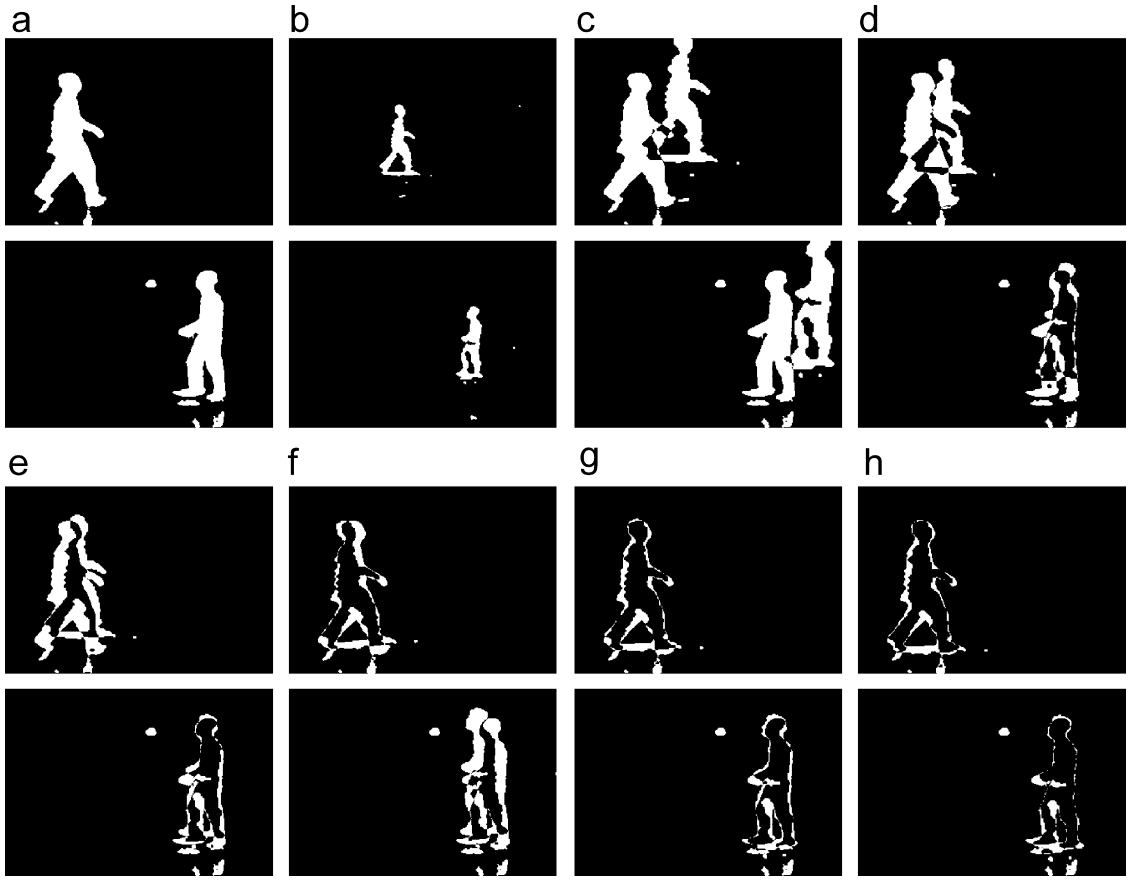


Fig. 10. Examples of estimated transformation results from bad initial correspondence: (a) original silhouettes from thermal images, (b) original silhouettes from color images, (c) matching error of the initial transformation, (d) matching error after the first search level, (e) after the second search level, (f) after the third search level, (g) after the fourth search level, (h) after the 23th search level.

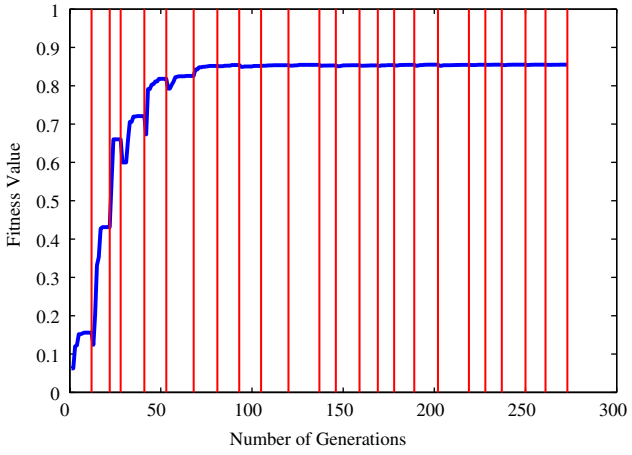


Fig. 11. Variation of fitness values from bad initial correspondence. The vertical line corresponds to the last generation at each search level. The curve between two adjacent vertical lines indicates the variation of GA fitness values at a search level.

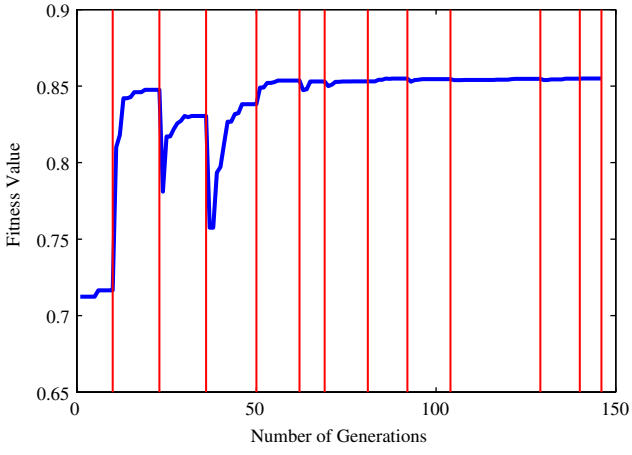


Fig. 12. Variation of fitness values from good initial correspondence. The vertical line corresponds to the last generation at each search level. The curve between two adjacent vertical lines indicates the variation of GA fitness values at a search level.

the transformation is uniquely determined for this pair of color and thermal image sequences, and it will be used to transform color images into the plane of thermal images.

3.4. Sensor fusion

To improve the accuracy of human silhouette extraction, we combine the information from the registered color and thermal

images. If the human silhouette extraction is viewed as a classification procedure, the commonly used classifier combination strategies can be employed here. Kittler et al. [26] demonstrate that the commonly used classifier combination schemes can be derived from a uniform Bayesian framework under different assumptions and using different approximations. The product rule assumes that the measurements used are conditionally statistically independent. The sum rule further assumes that the a posteriori probability computed by the respective classifiers will not deviate dramatically from the prior probabilities. The max rule is derived from the sum rule by approximating the sum by the maximum of the posterior probabilities under the assumption of equal priors. The min rule is derived from the product rule under the assumption of equal priors. Similar fusion strategies can be applied for human body silhouette detection by combining registered color and thermal images as follows:

- Product rule: $(X, Y) \in S$, if $P(S|\mathbf{c}(X, Y))P(S|t(X, Y)) > \tau_{product}$,
- Sum rule: $(X, Y) \in S$, if $P(S|\mathbf{c}(X, Y)) + P(S|t(X, Y)) > \tau_{sum}$,
- Max rule: $(X, Y) \in S$, if $\max\{P(S|\mathbf{c}(X, Y)), P(S|t(X, Y))\} > \tau_{max}$,
- Min rule: $(X, Y) \in S$, if $\min\{P(S|\mathbf{c}(X, Y)), P(S|t(X, Y))\} > \tau_{min}$,

where (X, Y) represents the 2-D image coordinate, S represents the human silhouette, $\mathbf{c}$ represents the color value vector, t represents the thermal value and $\tau_{product}$, τ_{sum}, τ_{max} and τ_{min} are thresholds described in the next section. The estimate of probability is computed as

$$P(S|c(X, Y)) = 1 - e^{-\|\mathbf{c}(X,Y)-\mu_{\mathbf{c}}(X,Y)\|^2}, \quad (9)$$

$$P(S|t(X, Y)) = 1 - e^{-|t(X,Y)-\mu_t(X,Y)|^2}, \quad (10)$$

where $\mu_{\mathbf{c}}$ represents the mean background color value vector, and μ_t represents the mean background thermal value.

3.5. Registration of EO/IR sequences with multiple objects

The proposed approach for registration of EO/IR imagery is presented for the single-object scenario. Without losing any generality, we can assume that both the EO and IR cameras are fixed for a period of time, and a pair of EO/IR image sequences, containing a single object, are available for registration at the beginning. Then, the estimated transformation model presented in this paper can be used for image registration from subsequent synchronous EO/IR sequence pairs under the same camera setup and it does not matter how many objects are present in these sequences. Therefore, multiple moving objects are allowed for registration and detection under the same camera setup.

Table 1
Confusion matrix

	Ground Truth foreground	Ground Truth background
Detected foreground	$N - \alpha$	β
Detected background	α	$B - \beta$

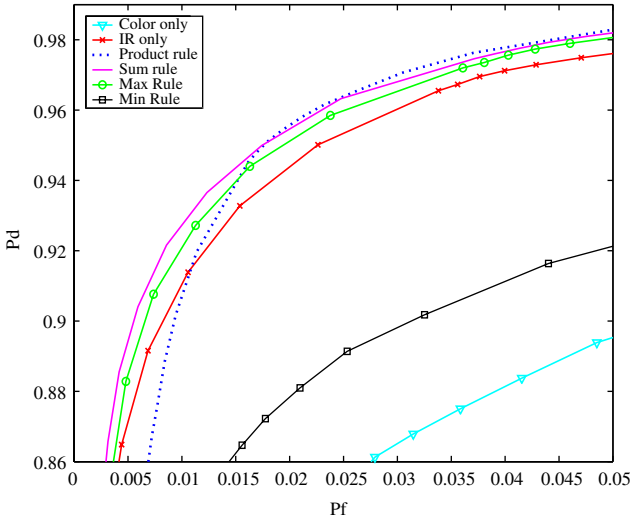


Fig. 13. ROC curves for detection performance evaluation of different fusion strategies for silhouette detection.

4. Experimental results

The image data used in our experiments are real human walking data recorded by the two cameras in the same indoor environment. Color images are recorded by a PC camera with image size of 240 × 360 as shown in the first row of Fig. 8. Thermal images are recorded by a long-wave IR camera with image size of 240 × 360 as shown in the third row of Fig. 8. Both cameras have fixed but different focal lengths. The IR camera has a narrower field-of-view and a higher resolution than the color camera. It has less distortion than the color camera, and, therefore, it is used as the base camera. The color images are transformed and then fused with the original thermal images in our experiments for human silhouette detection.

4.1. Image registration results

Three color and thermal images are selected for matching by Eq. (6). In our experiments, we choose $\alpha = \beta = 15$ in Eqs. (4) and (5). The initial search window is set as 16 × 16 pixels ($w_1 = 16$), and the final search window is 0.1 × 0.1 pixels ($w_l = 0.1$). In the GA at each search level, we use 6 bits to represent each coordinate value (totally 24 bits for 4 coordinate values); fitness function is the similarity between image pairs in Eq. (6); population size is 100; crossover rate is 0.9; crossover method is uniform crossover; mutation rate is 0.05; the GA will terminate if the fitness values have not changed for five successive steps.

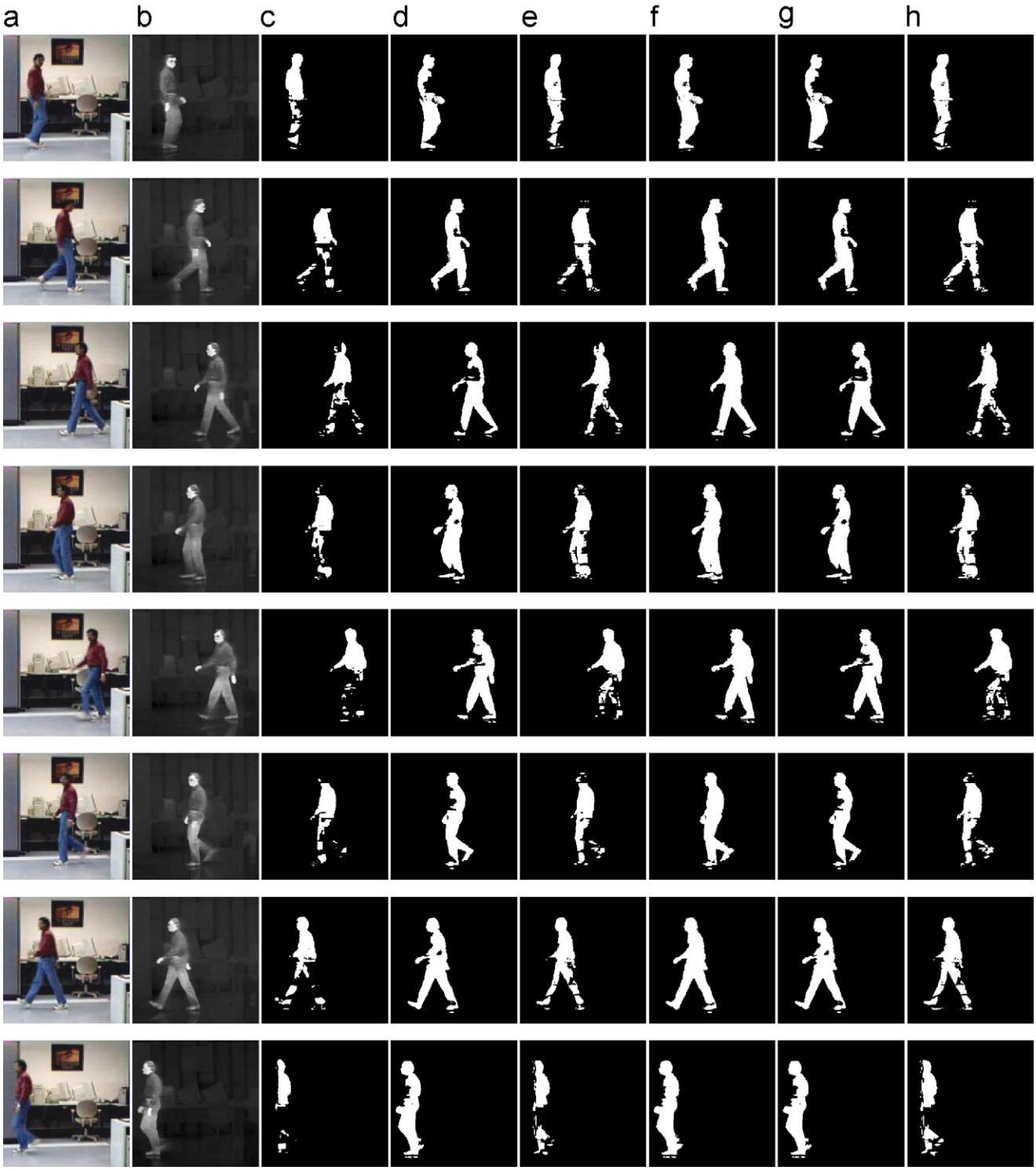


Fig. 14. Examples of fusion results: (a) transformed color images, (b) original thermal images, (c) silhouette from (a), (d) silhouette from (b), (e) silhouette from product rule fusion, (f) silhouette from sum rule fusion, (g) silhouette from max rule fusion, (h) silhouette from min rule fusion.

Fig. 9 shows examples of estimated transformation results from good initial corresponding points at different search levels, while Fig. 10 shows results from bad initial corresponding points. Even though the original transformation 10(c) is far away from the true transformation, the transformation results are improved gradually at successive search levels and finally converged around the real transformation. The variations of fitness values from bad and good initial correspondence are shown in Figs. 11 and 12, respectively. The vertical line corresponds to the last generation at each search level. The curve between

two adjacent vertical lines indicates the variation of GA fitness values in a search level. In the GA at each search level, the populations are randomly generated, leading to the drop of the fitness value at the beginning of each search level. We do not use the population from the previous search level because we hope to estimate transformation parameters more accurately as the window size decreases and diversify the population to avoid premature convergence. In general, our image registration approach is not sensitive to the location of initial correspondence if it is located inside or slightly outside of the initial search window depending on the size of the initial search window.

Fig. 8 shows the comparison of original color images, transformed color images and original thermal images. To evaluate the registration performance, we define the registration precision as $P(A, B) = (A \cap B)/(A \cup B)$, where A and B are manually labeled human silhouette pixel sets from the original thermal image and the transformed color image, respectively. According to this definition, the registration precision for the three image pairs in Fig. 8 is 78%, 80% and 85%, respectively. Considering that the color and thermal image pairs are not exactly synchronized, and there are possible human labeling errors due to the physical difference between color and thermal signals, our image registration still achieves good results.

4.2. Sensor fusion results

We evaluate the human silhouette extraction performance by the receiver operating characteristic (ROC) curves [27]. Let N be the number of moving object pixels in the Ground Truth images, α be the number of moving object pixels that the algorithm did not detect, B be the number of background pixels in the Ground Truth image, and β be the number of background pixels that were detected as foreground. The Ground Truth image in our experiments is manually labeled from the original thermal images. The confusion matrix is given in Table 1. We can define the probability of detection and probability of false alarms as

$$Pd = (N - \alpha)/N \quad \text{and} \quad Pf = \beta/B. \tag{11}$$

If we evaluate M images in their entirety, the equations become

$$Pd = \sum_{i=1}^{M}(N_i - \alpha_i) \Big/ \sum_{i=1}^{M} N_i \quad \text{and}$$
$$Pf = \sum_{i=1}^{M} \beta_i \Big/ \sum_{i=1}^{M} B_i. \tag{12}$$

Equation given in (12) are used to obtain the ROC curves for detection performance evaluation of different fusion strategies which are shown in Fig. 13. This figure shows that the product, sum and max fusion rules achieve better results than using color or thermal classifiers individually. Among these rules, the sum rule achieves the best results. Considering that the image resolution of the thermal camera is higher than that of the EO camera, the thermal classifier has much higher confidence than the color classifier. We believe that the main reason for the good performance achieved by sum rule is its robustness to errors (or noise) especially from the color classifier [26]. Product rule considers more color information, so it is sensitive to the noise from color classifier especially when the false alarm is low. Max rule considers less color information with low confidence, so its performance is higher than that of the thermal classifier but lower than sum rule. The performance of min rule is even worse than that of using thermal information only because it mainly focuses on the color information with low confidence. Fig. 14 shows the human silhouette extraction results by combining color and thermal image pairs with different strategies described in Section 3.4. The threshold for each of these rules is chosen as the smallest value such that shadows in both color and thermal images are eliminated. These thresholds are held constant ($\tau_{product} = 0.1$, $\tau_{sum} = 0.9$, $\tau_{max} = 0.9$ and $\tau_{min} = 0.1$) for all the experiments reported in this paper.

5. Conclusions

In this paper, we approach the task of human silhouette extraction from color and thermal image sequences using automatic image registration. A hierarchical Genetic Algorithm-(HGA) based scheme is employed to find correspondence so that the preliminary silhouettes form the color and thermal images are well matched. HGA estimates the model parameters within a series of windows with adaptively reduced size at different levels. The obtained correspondence and corresponding transformation are used for image registration in the same scene.

Registered color and thermal images are combined by probabilistic strategies to obtain better body silhouette extraction results. Experiments show that (1) the proposed approach achieves good performance for image registration between color and thermal image sequences, and (2) each of the product, sum and max fusion rules achieves better performance on silhouette detection than color or thermal images used individually. Among these rules, sum rule achieves the best results.

References

[1] S. Niyogi, E. Adelson, Analyzing and recognizing walking figures in XYT, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1994, pp. 469–474.
[2] J. Little, J. Boyd, Recognizing people by their gait: the shape of motion, Videre: J. Comput. Vision Res. 1 (2) (1998) 1–32.
[3] P. Huang, C. Harris, M. Nixon, Recognizing humans by gait via parametric canonical space, Artif. Intell. Eng. 13 (1999) 359–366.
[4] A. Kale, A. Rajagopalan, N. Cuntoor, V. Kruger, Gait-based recognition of humans using continuous HMMS, in: Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition, 2002, pp. 321–326.
[5] P. Phillips, S. Sarkar, I. Robledo, P. Grother, K. Bowyer, The gait identification challenge problem: data sets and baseline algorithm, in: Proceedings of the International Conference on Pattern Recognition, vol. 1, 2002, pp. 385–388.
[6] D. Tolliver, R. Collins, Gait shape estimation for identification, in: Proceedings of the Fourth International Conference on Audio- and Video-Based Biometric Person Authentication, Springer Lecture Notes in Computer Science, vol. 2688, Springer, Berlin, 2003, pp. 734–742.
[7] S. Nadimi, B. Bhanu, Physical models for moving shadow and object detection in video, IEEE Trans. Pattern Anal. Mach. Intell. 26 (8) (2004) 1079–1087.

[8] R. Santiago-Mozos, J. Leiva-Murillo, F. Perez-Cruz, A. Artes-Rodriguez, Supervised-PCA and SVM classifiers for object detection in infrared images, in: Proceedings of the IEEE Conference on Advanced Video and Signal Based Surveillance, 2003, pp. 122–127.
[9] H. Nanda, L. Davis, Probabilistic template based pedestrian detection in infrared videos, in: Proceedings of the IEEE Intelligent Vehicle Symposium, vol. 1, 2002, pp. 15–20.
[10] B. Bhanu, J. Han, Kinematic-based human motion analysis in infrared sequences, in: Proceedings of the IEEE Workshop on Applications of Computer Vision, 2002, pp. 208–212.
[11] H. Arlowe, Thermal detection contrast of human targets, in: Proceedings of the IEEE International Carnahan Conference on Security Technology, 1992, pp. 27–33.
[12] G. Clark, S. Sengupta, M. Buhl, R. Sherwood, P. Schaich, N. Bull, R. Kane, M. Barth, D. Fields, M. Carter, Detecting buried objects by fusing dual-band infrared images, 1993 Conference Record of the 27th Asilomar Conference on Signals, Systems and Computers, vol. 1, 1993, pp. 135–143.
[13] J. Perez-Jacome, V. Madisetti, Target detection from coregistered visual-thermal-range images, in: Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 4, 1997, pp. 2741–2744.
[14] H. Li, B. Manjunath, S. Mitra, A contour-based approach to multisensor image registration, IEEE Trans. Image Process. 4 (3) (1995) 320–334.
[15] J. Inglada, F. Adragna, Automatic multi-sensor image registration by edge matching using genetic algorithms, in: Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, vol. 5, 2001, pp. 2313–2315.
[16] M. Ali, D. Clausi, Automatic registration of SAR and visible band remote sensing images, in: Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, vol. 3, 2002, pp. 1331–1333.
[17] P. Dare, I. Dowman, Automatic registration of SAR and spot imagery based on multiple feature extraction and matching, in: Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, vol. 7, 2000, pp. 2896–2898.
[18] J. Wilder, P. Phillips, C. Jiang, S. Wiener, Comparison of visible and infra-red imagery for face recognition, in: Proceedings of the International Conference on Automatic Face and Gesture Recognition, 1996, pp. 182–187.
[19] Y. Yoshitomi, S.-I. Kim, T. Kawano, T. Kilazoe, Effect of sensor fusion for recognition of emotional states using voice, in: Proceedings of the IEEE International Workshop on Robot and Human Interactive Communication, 1996, pp. 178–183.
[20] Q. Zheng, R. Chellappa, A computational vision approach to image registration, IEEE Trans. Image Process. 2 (3) (1993) 311–326.
[21] H. Li, Y. Zhou, Automatic EO/IR sensor image registration, in: Proceedings of the International Conference on Image Processing, vol. 3, 1995, pp. 240–243.
[22] H. Li, Y. Zhou, R. Chellappa, SAR/IR sensor image fusion and real-time implementation, Record of the 29th Asilomar Conference on Signals, Systems and Computers, vol. 2, 1995, pp. 1121–1125.
[23] V. Mandava, J. Fitzpatrick, D.I. Pickens, Adaptive search space scaling in digital image registration, IEEE Trans. Med. Imaging 8 (3) (1989) 251–262.
[24] P. van den Elsen, E.-J. Pol, M. Viergever, Medical image matching—a review with classification, IEEE Eng. Med. Biol. Mag. 12 (1) (1993) 26–39.
[25] J. Yao, Image registration based on both feature and intensity matching, in: Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 3, 2001, pp. 1693–1696.
[26] J. Kittler, M. Hatef, R. Duin, J. Matas, On combining classifiers, IEEE Trans. Pattern Anal. Mach. Intell. 20 (3) (1998) 226–239.
[27] S. Nadimi, B. Bhanu, Multistrategy fusion using mixture model for moving object detection, in: Proceedings of the International Conference on Multisensor Fusion and Integration for Intelligent Systems, 2001, pp. 317–322.

About the Author—JU HAN received the Ph.D. degree in electrical engineering from the University of California at Riverside in 2005. Currently, he is a postdoctoral fellow in the Life Science Division at Lawrence Berkeley National Laboratory. His research interests include biometrics, biological image understanding and computational biology.

About the Author—BIR BHANU received the S.M. and E.E. degrees in electrical engineering and computer science from the Massachusetts Institute of Technology, Cambridge, the Ph.D. degree in electrical engineering from the Image Processing Institute, University of Southern California, Los Angeles and the M.B.A. degree from the University of California, Irvine.

Dr. Bhanu has been the founding professor of electrical engineering and served its first Chair at the University of California at Riverside (UCR). He has been the cooperative professor of computer science and engineering and director of Visualization and Intelligent Systems Laboratory (VISLab) since 1991. Currently, he also serves as the founding director of an interdisciplinary Center for Research in Intelligent Systems (CRIS) at UCR. Previously, he was a Senior Honeywell Fellow at Honeywell Inc., Minneapolis, MN. He has been on the faculty of the Department of Computer Science at the University of Utah, Salt Lake City, and has worked at Ford Aerospace and Communications Corporation, CA, INRIA-France, and IBM San Jose Research Laboratory, CA. He has been the principal investigator of various programs for DARPA, NASA, NSF, AFOSR, ARO and other agencies and industries in the areas of learning and vision, image understanding, pattern recognition, target recognition, biometrics, navigation, image databases and machine vision applications. He is the coauthor of *Evolutionary Synthesis of Pattern Recognition Systems* (New York: Springer, 2005), *Computational Algorithms for Fingerprint Recognition* (Norwell, MA: Kluwer, 2004), *Genetic Learning for Adaptive Image Segmentation* (Norwell, MA: Kluwer, 1994) and *Qualitative Motion Understanding* (Norwell, MA; Kluwer, 1992), and the coeditor of *Computer Vision Beyond the Visible Spectrum* (New York: Springer, 2004). He holds 11 U.S. and international patents and over 250 reviewed technical publications in the areas of his interest.

Dr. Bhanu has received two outstanding paper awards from the Pattern Recognition Society and has received industrial and university awards for research excellence, outstanding contributions and team efforts. He has been on the editorial board of various journals and has edited special issues of several IEEE transactions and other journals. He has been General Chair for the IEEE Conference on Computer Vision and Pattern Recognition, IEEE Workshops on Applications of Computer Vision, IEEE Workshops on Learning in Computer Vision and Pattern Recognition; Chair for the DARPA Image Understanding Workshop, and Program Chair for the IEEE Workshops on Computer Vision Beyond the Visible Spectrum. He is a Fellow of the American Association for the Advancement of Science (AAAS), Institute of Electrical and Electronics Engineers (IEEE), International Association of Pattern Recognition (IAPR) and the International Society for Optical Engineering (SPIE).