

Article

Multi-Sensor Face Registration Based on Global and Local Structures

Wei Li ^{1,*}, Mingli Dong ^{2,*}, Naiguang Lu ², Xiaoping Lou ² and Wanyong Zhou ¹

¹ Hebei Engineering Research Center for Assembly and Inspection Robot, North China Institute of Aerospace Engineering, Langfang 065000, China; zhouwanyong@nciae.edu.cn

² Key Laboratory of the Ministry of Education for Optoelectronic Measurement Technology and Instrument, Beijing Information Science and Technology University, Beijing 100192, China; nglv2002@sina.com (N.L.); louxiaoping@bistu.edu.cn (X.L.)

* Correspondence: liweikilary@163.com (W.L.); dongml@bistu.edu.cn (M.D.); Tel.: +86-10-8242-6892 (W.L.)

Received: 10 September 2019; Accepted: 28 October 2019; Published: 30 October 2019



Abstract: The work reported in this paper aims at utilizing the global geometrical relationship and local shape feature to register multi-spectral images for fusion-based face recognition. We first propose a multi-spectral face images registration method based on both global and local structures of feature point sets. In order to combine the global geometrical relationship and local shape feature in a new Student's t Mixture probabilistic model framework. On the one hand, we use inner-distance shape context as the local shape descriptors of feature point sets. On the other hand, we formulate the feature point sets registration of the multi-spectral face images as the Student's t Mixture probabilistic model estimation, and local shape descriptors are used to replace the mixing proportions of the prior Student's t Mixture Model. Furthermore, in order to improve the anti-interference performance of face recognition techniques, a guided filtering and gradient preserving image fusion strategy is used to fuse the registered multi-spectral face image. It can make the multi-spectral fusion image hold more apparent details of the visible image and thermal radiation information of the infrared image. Subjective and objective registration experiments are conducted with manual selected landmarks and real multi-spectral face images. The qualitative and quantitative comparisons with the state-of-the-art methods demonstrate the accuracy and robustness of our proposed method in solving the multi-spectral face image registration problem.

Keywords: multi-sensor; face registration; inner-distance; Student's-t Mixtures Model; image fusion

1. Introduction

Image fusion can often analyze and extract the complementary information of multi-sensor data, and compose a robust or informative image, which can provide more complex and detailed target scene representation [1–3]. Since the fusion process of multi-sensor images is in accordance with human visual perception, it has an important research significance for object detection and target recognition in the areas of remote sensing image, medical imaging analysis, military target detection and video surveillance [4–6]. E.g., face recognition is one of the major applications of multi-sensor data, such as visible and thermal infrared images. Visible images can provide face texture details with high spatial resolution, while thermal infrared images are not likely to be interfered with by illumination variation or face disguise with high thermal contrast. Therefore, it is beneficial for face recognition to fuse the multi-sensor data, which can combine the advantages of texture detail and thermal radiation information in the multi-spectral face images.

However, image registration is a prerequisite for the success of multi-spectral image fusion, which is an essential and challenging step in the process of image fusion research [7]. Currently, multi-spectral

images registration can generally be implemented in two ways: Hardware-based registration and software-based registration. In order to ensure that the multi-spectral images are strictly geometrically aligned, hardware-based registration can be realized based on the coaxial catadioptric optical system through the beam splitter [8]. But the sophisticated and cumbersome imaging sensor equipment for generating co-registered image pairs may not be practical in many face recognition scenarios, due to its high cost and low availability. In contrast, software-based registration can capture dual band images simultaneously and independently by utilizing different off-the-shelf low cost wavebands sensors. Since no additional hardware is needed, it may be more appropriate for practical detection and recognition scenarios compared to the hardware-based registration [9]. In general, software-based registration can be classified into two categories: Area-based and feature-based approaches [10]. Area-based methods attempt to deal directly with the image intensity values to detect, for instance, cross-correlation [11], Fourier transform and mutual information [12,13]. On the contrary, feature-based methods attempt to indirectly extract salient structures or features in the images, such as points of high curvature, corners, strong edges, intersections of lines, structural contours and silhouettes within the images [14–16]. For visible and thermal infrared image pairs usually involve quite different intensity values, and the area-based methods are good usually only on a selected small region of the images. Instead, salient structures with a strong edge could often represent a significant common feature of the heterogeneous images. In view of this, this paper focuses on the feature-based registration methods for the visible and thermal infrared face images.

Feature-based registration methods first extract different feature point sets of salient structures in the multi-spectral images, and then the registration problem can be simplified to determine the correct correspondence and the inherent spatial transformation between two point sets of extracted features. Unlike the face recognition problem [17,18], the multi-sensor face registration problem has not enough common features available to build a convolutional neural network (CNNs) framework to predict the classification model; a popular strategy is to regard the alignment of two point sets as probability density estimation, e.g., the Gaussian Mixture Model (GMM) [19–21] and Student's t Mixture Model (SMM) [22–24]. Because GMM formulations are intensively used, and the heavily tailed SMM is more robust and precise against noise and outliers than the classical GMM, it is a potential research direction to take advantage of SMM for image registration methods [25–27]. In general, both GMM and SMM probabilistic methods are intended to exploit global relationships in the point sets. However, the neighborhood structures among the feature points also contribute to the alignment of two point sets. Thus, another interesting point matching strategy would be aimed first at using local neighborhood structures as the feature descriptor to recover the feature point correspondences, e.g., Shape Context (SC) [28] and Inner-Distance Shape Context (IDSC) [29]. But for multi-spectral face images registration, the local neighborhood structures of feature points are not discriminative enough, because it will inevitably include a number of mismatching points. To address this issue, in this paper we modify the SMM probabilistic framework and take full advantage of the global and local structures during the multi-spectral face images registration process.

Our contribution in this paper includes the following three aspects: Firstly, in order to keep the global and local structures of feature point sets in the registration process, we use IDSC as the local shape descriptors of feature point sets, and treat the confidence of feature matching as the mixing proportion of the finite mixture model, which aims to describe the local shape and global geometrical relationship of feature point sets in a new SMM probabilistic framework. Secondly, in order to make the multi-spectral fusion image hold more apparent details of a visible image and the thermal radiation information of an infrared image, we propose a guided filtering and gradient preserving image fusion strategy for multi-spectral face images fusion. It can improve the robustness of face recognition techniques in varying illumination conditions. Thirdly, we construct a simple multi-spectral imaging system in order to acquire higher resolution multi-spectral face images from visible and thermal infrared cameras simultaneously.

2. The Proposed Registration Method for Visible and Infrared Face Images

For the software-based registration problem, a basic process of infrared and visible images fusion mainly includes the following procedures: Feature extraction and description, feature point sets registration, image transformation and interpolation and finally image fusion [1,30]. As is shown in Figure 1, the first step is to extract robust common salient features that might be preserved in the multi-spectral face images. Considering the fact that visible and infrared images are characterizations of two different modalities, we use edge maps to represent the common salient features of multi-spectral face images. After obtaining the salient features of multi-spectral images, the edge maps can usually be discretized as two feature point sets. Thus, the registration procedure is converted into determining the correct correspondence and estimating the spatial transformation between the feature point sets for multi-spectral face images. Furthermore, the other visible image points can be interpolated using a smoothing Thin-Plate Spline (TPS) [31] based on the pixel coordinates of the thermal infrared image. Finally, the multi-spectral image fusion process takes place in the output image, and the fused image can preserve both the apparent details of the visible image and the thermal radiation information of the infrared image, which will greatly improve the accuracy and efficiency of detection and recognition. In this paper, our attention mainly focuses on the following two aspects: Feature point sets registration and multi-spectral face images fusion.

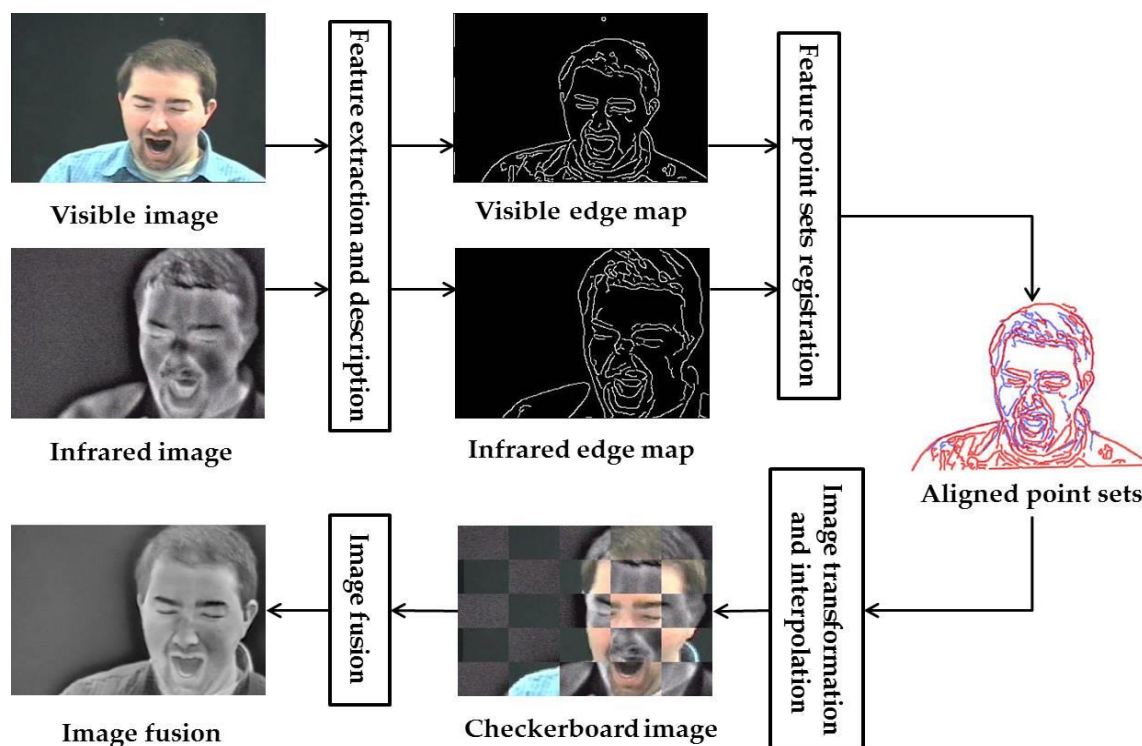


Figure 1. Visible and infrared face images fusion basic process for software-based registration.

2.1. Inner-Distance Shape Context for 2D Feature Descriptors

In order to establish the initial correspondences between the two feature point sets of visible and infrared face images, we first use IDSC as the local feature descriptors, and solve the point sets matching problem through a bipartite graph method [32]. Compared to the SC feature matching method, the IDSC feature matching method uses inner-distance to replace the Euclidean distance, which is robust to articulation shape and part structure. For example, Figure 2 shows the schematic illustration of face silhouette feature descriptors based on IDSC, in the IDSC feature histograms from Figure 2c, the horizontal axis $n_{\theta} = 12$ denotes the numbers of orientation bins and the vertical axis $n_d = 5$ denotes the numbers of logarithm distance bins.

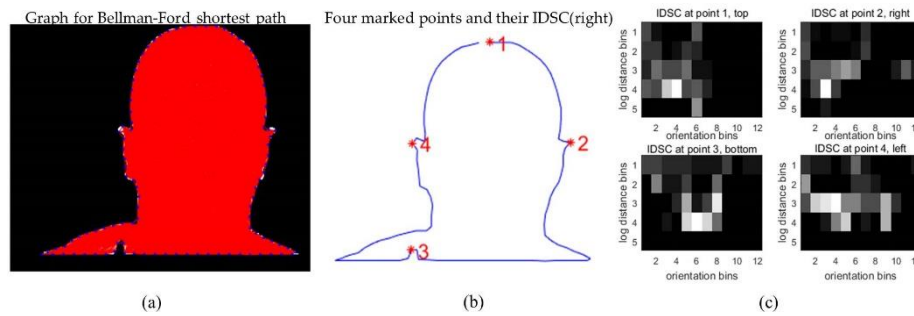


Figure 2. Schematic illustration of face silhouette feature descriptors based on Inner-Distance Shape Context (IDSC). (a) Bellman-Ford shortest path graph built using face silhouette landmark points; (b,c) Four marked points and their corresponding IDSC feature histograms.

As can be seen from Figure 2c, the IDSC feature histograms for all marked points are quite different, and the excellent discriminative performance demonstrates the inner-distance’s ability to capture local structures in the point sets. Consider a point p_i in the point set X and a point q_j in the point set Y . Let $C_{i,j}$ denote the cost of matching p_i to q_j , $h_{X,p_i}(k)$ and $h_{Y,q_j}(k)$ denote the normalized histogram of the relative coordinates of the remaining points, respectively, and K be the number of histogram bins. As the IDSC is represented based on the histogram, it can be measured by the χ^2 test statistic:

$$C_{i,j} = C(p_i, q_j) = \frac{1}{2} \sum_{k=1}^K \frac{[h_{X,p_i}(k) - h_{Y,q_j}(k)]^2}{h_{X,p_i}(k) + h_{Y,q_j}(k)}, \tag{1}$$

In general, the smaller the matching cost $C_{i,j}$ is, the more similar the local appearance at points p_i and q_j are. Once the cost matrix C for all correspondence points in the point set X and Y is obtained, the correspondences Ω between the two point sets are an instance of an assignment problem, which can be solved by the Hungarian method.

2.2. Student’s t Mixtures for Feature Point Set Registration

In order to fully exploit the global structures in the feature point sets, we treat the alignment of two point sets as a probability density estimation of an SMM distribution with the IDSC local feature descriptors. Compared to the GMM distribution, the main advantage of SMM is that Student’s t distribution has a heavier tail than the exponentially decaying tail of Gaussian distribution. As is shown in the Figure 3, each component of the SMM has an additional parameter ν called the degrees of freedom, which can change the shape of a Student’s t distribution curve, and hence SMM provides a more robust model than the classical GMM.

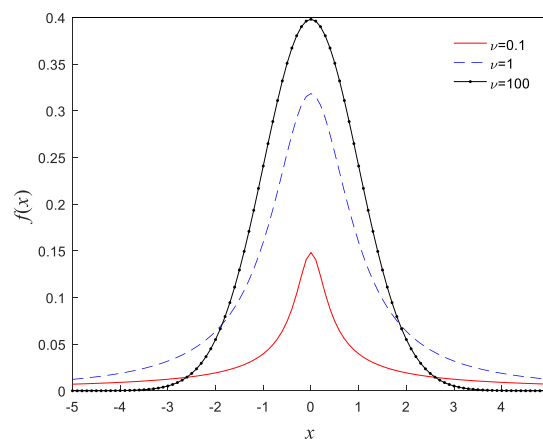


Figure 3. The Student’s t distribution for various degrees of freedom.

Given a point set $X_{N \times D} = (x_1, \dots, x_N)^T$, which can be considered as an observation datum, the other point set $Y_{M \times D} = (y_1, \dots, y_M)^T$ is treated as the SMM centroids, where N and M represent the number of points in X and Y , respectively, and D represents the dimension of each point. Then the goal of registration is to align the centroids point set Y to the observation data point set X . Typically, the point sets contain noise and outliers, which can be supposed to be a uniform distribution $1/N$, and the weight of the uniform distribution is denoted as $\eta \in [0, 1]$. Let $w_{ij} \in [0, 1]$ denote a mixing proportion corresponding to the i^{th} component of the SMM for point x_j ($\sum_{i=1}^M w_{ij} = 1$), then the mixture probability density function can take the form:

$$f(x|\psi) = \eta \frac{1}{N} + (1 - \eta) \sum_{i=1}^M \omega_{ij} f(x_j|y_i, \Sigma_i, v_i), \tag{2}$$

where $\psi = (\psi_1, \psi_2, \dots, \psi_M)$ is mixture parameter set with $\psi_i = (w_{ij}, y_i, \Sigma_i, v_i)$, and $f(x_j|y_i, \Sigma_i, v_i)$ is the Student's t distribution probability density function for the i^{th} component of SMM; we express that:

$$f(x_j|y_i, \Sigma_i, v_i) = \frac{\Gamma\left(\frac{v_i+D}{2}\right) |\Sigma_i|^{-\frac{1}{2}}}{(\pi v_i)^{\frac{1}{2}D} \Gamma\left(\frac{v_i}{2}\right) \left[1 + \frac{d(x_j, y_i; \Sigma_i)}{v_i}\right]^{\frac{v_i+D}{2}}}, \tag{3}$$

where $d(x_j, y_i; \Sigma_i)$ is the Mahalanobis squared distance between observation point x_j and centroid point y_i , while Σ_i , Γ and v represent the covariance matrix, Gamma function and degrees of freedom, respectively.

2.3. Multi-Spectral Face Registration Using Global and Local Structures

In order to estimate the mixture parameter ψ in the mixture probability density function (2), according to [22,23], we first rewrite the function (2) as a complete data logarithm likelihood function $\ln L_C(\psi) = \sum_{j=1}^N \ln \sum_{i=1}^M w_{ij} f(x_j|y_i, \Sigma_i, v_i)$, and then use the Expectation Maximization (EM) to train the logarithm likelihood function. In general, the EM algorithm can be divided into the following two steps: Expectation step (E-step) and Maximization step (M-step), and the iterative process proceeds by alternating between the E- and M-steps until convergence.

E-step: We first use the current parameter $\psi^{(k)}$ to calculate the posterior probability $\tau_{ij}^{(k)}$, and according to the Bayesian Theorem, it can be expressed in terms of the observation data x_j belonging to the i^{th} component of the SMM

$$\tau_{ij}^{(k)} = \frac{w_{ij}^{(k)} f(x_j|y_i^{(k)}, \Sigma_i^{(k)}, v_i^{(k)})}{f(x_j|\psi^{(k)})}, \tag{4}$$

Next we need to calculate the k^{th} iteration conditional expectation of the logarithm likelihood function $\ln L_C(\psi)$, which can be calculated as follows:

$$Q(\psi|\psi^{(k)}) = Q_1(w|\psi^{(k)}) + Q_2(v|\psi^{(k)}) + Q_3(y, \Sigma|\psi^{(k)}), \tag{5}$$

where Q_1 , Q_2 and Q_3 are respectively written as follows:

$$\begin{aligned}
 Q_1(\boldsymbol{w}|\boldsymbol{\psi}^{(k)}) &= \sum_{j=1}^N \sum_{i=1}^M \tau_{ij}^{(k)} \ln w_{ij} \\
 Q_2(\boldsymbol{v}|\boldsymbol{\psi}^{(k)}) &= \sum_{j=1}^N \sum_{i=1}^M \tau_{ij}^{(k)} \left(-\ln \Gamma\left(\frac{v_i}{2}\right) + \frac{v_i}{2} \ln\left(\frac{v_i}{2}\right) + \frac{v_i}{2} \left\{ \sum_{j=1}^N (\ln u_{ij}^{(k)} - u_{ij}^{(k)}) + \gamma\left(\frac{v_i^{(k)}+D}{2}\right) - \ln\left(\frac{v_i^{(k)}+D}{2}\right) \right\} \right), \\
 Q_3(\boldsymbol{y}, \boldsymbol{\Sigma}_i|\boldsymbol{\psi}^{(k)}) &= \sum_{j=1}^N \sum_{i=1}^M \tau_{ij}^{(k)} \left(-\frac{D}{2} \ln(2\pi) - \frac{1}{2} \ln|\boldsymbol{\Sigma}_i| + \frac{D}{2} \ln u_{ij}^{(k)} - \frac{u_{ij}^{(k)}\|x_j - y_i\|^2}{2\boldsymbol{\Sigma}_i} \right)
 \end{aligned}$$

with the gamma in Q_2 that denotes the Digamma function, and the $u_{ij}^{(k)}$ in Q_3 that represents the conditional expectation about additional missing data in the complete-data; it can be written as:

$$u_{ij}^{(k)} = \frac{v_i^{(k)} + D}{v_i^{(k)} + D(x_j, y_i^{(k)}; \boldsymbol{\Sigma}_i^{(k)})}, \tag{6}$$

In addition, we use the displacement function ρ to denote the non-rigid transformation in the point set: $X = T(Y, \rho) = Y + \rho(Y)$, and add the regularization term $\varphi(\rho)$ to the Q_3 , which can enforce the smoothness of the displacement function in the alignment of two point sets. So this Q_3 term can be rewritten as:

$$\tilde{Q}_3 = Q_3(\rho, \boldsymbol{\Sigma}_i^{(k+1)}|\boldsymbol{\psi}^{(k)}) + \frac{\lambda}{2} \varphi(\rho), \tag{7}$$

where λ is the regularization parameter and controlling the trade-off of two terms in Equation (7). By reproducing Kernel Hilbert Space and Fourier transformation, the displacement function ρ can be denoted as:

$$\rho(y_j) = \sum_{i=1}^M h_i G(y_i, y_j), \quad G(y_i, y_j) = e^{-\frac{1}{2} \left\| \frac{y_i - y_j}{\beta} \right\|^2}, \tag{8}$$

where G is a Gaussian kernel matrix, β determines the width of the smoothing Gaussian filter, and the Equation (8) can be conveniently denoted in matrix form as: $\rho(Y) = GH$, where $H_{M \times D} = (h_1, \dots, h_M)$ is a matrix of coefficients. Thus, the Equation (7) can be expanded in the following matrix form:

$$\tilde{Q}_3 = \sum_{j=1}^N \sum_{i=1}^M \tau_{ij}^{(k)} \left(-\frac{D}{2} \ln(2\pi) - \frac{1}{2} \ln|\boldsymbol{\Sigma}_i| + \frac{D}{2} \ln u_{ij}^{(k)} - \frac{u_{ij}^{(k)}\|x_j - y_i\|^2}{2\boldsymbol{\Sigma}_i} \right) + \frac{\lambda}{2} \text{tr}\left(\left(H^T\right)^{(k)} GH^{(k)}\right), \tag{9}$$

M-step: On the M-step at the $(k + 1)^{\text{th}}$ iteration of the EM algorithm, considering that Q_1 , Q_2 and Q_3 in Equation (5) can be computed independently of each other, thus the maximization of the objective function Q_1 , Q_2 and Q_3 with respect to the parameters w , v , $\boldsymbol{\Sigma}$, ρ can be operated separately. The mixing proportion for SMM is updated by our consideration of the first term Q_1 . Given the feature descriptors in Section 2.1, it can obtain the coarse correspondences Ω between the point sets X and Y . For an observation data x_j , we define ι ($0 \leq \iota \leq 1$) as a confidence by IDSC feature matching, and then the mixing proportion w_{ij} by incorporating the local structures among neighboring points can be updated by the following rule:

$$w_{ij}^{(k+1)} = \begin{cases} \sum_{j=1}^N \frac{\tau_{ij}^{(k)}}{N}, & \exists(x_j \rightarrow y_i) \in \Omega \\ \iota, & \exists(x_j \rightarrow y_i) \in \Omega \end{cases}, \tag{10}$$

If observation data x_j does not have a corresponding point y_i in the label Ω , the mixing proportion $w_{ij}^{(k+1)}$ is given by the average of the posterior probabilities $\tau_{ij}^{(k)}$ of the SMM component membership. If the observation datum x_j corresponds to the centroid point y_i in the label Ω , the $w_{ij}^{(k+1)}$ is given by a constant confidence ι .

Note that in Equation (10) the mixing proportion w_{ij} does not only depend on prior assignment by local structures, but also depends upon the posterior probability of observation data x_j belonging to the i^{th} component of the SMM by global structures. In addition, we obtain degree of freedom ν by taking the corresponding derivative of Q_2 to zero, and the updated value $\nu^{(k+1)}$ is a solution of the following equation:

$$1 - \gamma \left(\frac{\nu_i^{(k+1)}}{2} \right) + \ln \left(\frac{\nu_i^{(k+1)}}{2} \right) + \frac{\sum_{j=1}^N \tau_{ij}^{(k)} (\ln u_{ij}^{(k)} - u_{ij}^{(k)})}{\sum_{j=1}^N \tau_{ij}^{(k)}} + \gamma \left(\frac{\nu_i^{(k)} + D}{2} \right) - \ln \left(\frac{\nu_i^{(k)} + D}{2} \right) = 0, \quad (11)$$

Then, to update the estimates of covariance matrices Σ and displacement function ρ in Equation (7), we also need to take the derivative of Q_3 , and the update values can be denoted as:

$$\begin{aligned} H^{(k+1)} &= \left(\text{diag}(P^{(k)} \mathbf{1}) G + \lambda \Sigma_i^{(k)} I \right)^{-1} (P^{(k)} X - \text{diag}(P^{(k)} \mathbf{1}) Y), P_{ij}^{(k)} = \tau_{ij}^{(k)} u_{ij}^{(k)} \\ \Sigma_i^{(k+1)} &= \frac{\sum_{j=1}^N \tau_{ij}^{(k)} u_{ij}^{(k)} \|x_j - T(y_i, \rho)\|^2}{\sum_{j=1}^N \tau_{ij}^{(k+1)}}, \quad (i = 1, \dots, N) \end{aligned} \quad (12)$$

where $\mathbf{1}$ is a column vector of all ones, and I is an identity matrix, $\text{diag}(\bullet)$ denotes a diagonal matrix and $G(i, \bullet)$ denotes the column vector in the kernel matrix G . Furthermore, for our centroid point y_i in the point set Y , the non-rigid transformation $T(Y, \rho) = Y + \rho(Y)$ can be expressed by Equation (8) as follows:

$$T(y_i, \rho) = y_i + G(i, \cdot) H_i^{(k+1)}, \quad (13)$$

At the end, the iterative process alternates between E-steps and M-steps until satisfying the following convergence condition:

$$\left\| \frac{L_C(\psi^{(k+1)}) - L_C(\psi^{(k)})}{L_C(\psi^{(k+1)})} \right\| \leq \varepsilon, \quad (14)$$

where ε is a convergence threshold. After a number of repeatability testing, we use the following setting for the parameters $\eta = 0.1$, $\iota = 0.8$, $\lambda = 3$, $\beta = 2$, $\varepsilon = 10^{-5}$ in our experiment. Furthermore, to register the visible and infrared images accordingly, the image transformation and interpolation is performed on the visible image. Since our multi-spectral face registration method is implemented by using global and local structures, then the registration method is to be named as Face Registration using the Global and Local Structure (FR-GLS) in the rest of this paper.

3. The Proposed Fusion Strategies for Visible and Infrared Face Images

After obtaining the accurate alignment of the visible and infrared images, the second problem is then to solve the multi-spectral images fusion. In order to make sure that the fused image can preserve both the apparent details of the visible image and thermal radiation information of the infrared image, we propose the fusion strategies based on Guided Filtering and Gradient preserving (GF-GP). Firstly, we use two-scale image decomposition and reconstruction with a guided filtering algorithm to get an initial fusion image, and then the output fusion image is optimized by combining the intensity distribution of the infrared image and the intensity variation of the visible image. According to [33], since each source image can be separated into a base layer containing the large-scale details, and the detail layer containing the small-scale details, the two-scale image fusion F_G is given by the guided filtering algorithm as follows:

$$F_G = \bar{B} + \bar{D}, \quad (15)$$

where the fused base layer \bar{B} and the fused detail layer \bar{D} can be denoted as a weighted average cost of the base B_n and detail D_n layers of different source images:

$$\bar{B} = \sum_{n=1}^N W_n^B B_n, \bar{D} = \sum_{n=1}^N W_n^D D_n, \quad (16)$$

where W_n^B and W_n^D represent the refined weight maps of the base B_n and detail layers D_n via guided image filtering respectively. In order to optimize the initial fusion image F_G , here the visible, infrared and fused images are defined by V , I and F , respectively. Given that the thermal radiation information in the infrared image is typically characterized by the pixel intensity values, the pixel intensity values are quite different between the target and background. Thus, we constrain the fused image F to have the similar pixel intensity distribution with the infrared source image I , and the optimization problem can be formulated as:

$$\delta_1 = \|F - I\|, \quad (17)$$

where $\|\bullet\|$ denotes the l_1 norm. Besides, given that the pixel intensity distribution in the same physical location might be discrepant for infrared and visible face images, they are manifestations of two different phenomena. Note that the detailed information of object edge and texture is mainly characterized by the pixel gradient values in the visible image. Hence, we constrain the fused image F to have the similar pixel gradient values with the visible source image V , the optimization problem can also be formulated as:

$$\delta_2 = \|\nabla F - \nabla V\|, \quad (18)$$

where ∇ denotes the gradient operator. By combining the Equations (17) and (18), the optimization problem of the fusion image F can be formulated as minimizing the following objective function:

$$\delta_F = \delta_1 + \alpha \delta_2 = \|F - I\| + \alpha \|\nabla F - \nabla V\|, \quad (19)$$

where α is a weighting factor, the optimal F could be found by using gradient descent strategy. Furthermore, it is important to note that if the α is small, the fusion image preserves the more thermal radiation information of the infrared image, otherwise, the fusion image preserves more edge and texture information of the visible image. We set the weighting coefficient $\alpha = 5$ in our experiment, because it can achieve good subjective visual quality in most cases.

4. Experimental Results

To demonstrate the robustness and efficiency of our proposed registration method in solving the multi-spectral face image registration problem, subjective and objective evaluation experiments are conducted with the UTK-IRIS standard multi-spectral face database [34] and self-constructed multispectral face data sets. The experiments are performed on a desktop with 3.3 GHz Intel Core™ i5-4590 CPU, 8 GB memory and MATLAB Code.

4.1. Registration on Real Face Images with the UTK-IRIS Database

In this section, the performance of face image registration with our FR-GLS method is evaluated by using the public UTK-IRIS Database. The images in the database have a spatial resolution of 320×240 pixels and the database contains individuals with various poses, facial expressions and illumination variations. We first give an intuitive impression of registration and fusion results, and then we provide a quantitative comparison with three typical methods such as Coherent Point Drift (CPD) [19], Regularized Gaussian Fields (RGF) [20] and SMM [25], based on the specified landmarks of visible and thermal infrared image pairs. The reasons for choosing these three comparison methods rest on the following two considerations. On the one hand, CPD and SMM use both global structures based on finite mixture models to parameterize the transformations. On the other hand, besides the global

relationships in the feature point sets, RGF uses local features descriptor such as SC to recover the accurate transformation.

However, our FR-GLS has two major advantages compared to RGF: (i) Unlike the local features descriptor used in RGF, we use a more robust local features descriptor such as IDSC to initialize the correspondences between the two shape features; (ii) In order to preserve the global and local structures of feature point sets during the registration process, a heavy tail distribution in our FR-GLS is used to replace the traditional Gaussian distribution in the RGF method. Furthermore, we use the IDSC descriptor to assign the mixing proportion w_{ij} of the SMM.

4.1.1. Qualitative Evaluation

In order to obtain a qualitative evaluation of the registration results issued by the software-based registration process in Figure 1, several typical frame pairs of four individuals involving different poses and illumination changes are given in Figure 4a–d. Specifically, the original visible images and thermal infrared images are shown in the first and second row, respectively. Then the middle row shows the edge maps extracted by the canny edge detector. We use the red point set to denote the edge map of the thermal infrared images, and the blue point set corresponds to the edge map of the visible images. Subsequently, the visible image is aligned to the thermal infrared image by our FR-GLS method, and the registration performance is demonstrated by the checkerboard in the fourth row, where the aligned visible image is superimposed onto the corresponding thermal infrared image. We can see that the seams between two grids in the checkerboard are natural, and this demonstrates the feasibility and effectiveness of the proposed registration method qualitatively. Finally, the last row represents the fusion results of the visible and thermal IR images by our GF-GP method. We can see that the fused images in the last row contain both the apparent details of the visible image and the thermal radiation information of the infrared image, which can be clearly seen from noses, cheeks, ears, mouths and glasses. Therefore, it will be beneficial for the performance of follow-up face detection and recognition under the case of illumination changes or disguise.

4.1.2. Quantitative Evaluation

In order to give a quantitative comparison of our FR-GLS and other typical methods, such as CPD [19], RGF [20] and SMM [25], we further utilize several visible and infrared image pairs of the four individuals in Figure 4, and manually selected several landmarks and their correspondences (about 20 pairs) in each visible and infrared image pairs as ground truth. For example, the landmarks involve the salient features like the cheeks, eyes, nose, mouth, eyebrows, ears and glasses, which are highly distinguished. Referring to the metric used in [20,21,35], the recall rate is proposed to be a qualitative indicator for evaluating the registration results on all landmarks set pairs of an individual. Here the recall rate, or true positive rate, is defined as the proportion of the true positive correspondences between the landmark pairs to the ground truth correspondences, and a true positive correspondence is counted when the pair falls within a given accuracy threshold of Euclidean distance. Here the Euclidean distance can be defined as 2-norm between a landmark in the aligned feature point set of visible images and the corresponding landmark in the feature point set of infrared images. The recall curves plot the recall rates under different threshold values.

Figure 5 shows the recall rates against different threshold values using the registration methods of CPD, SMM, RGF and FR-GLS, considering four individuals. Each individual contains about 100 hundred-image pairs corresponding to every spectrum. By the definition of recall, we can see that the FR-GLS and RGF methods for the global and local structures are significantly superior to the CPD and SMM methods for the global structures only based on finite mixture models. Moreover, the curves of our FR-GLS method are consistently above the curves of the RGF method for different accuracy thresholds and individuals, which could be attributable to the combination of the robust SMM and IDSC local feature descriptor.

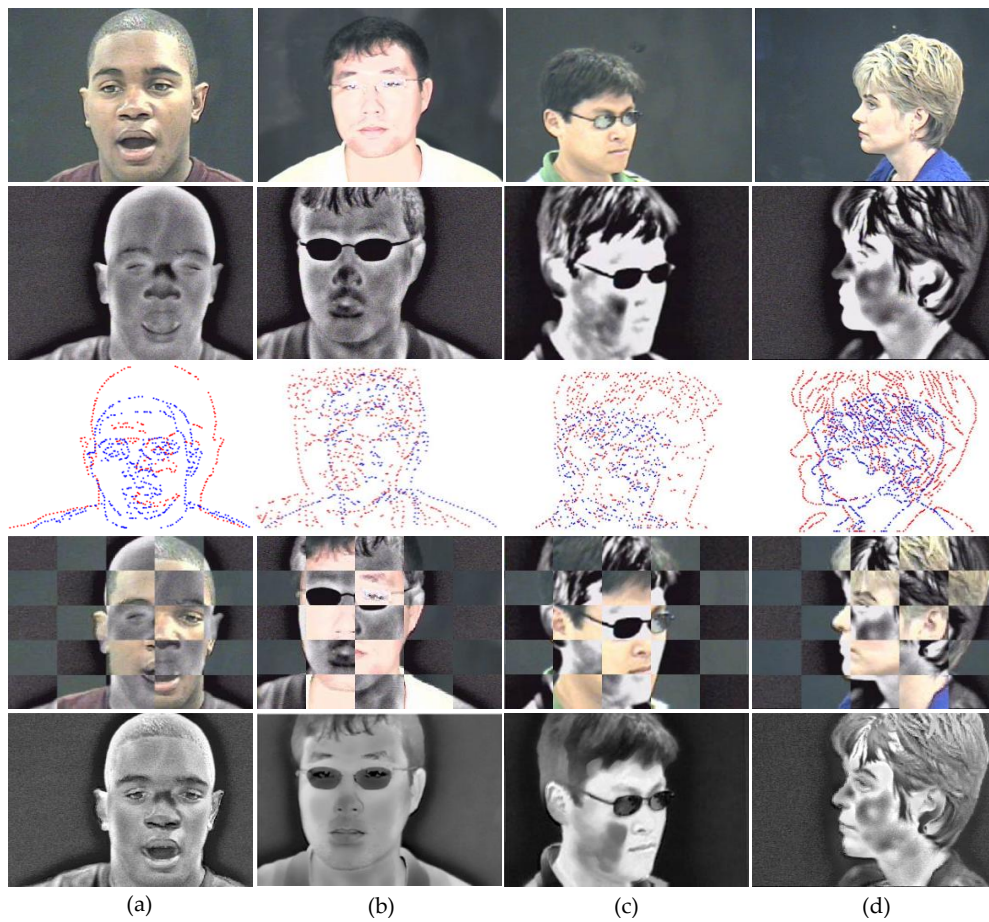


Figure 4. Registration and fusion results of our method on four typical unregistered visible/infrared image pairs in the database of UTK-IRIS. (a) Charles; (b) Heo; (c) Meng; (d) Sharon.

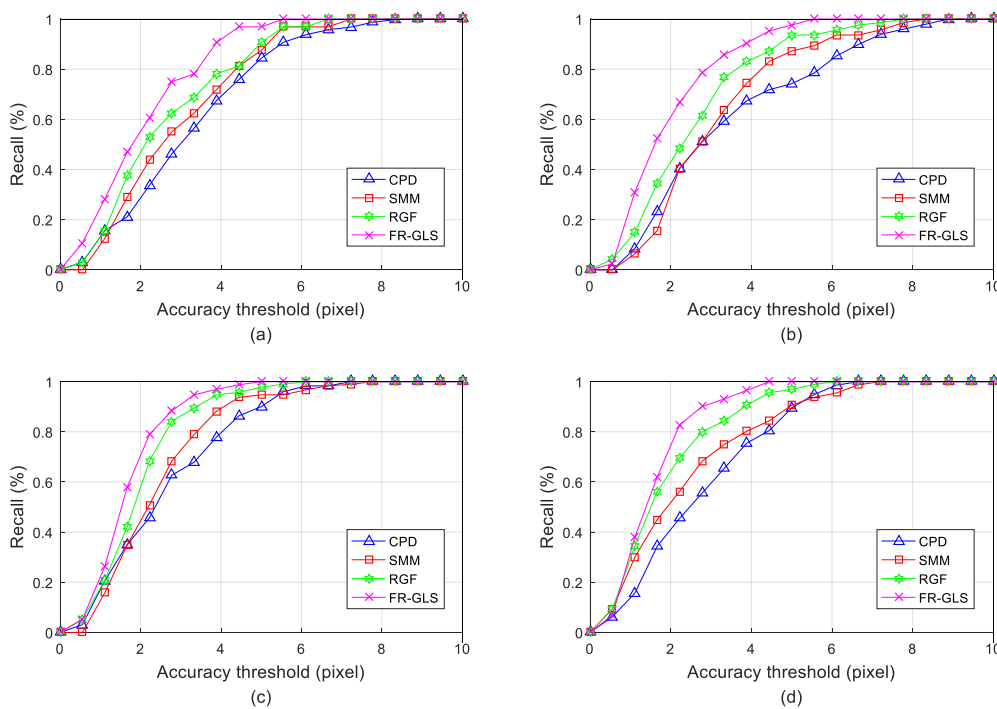


Figure 5. Quantitative comparisons of multispectral face image pairs with four individuals in Figure 4. (a) Charles; (b) Heo; (c) Meng; (d) Sharon.

The average registration errors of each individual for different registration methods are summarized in Table 1. We again see that the FR-GLS and RGF methods with the global and local structures significantly outperform the CPD and SMM methods just with finite mixture models.

Moreover, our FR-GLS method can achieve consistently better performance compared to the RGF method. The average registration errors of this RGF method are about 2.40, 2.24, 2.24 and 1.92 pixels on the four individuals, respectively. By contrast, the average registration errors of our FR-GLS method are reduced to about 2.25, 2.06, 1.99 and 1.78 pixels, respectively. In addition, The SMM method with SMM has a slightly better performance than the CPD method with GMM, where the average registration errors are 2.88, 2.40, 2.56 and 2.13 pixels, respectively. Whereas, the average registration errors of the CPD method are up to 3.04, 2.58, 2.73 and 2.35 pixels, respectively. It can quantitatively verify the heavy-tailed property of SMM.

Table 1. The average registration errors comparison of CPD, SMM, RGF and our FR-GLS method on four typical individuals in Figure 4.

Methods	The Average Registration Errors			
	Charles	Heo	Meng	Sharon
CPD [19]	3.04	2.58	2.73	2.35
SMM [25]	2.88	2.40	2.56	2.13
RGF [20]	2.40	2.24	2.24	1.92
FR-GLS	2.25	2.06	1.99	1.78

4.2. Registration on Real Face Images of Self-Built Multi-Spectral Database

In order to demonstrate the performance of our registration in the high-resolution multi-spectral imaging system, we built a four degree of freedom (DOF) pan and tilt head vision platform for multispectral database acquisition, which can be seen in Figure 6.

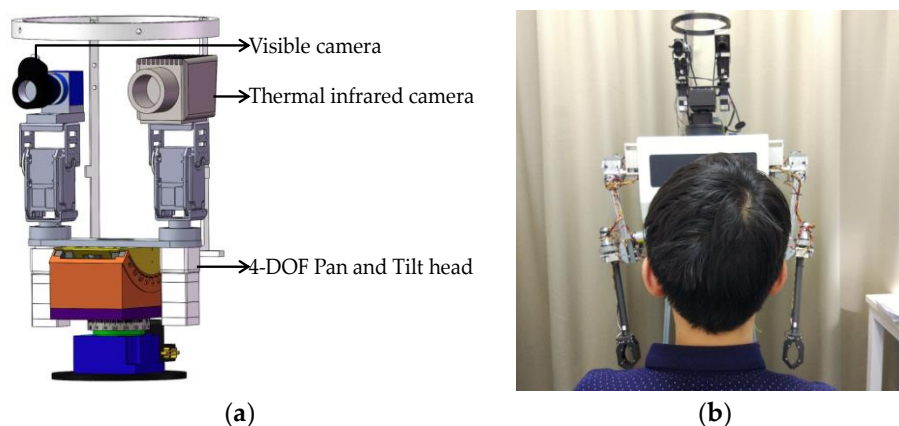


Figure 6. 4-DOF pan and tilt head vision platform for multi-spectral database acquisition: (a) Design mechanical model of vision platform; (b) Image acquisition scenario.

This 4-DOF pan and tilt head vision platform can realize up-down pitch movement and left-right yaw movement, respectively. Moreover, a Belgium Raven-640-Analog thermal infrared camera with 10 mm short lens is used to acquire a $74^\circ \times 59^\circ$ field of view (FOV) infrared image (Figure 6a left), while a DAHENG MER-310-12UC color camera with an Optotune EL-10-30-Ci variable-focus liquid lens is used to acquire a $28^\circ \times 22^\circ$ FOV visible image (Figure 6a right). Meanwhile, a visible image is always taken simultaneously with the infrared image to form a multi-spectral image pair, and the visible and infrared images are both compressed into 640×480 pixels. The multi-spectral vision equipment and image acquisition scenario is shown in Figure 6b. The self-built multi-spectral database is composed of 20 different individuals involving different illumination conditions and pose variations.

In addition, for each individual, the sampled images are captured from 13 different angles (tilt head from -90° to $+90^\circ$) and two illumination conditions (dark and light).

4.2.1. Qualitative Comparison

Referring to the qualitative evaluation in Section 4.1, Figure 7 shows an intuitive impression of registration and fusion results on six typical frame pairs in the self-built multi-spectral database.

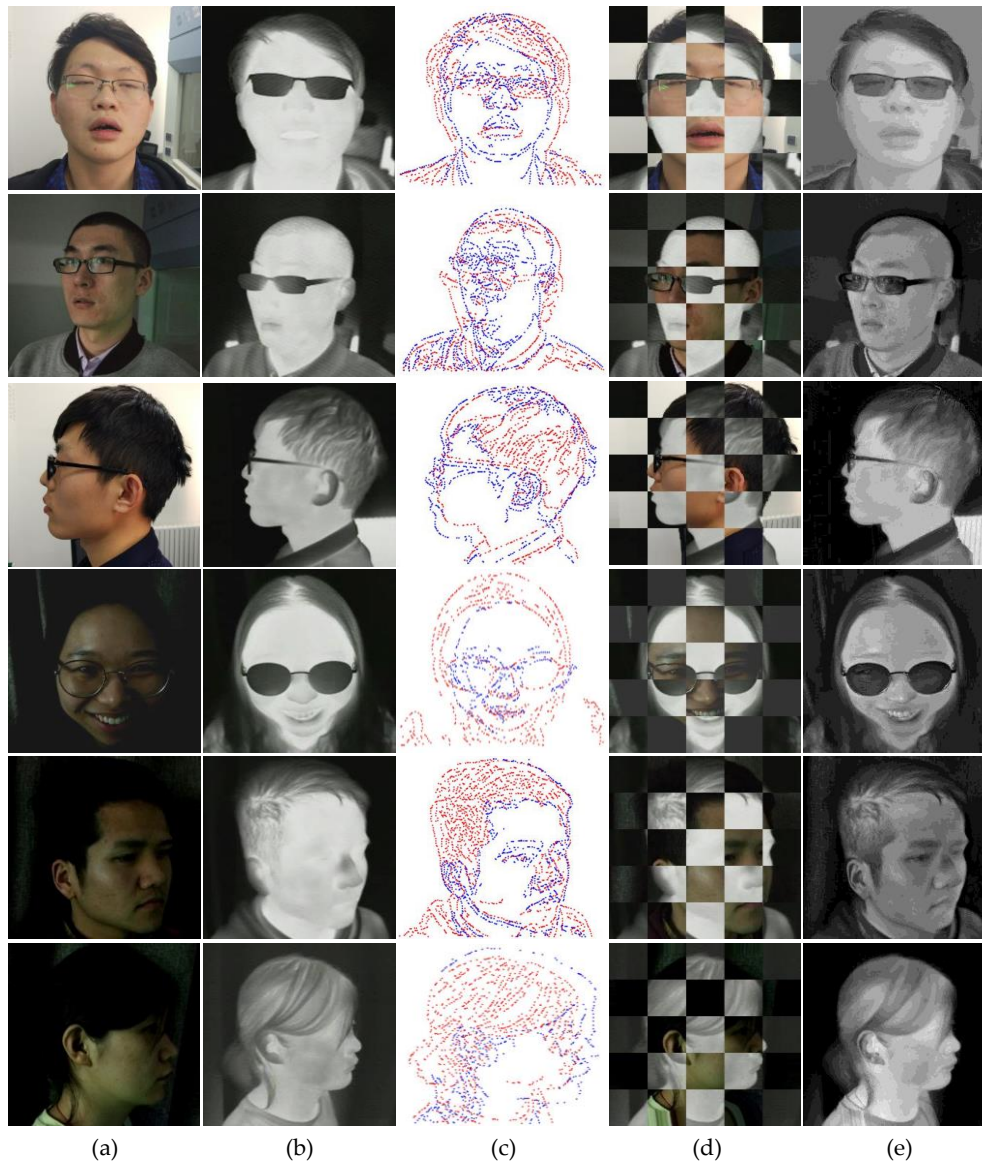


Figure 7. Registration and fusion results of our method considering six typical unregistered visible and infrared image pairs in the self-built multi-spectral face datasets. (a) Visible image; (b) Thermal infrared image; (c) Canny edge maps; (d) Superimposed checkerboard pattern of visible and infrared images; (e) Fusion results.

As shown in the Figure 7 above. Firstly, several original visible images and thermal infrared images of six individuals are given in Figure 7a,b, respectively, and those face images involve different poses and illumination changes. In addition, the image pairs in the bright light are shown in the first three rows, while the image pairs in the dark condition are shown in the last three rows with different poses. Then, Figure 7c shows the discrete point set of edge maps, which are extracted by the canny edge detector. For distinguishing each spectral image, we also use the blue point set to denote the edge

map of the visible image, and the red point set corresponds to the edge map of the thermal infrared image. We can see that contour point sets of the face images in the edge map pairs are significantly different for the thermal infrared and visible images. Those discrepancies will result in a lot of outlier and noise for the edge maps alignment, especially in the textured regions of hairs and clothes. Hence, it is hard to match the two edge map pairs accurately.

Furthermore, note that we just focus on the registration results of the face region, the redundant information, such as the edge points of clothes, which is not related to face detection, might be ignored during the matching. Subsequently, the visible image is aligned to the thermal infrared image by our FR-GLS method, and the registration performance is demonstrated by checkerboard images in Figure 7d, where the aligned visible image is superimposed to the corresponding thermal infrared image. In addition to the uninterested regions of hairs and clothes, we can see from Figure 7d that the seams between two grids in the checkerboard images are natural, and this can demonstrate the feasibility and effectiveness of the proposed registration method qualitatively. Finally, for further verification, Figure 7e shows the fusion results of the visible and thermal infrared images by our GF-GP method. Compared to the original thermal infrared images in the second column (Figure 7b), we can see that the fused images in the last column (Figure 7e) seem to be sharpened, which contain both the apparent details of the visible image and the thermal radiation information of the infrared image. Thus, it will be beneficial for follow-up face detection and recognition in case of illumination changes, e.g., the outdoor environment illumination changed from bright to dark.

4.2.2. Quantitative Comparison

To have a quantitative evaluation, we also manually select a set of landmarks in the multi-spectral images (about 40 pairs of landmarks for each spectral image), and then treat the recall on all real face images of an individual as the metric. Figure 8 shows the recall comparison results of the four methods on six individuals, where each individual contains about 52 image pairs.

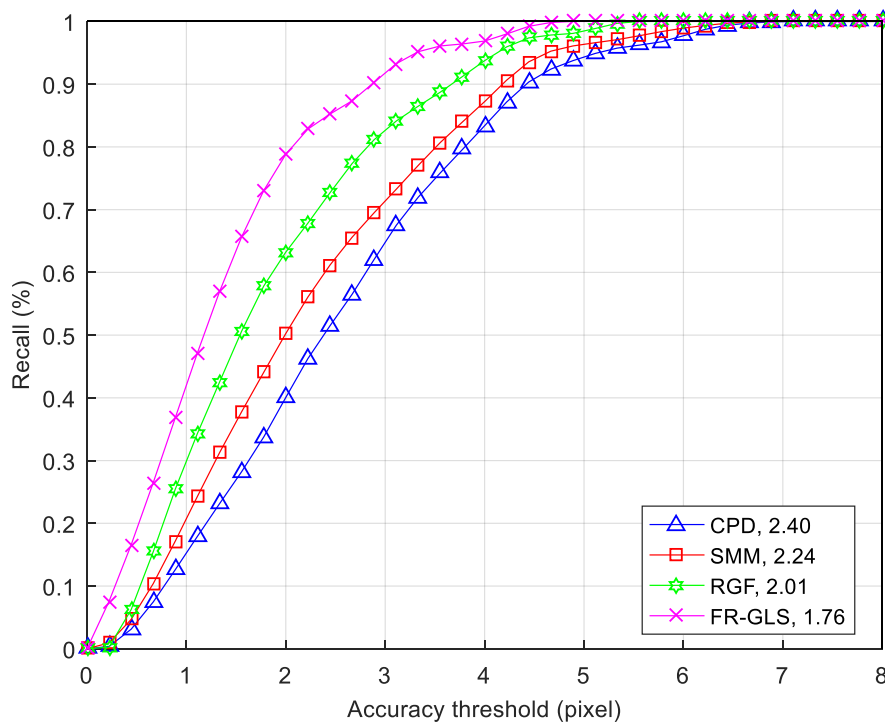


Figure 8. Quantitative comparisons of multi-spectral face image pairs with 6 individuals in Figure 7.

We can see from the Figure 8 above, FR-GLS (mauve fork) and RGF (green star) methods with the global and local structures are significantly superior to the CPD and SMM methods with global structures, and the curve of the Student's t distribution model (red square) is mainly above the curve of the Gaussian distribution model (blue triangle). Furthermore, the curve of our FR-GLS method, combined with SMM and IDSC, is consistently above those of the other three methods. In conclusion, for the six pairs of typical multi-spectral images in Figure 7, the average registration errors of CPD, SMM, RGF and our FR-GLS method are about 2.40, 2.24, 2.01 and 1.76 pixels, respectively.

5. Conclusions

In this paper, we have introduced a novel multi-sensor face images registration method. It uses the IDSC to describe the local feature of a face image, which is more stable than SC, and then the Student's t Mixture probabilistic model is used to estimate the transformation between visible and infrared images by combining global and local structures of feature point sets. In order to verify the correctness of the registration method intuitively, we propose a multi-spectral image fusion strategy based on guided filtering and gradient preserving. Experimental results on a standard real face database and self-built multi-spectral face database demonstrate that the proposed method is able to achieve much more registration accuracy compared to other state-of-the-art registration methods. Therefore, it will be beneficial to improve the reliability of a fusion-based face recognition system. In the future, given that the ultimate goal of registration and fusion is to enhance the recognition rate, a series of more comprehensive and scientific evaluations for multi-sensor face images registration will be conducted through some combination of the latest deep neural network face recognition method.

Author Contributions: Conceptualization, W.L. and W.Z.; Methodology, W.L.; Software, W.Z.; Validation, N.L. and W.Z.; Formal Analysis, W.L.; Investigation, W.L.; Resources, M.D.; Data Curation, W.L.; Writing-Original Draft Preparation, W.L.; Writing-Review & Editing, W.L. and W.Z.; Visualization, W.L.; Supervision, N.L. and M.D.; Project Administration, X.L.; Funding Acquisition, M.D and X.L.

Funding: This research was funded by National Natural Science Foundation of China, grant number No. 51475046, No. 51475047 and the National High-tech R&D Program of China. grant number No. 2015AA042308.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ma, J.; Ma, Y.; Li, C. Infrared and visible image fusion methods and applications: A survey. *Inf. Fusion* **2019**, *45*, 153–178. [[CrossRef](#)]
2. Ma, J.; Yu, W.; Liang, P. FusionGAN: A generative adversarial network for infrared and visible image fusion. *Inf. Fusion* **2019**, *48*, 11–26. [[CrossRef](#)]
3. Huang, Y.; Bi, D.; Wu, D. Infrared and visible image fusion based on different constraints in the non-subsampled shearlet transform domain. *Sensors* **2018**, *18*, 1169. [[CrossRef](#)]
4. Ma, J.; Chen, C.; Li, C. Infrared and visible image fusion via gradient transfer and total variation minimization. *Inf. Fusion* **2016**, *31*, 100–109. [[CrossRef](#)]
5. Paramanandham, N.; Rajendiran, K. Infrared and visible image fusion using discrete cosine transform and swarm intelligence for surveillance applications. *Infrared Phys. Technol.* **2018**, *88*, 13–22. [[CrossRef](#)]
6. Zhu, Z.; Qi, G.; Chai, Y. A geometric dictionary learning based approach for fluorescence spectroscopy image fusion. *Appl. Sci.* **2017**, *7*, 161. [[CrossRef](#)]
7. Singh, R.; Vatsa, M.; Noore, A. Integrated multilevel image fusion and match score fusion of visible and infrared face images for robust face recognition. *Pattern Recognit.* **2008**, *41*, 880–893. [[CrossRef](#)]
8. Vizgaitis, J.N.; Hastings, A.R. Dual band infrared picture-in-picture systems. *Opt. Eng.* **2013**, *52*, 061306. [[CrossRef](#)]
9. Kong, S.G.; Heo, J.; Boughorbel, F. Multiscale fusion of visible and thermal IR images for illumination-invariant face recognition. *Int. J. Comput. Vis.* **2007**, *71*, 215–233. [[CrossRef](#)]

10. Oliveira, F.P.M.; Tavares, J.M.R.S. Medical image registration: A review. *Comput. Method Biomech.* **2014**, *17*, 73–93. [[CrossRef](#)]
11. Avants, B.B.; Epstein, C.L.; Grossman, M. Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Med. Image Anal.* **2008**, *12*, 26–41. [[CrossRef](#)]
12. Pan, W.; Qin, K.; Chen, Y. An adaptable-multilayer fractional fourier transform approach for image registration. *IEEE Trans. Pattern Anal.* **2008**, *31*, 400–414. [[CrossRef](#)]
13. Zhuang, Y.; Gao, K.; Miu, X. Infrared and visual image registration based on mutual information with a combined particle swarm optimization—Powell search algorithm. *Optik Int. J. Light Electron Opt.* **2016**, *127*, 188–191. [[CrossRef](#)]
14. Sotiras, A.; Davatzikos, C.; Paragios, N. Deformable medical image registration: A survey. *IEEE Trans. Med. Imaging* **2013**, *32*, 1153–1190. [[CrossRef](#)]
15. Viergever, M.A.; Maintz, J.B.A.; Klein, S. A survey of medical image registration—under review. *Med. Image Anal.* **2016**, *33*, 140–144. [[CrossRef](#)]
16. Moghbel, M.; Mashohor, S.; Mahmud, R. Review of liver segmentation and computer assisted detection/diagnosis methods in computed tomography. *Artif. Intell. Rev.* **2018**, *50*, 497–537. [[CrossRef](#)]
17. Zhang, Y.; Zhao, D.; Sun, J. Adaptive convolutional neural network and its application in face recognition. *Neural Process. Lett.* **2016**, *43*, 389–399. [[CrossRef](#)]
18. Zhang, K.; Zhang, Z.; Li, Z. Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal. Proc. Lett.* **2016**, *23*, 1499–1503. [[CrossRef](#)]
19. Myronenko, A.; Song, X. Point set registration: Coherent point drift. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *32*, 2262–2275. [[CrossRef](#)]
20. Ma, J.; Zhao, J.; Ma, Y. Non-rigid visible and infrared face registration via regularized Gaussian fields criterion. *Pattern Recognit.* **2015**, *48*, 772–784. [[CrossRef](#)]
21. Tian, T.; Mei, X.; Yu, Y. Automatic visible and infrared face registration based on silhouette matching and robust transformation estimation. *Infrared Phys. Technol.* **2015**, *69*, 145–154. [[CrossRef](#)]
22. Peel, D.; McLachlan, G.J. Robust mixture modelling using the t distribution. *Stat. Comput.* **2000**, *10*, 339–348. [[CrossRef](#)]
23. Gerogiannis, D.; Nikou, C.; Likas, A. Robust Image Registration using Mixtures of t-distributions. In Proceedings of the 2007 IEEE 11th International Conference on Computer Vision, Rio de Janeiro, Brazil, 14–21 October 2007; pp. 1–8.
24. Gerogiannis, D.; Nikou, C.; Likas, A. The mixtures of Student’s t-distributions as a robust framework for rigid registration. *Image Vis. Comput.* **2009**, *27*, 1285–1294. [[CrossRef](#)]
25. Zhou, Z.; Zheng, J.; Dai, Y. Robust non-rigid point set registration using student’s-t mixture mode. *PLoS ONE* **2014**, *9*, e91381.
26. Zhou, Z.; Tu, J.; Geng, C. Accurate and robust non-rigid point set registration using student’s t mixture model with prior probability modeling. *Sci. Rep. UK* **2018**, *8*, 8742. [[CrossRef](#)]
27. Maiseli, B.; Gu, Y.; Gao, H. Recent developments and trends in point set registration methods. *J. Vis. Commun. Image Represent.* **2017**, *46*, 95–106. [[CrossRef](#)]
28. Belongie, S.; Malik, J.; Puzicha, J. Shape matching and object recognition using shape contexts. *IEEE Trans. Pattern Anal. Mach. Intell.* **2002**, *24*, 509–522. [[CrossRef](#)]
29. Ling, H.; Jacobs, D.W. Shape classification using the inner-distance. *IEEE Trans. Pattern Anal. Mach. Intell.* **2007**, *29*, 286–299. [[CrossRef](#)]
30. Li, H.; Wu, X.J. DenseFuse: A Fusion Approach to Infrared and Visible Images. *IEEE Trans. Image Process.* **2018**, *28*, 2614–2623. [[CrossRef](#)]
31. Chui, H.; Rangarajan, A. A new point matching algorithm for non-rigid registration. *Comput. Vis. Image Underst.* **2003**, *89*, 114–141. [[CrossRef](#)]
32. Riesen, K.; Bunke, H. Approximate graph edit distance computation by means of bipartite graph matching. *Image Vis. Comput.* **2009**, *27*, 950–959. [[CrossRef](#)]
33. Li, S.; Kang, X.; Hu, J. Image fusion with guided filtering. *IEEE Trans. Image Process.* **2013**, *22*, 2864–2875. [[PubMed](#)]

34. UTK-IRIS Database. Available online: <http://www.cse.ohio-state.edu/otcbvs-bench/> (accessed on 10 September 2015).
35. Dwith, C.; Ghassemi, P.; Pfefer, T. Free-form deformation approach for registration of visible and infrared facial images in fever screening. *Sensors* **2018**, *18*, 125. [[CrossRef](#)] [[PubMed](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).