Proceedings Paper:
Zhu, H., Tang, C., De Freitas, A. et al. (1 more author) (2020) A maximum likelihood approach to joint groupwise image registration and fusion by a Student-t mixture model. In: 2019 22th International Conference on Information Fusion (FUSION). 22nd International Conference on Information Fusion, 02-05 Jul 2019, Ottawa, Canada. IEEE . ISBN 9781728118406

A Maximum Likelihood Approach to Joint Groupwise Image Registration and Fusion by a Student-*t* Mixture Model

Hao Zhu
College of Automation
Chongqing University of Posts and Telecommunications
Chongqing, China
haozhu1982@gmail.com

Chunxia Tang
College of Automation
Chongqing University of Posts and Telecommunications
Chongqing, China
ttangcx@163.com

Allan De Freitas
Department of Electrical, Electronic and Computer Engineering
University of Pretoria
South Africa
Allandefreitas1@gmail.com

Lyudmila Mihaylova
Department of Automatic Control and Systems Engineering
University of Sheffield
Sheffield, UK
Mihaylova@sheffield.ac.uk

***Abstract*—In this paper, we propose a Student-*t* mixture model (SMM) to approximate the joint intensity scatter plot (JISP) of the groupwise images. The problem of joint groupwise image registration and fusion is considered as a maximum likelihood (ML) formulation. The parameters of registration and fusion are estimated simultaneously by an expectation maximization (EM) algorithm. To evaluate the performance of the proposed method, experiments on several types of multimodal images are performed. Comprehensive experiments demonstrate that the proposed approach has better performance than other methods.**

Keywords—image registration, image fusion, Student-t mixture model, expectation maximization

I. INTRODUCTION

Image fusion is a process of integrating redundant information present in multiple images into a single image [1]. It aims at the integration of disparate and complementary data to enhance the information apparent in the images as well as to increase the reliability of the interpretation. It has a wide variety of applications in computer vision, remote sensing, medical diagnosis, and intelligent transportation systems [2]. Furthermore, the issue to be solved in image fusion is the accurate image registration.

The image registration can be categorized into pairwise image registration and groupwise image registration [3]. Pairwise image registration methods align two images at a time, and groupwise image registration methods register multiple images at a time. The problem of pairwise registration is that the performance depends highly on the template [4], [5]. Therefore, many groupwise registration algorithms have been proposed [6], [7]. In [8], a joint intensity scatter plot (JISP) of multiple images is approximated by a finite Gaussian mixture model (GMM). The groupwise registration can be performed by minimizing the JISP. In [9], an infinite GMM is proposed to model the JISP, a variational Bayesian algorithm is then used to estimate the unknown parameters. Furthermore, in order to model the relationship between neighboring pixels in each image, a joint GMM and Markov random field, called a spatially constrained Gaussian mixture model, to model the JISP of the unregistered images [10].

The image registration and image fusion are two dependent processes. Some algorithms have been proposed to perform image registration and fusion simultaneously [11]. In [12], the problem of pairwise image registration and fusion is considered as a maximum likelihood (ML) problem. An expectation maximization (EM) algorithm is used to solve the ML problem. In [13], an iterative optimization approach, which jointly considers the registration and fusion processes, is proposed. However, the above literature only considers joint pairwise image registration and fusion. The pairwise image registration and fusion methods are problematic since results depend highly on which image is chosen as the template and there is no guarantee that redundancy in the solution is consistent [9]. To address this issue, a GMM is proposed to approximate the JISP of multiple images and the mapping from the fused image to the source images [14]. However, this algorithm is sensitive to outliers.

As the Student-*t* distribution has a heavy tail compared to the Gaussian distribution and is robust to outliers, a Student-*t* mixture model (SMM) is proposed to approximate the JISP of unregistered images and the mapping from the fused image to the source images to improve the robustness. The problem of groupwise image registration and fusion can be performed by minimizing the

This work is jointly supported by Key Research and Development Project of Chongqing Science and Technology Commission (cstc2017zdcy-zdyfX0004), by the Scientific and Technological Research Program of Chongqing Municipal Education Commission (Grant No. KJ1704078), by the Research Funds of Chongqing Science and Technology Commission (Grant No. cstc2017jcyjAX0293), by the National Natural Science Foundation of China (Grant No. 61773082), by the Foundation of Chongqing (Grant No. CSTC2017JCYJBX0018 and No. CX2017044), by the Key Project of Crossing and Emerging Area of CQUPT (Grant No. A2018-02), by the Research Fund of young-backbone university teacher in Chongqing province, by Chongqing Overseas Scholars Innovation Program, by Wenfeng Talents of Chongqing University of Posts and Telecommunications, by Innovation Team Project of Chongqing Education Committee (Grant No.CXTDX201601019), by the Research Funds of Chongqing University of Posts and Telecommunications (Grant No. A2018-174), by the National Key Research and Development Program (Grant No. 2016YFB0100906), by the Research and Innovation Project of Chongqing Postgraduate, by the Lilong Innovation and Entrepreneurship Fund of Chongqing University of Posts and Telecommunications (Grant No. 2019-1-01).

JISP. In order to solve this problem, an EM algorithm is used to estimate the unknown parameters in the proposed method.

This paper has organized the rest of this paper as follows: In section 2, the problem of joint groupwise images registration and fusion is formulated. In section 3, an EM algorithm is presented to estimate the parameters, and the experimental results are given in section 4. In the end, we conclude the paper with future work.

II. Problem Formulation

Assume that *D* images from different sensors will be simultaneously registered and fused into one image. Each pixel *x* corresponded to *D* intensity values of an image. A vector of intensities *I*(*x*) is used to represent each pixel and each fused image pixel is *F*(*x*). *I*(*x*) and *F*(*x*) are combined into a *D* + 1 dimensional vector *L*(*x*) to model a JISP of the unregistered image.

$$L(x) = \left[I(x); F(x) \right] \tag{1}$$

Groupwise registration is to give these images a set of motion parameters which refer to the transforms applied to them. If each image of the group has *M* motion parameters, the total number of motion parameters is *MD*+*M* (the fused image also has *M* motion parameters) [14]. We use θ to represent the motion parameter set, the JISP with motion parameters becomes

$$L^{\theta}(x) = \left[I^{\theta}(x); F^{\theta}(x) \right] \tag{2}$$

An SMM is used to model the probability distribution of JISP components of a fused image. The mixture consists of *K* Student-*t* components. For each pixel location *x*, the likelihood of the JISP could be written as

$$p\left(L^{\theta}(x) | u_k, \Sigma_k, v_k \right) = \sum_{k=1}^{K} \pi_k S\left(L^{\theta}(x) | u_k, \Sigma_k, v_k \right) \tag{3}$$

where u_k and Σ_k denote the mean and covariance of *k*th Student-*t* component in SMM, respectively, and v_k is the degree of freedom, the mixing coefficients π_k satisfy $\pi_k \geq 0$ and a constraint $\sum_{k=1}^{K} \pi_k = 1$. *S*(.) denotes the Student-*t* distribution function as

$$\begin{aligned} S\left(L^{\theta}(x) \mid u_k, \Sigma_k, v_k \right) = & \frac{\Gamma\left(\frac{D+v_k}{2} \right)}{\Gamma\left(\frac{v_k}{2} \right) (v_k \pi)^{\frac{D}{2}}} \mid \Sigma_k \mid^{-\frac{1}{2}} \\ & \times \left[1 + \frac{1}{v_k} \left(L^{\theta}(x) - u_k \right)^T \left(\Sigma_k \right)^{-1} \left(L^{\theta}(x) - u_k \right) \right]^{-\frac{D+v_k}{2}} \end{aligned} \tag{4}$$

where $\Gamma(y) = \int_0^{\infty} x^{y-1} e^{-x} dx$ represents a Gamma function.

Based on the formation model of a sensory image [11], the mapping from $F^{\theta}(x)$ to $I^{\theta}(x)$ can be written as

$$I^{\theta}(x) = \beta_x F^{\theta}(x) + \alpha_x + w_x \tag{5}$$

where β_x is a vector of sensor gain which includes the effects of local polarity reversals and complementarity, α_x is a vector of sensor offsets, and w_x is a random distortion vector which is modeled by an SMM. The constraint as proposed in [15] is set so that the value of β_x can be taken from {-1,0,1}. The constraint acknowledges that a sensor may be able to "see" certain objects ($\beta_x = 1$), may to "see" other objects ($\beta_x = 0$), or may "see" certain objects with polarity reversed representation ($\beta_x = -1$). A vector $A^{\theta}(x)$ is used to represent the relationship between $F^{\theta}(x)$ and $I^{\theta}(x)$, and it is consistent with the noise distribution:

$$A^{\theta}(x) = I^{\theta}(x) - \beta_x F^{\theta}(x) - \alpha_x \tag{6}$$

$A^{\theta}(x)$ is assumed to follow an SMM with a zero mean. Therefore, the likelihood of $A^{\theta}(x)$ can be represented as

$$p\left(A^{\theta}(x) \middle| \Sigma_r, v_r \right) = \sum_{r=1}^{R} \pi_r S\left(A^{\theta}(x) \middle| 0, \Sigma_r, v_r \right) \tag{7}$$

where *R* is the number of Student-*t* components, Σ_r and π_r denote the covariance and component weight of the *r*th Student-*t* distribution, respectively, v_r is the degree of freedom.

The complete data set of the joint groupwise image registration and fusion problem can be defined as $C = \left\{ L^{\theta}(x), A^{\theta}(x) \right\}$. Correspondingly, the incomplete data set is $Z = \left\{ I^{\theta}(x) \right\}$, and the unknown parameters are denoted as $\rho = \left\{ F^{\theta}(x), \beta_x, \alpha_x, u_k, \Sigma_k, v_k, \pi_k, \Sigma_r, v_r, \pi_r, \theta \right\}$ with $x = 1, \ldots, X$, $k = 1, \ldots, K$ and $r = 1, \ldots, R$, where *X* is the total number of pixels. Since each sensor image is an independent observation on the true scene, the log- likelihood function of the complete data of joint groupwise image registration and fusion can be written as

$$\begin{aligned} \log \mathcal{L}(C|\rho) &= \log \prod_{x=1}^{X} p\left(L^{\theta}(x), A^{\theta}(x) | \rho \right) \\ &= \sum_{x=1}^{X} \left[\log p\left(L^{\theta}(x) | \rho \right) + \log p\left(A^{\theta}(x) | \rho \right) \right] \end{aligned} \tag{8}$$

Our goal is to find out the appropriate choice of ρ to maximize $\log \mathcal{L}(C|\rho)$. The hidden random variables z_{xk} and z_{xr} are introduced to indicate the membership of $L^{\theta}(x)$ and $A^{\theta}(x)$ respectively among their own clusters. The log-likelihood function can be written as (9).

III. Joint Image Registration and Fusion using EM

The EM algorithm is employed to estimate the model parameters and to produce the fused image. There are two steps in the EM algorithm:

$$\begin{aligned}\log \mathcal{L}(C|\rho) &= \log \prod_{x=1}^{X} p\left(L^{\theta}(x), A^{\theta}(x), z_{xk}, z_{xr}|\rho\right) \\ &= \sum_{x=1}^{X}\left[\log p\left(L^{\theta}(x), z_{xk}|\rho\right)+\log p\left(A^{\theta}(x), z_{xr}|\rho\right)\right] \\ &= \sum_{x=1}^{X}\left[\log p\left(L^{\theta}(x)|z_{xk}, \rho\right)p(z_{xk}|\rho)+\log p\left(A^{\theta}(x)|z_{xr}, \rho\right)p(z_{xr}|\rho)\right]\end{aligned} \tag{9}$$

E-step: $Q\left(\rho,\rho^{(t)}\right)=E\left[\mathcal{L}(C|\rho)\middle|Z,\rho^{(t)}\right]$

M-step: $\rho^{(t+1)}=\max_{\rho} Q\left(\rho,\rho^{(t)}\right)$

where t represents the tth iteration. By iterating these two steps, the parameters are determined with monotonically non-decreasing observed-data likelihood value.

A. E-step

In the E-step, the conditional expectation of (9) can be evaluated as (10) at the bottom of the page. Where C is a term that unrelated to ρ . The conditional probability γ_{xk} of z_{xk} can be computed as

$$\gamma_{xk}=p(z_{xk}|Z,\rho)=\frac{\pi_k S\left(L^{\theta}(x)|u_k,\Sigma_k,v_k\right)}{\sum_{k=1}^{K}\pi_k S\left(L^{\theta}(x)|u_k,\Sigma_k,v_k\right)} \tag{11}$$

the conditional expectation values γ_{xr} of z_{xr} is computed as

$$\gamma_{xr}=p(z_{xr}|Z,\rho)=\frac{\pi_r S\left(A^{\theta}(x)\middle|0,\Sigma_r,v_r\right)}{\sum_{r=1}^{R}\pi_r S\left(A^{\theta}(x)\middle|0,\Sigma_r,v_r\right)} \tag{12}$$

B. M-step

In the M-step, the parameters are determined by maximizing $Q\left(\rho,\rho^{(t)}\right)$. The fused image $F^{\theta}(x)$ is computed by

$$F^{\theta^{(t)}}(x)=\frac{\sum_{k=1}^{K}\gamma_{xk}\frac{u_k-L^{\theta}(x)}{(\Sigma_k)^{-1}}+\sum_{r=1}^{R}\gamma_{xr}\beta_x^T(\Sigma_r)^{-1}\left(I^{\theta}(x)-\alpha_x\right)}{\sum_{k=1}^{K}\frac{\gamma_{xk}}{(\Sigma_k)^{-1}}+\sum_{r=1}^{R}\gamma_{xr}\beta_x^2(\Sigma_r)^{-1}} \tag{13}$$

Similarly, we can update u_k , Σ_k , π_k , Σ_r , π_r as

$$\begin{aligned}&u_k^{(t)}=\frac{\sum_{k=1}^{K}\gamma_{xk}L^{\theta}(x)}{\sum_{x=1}^{X}\gamma_{xk}}, \pi_k^{(t)}=\frac{\sum_{k=1}^{K}\gamma_{xk}}{\sum_{k=1}^{K}\sum_{x=1}^{X}\gamma_{xk}} \\ &\Sigma_k^{(t)}=\frac{\sum_{k=1}^{K}\gamma_{xk}\left(L^{\theta}(x)-u_k^{(t)}\right)\left(L^{\theta}(x)-u_k^{(t)}\right)^T}{\sum_{x=1}^{X}\gamma_{xk}} \\ &\pi_r^{(t)}=\frac{\sum_{r=1}^{R}\gamma_{xr}}{\sum_{r=1}^{R}\sum_{x=1}^{X}\gamma_{xr}}, \Sigma_r^{(t)}=\frac{\sum_{r=1}^{R}\gamma_{xr}\left(A^{\theta}(x)\right)\left(A^{\theta}(x)\right)^T}{\sum_{x=1}^{X}\gamma_{xr}}\end{aligned} \tag{14}$$

The updating of v_k is the result of the equation

$$\begin{aligned}&-\Psi\left(\frac{v_k^{(t\text{-}1)}}{2}\right)+\log\left(\frac{v_k^{(t\text{-}1)}}{2}\right)+1+\frac{\sum_{x=1}^{X}\gamma_{xk}\left(\log u_k^{(t)}-u_k^{(t)}\right)}{\sum_{x=1}^{X}\gamma_{xk}} \\ &+\Psi\left(\frac{v_k^{(t)}+D}{2}\right)-\log\left(\frac{v_k^{(t)}+D}{2}\right)=0\end{aligned} \tag{15}$$

where $\Psi(x)=\frac{\partial(\ln\Gamma(x))}{\partial x}$ is the digamma function.

$$\begin{aligned}Q\left(\rho,\rho^{(t)}\right)&=E\left[\mathcal{L}(C|\rho)\middle|Z,\rho^{(t)}\right] \\ &=C+\sum_{x=1}^{X}\sum_{k=1}^{K}\gamma_{xk}\left\{\log\pi_k+\log\Gamma\left(\frac{D+v_k}{2}\right)-\log\Gamma\left(\frac{v_k}{2}\right)-\frac{1}{2}\log|\Sigma_k|-\frac{D}{2}\log(v_k)\right\} \\ &\quad-\sum_{x=1}^{X}\sum_{k=1}^{K}\gamma_{xk}\left\{\frac{D+v_k}{2}\log\left[1+\frac{1}{v_k}\left(L^{\theta}(x)-u_k\right)^T\left(\Sigma_k\right)^{-1}\left(L^{\theta}(x)-u_k\right)\right]\right\} \\ &\quad+\sum_{x=1}^{X}\sum_{r=1}^{R}\gamma_{xr}\left\{\log\pi_r+\log\Gamma\left(\frac{D+v_r}{2}\right)-\log\Gamma\left(\frac{v_r}{2}\right)-\frac{1}{2}\log|\Sigma_r|-\frac{D}{2}\log(v_r)\right\} \\ &\quad-\sum_{x=1}^{X}\sum_{r=1}^{R}\gamma_{xr}\left\{\frac{D+v_r}{2}\log\left[1+\frac{1}{v_r}\left(A^{\theta}(x)\right)^T\left(\Sigma_r\right)^{-1}\left(A^{\theta}(x)\right)\right]\right\}\end{aligned} \tag{10}$$

The estimate of the degree of freedom v_r is given by the solution of the equation.

$$-\Psi\left(\frac{v_r^{(t\text{-}1)}}{2}\right)+\log\left(\frac{v_r^{(t\text{-}1)}}{2}\right)+1+\Psi\left(\frac{v_r^{(t)}+D}{2}\right)-\log\left(\frac{v_r^{(t)}+D}{2}\right)=0 \tag{16}$$

However, the maximization of $Q\left(\rho,\rho^{(t)}\right)$ is difficult to perform analytically due to the nonlinear coupling between the parameters α_x and β_x, this problem can be solved by the SAGE algorithm [11]. The calculations will employ the pixels in an $W=h\times h$ window around x. We assume that the parameters α_x and β_x are the same for each pixel in the window. $Q\left(\rho,\rho^{(t)}\right)$ can be maximized by setting the differential with respect to α_x and β_x to zero, while other parameters are unchanged. We have

$$\alpha_x^{(t)}=\frac{\sum_{x=1}^{W}\left(I^{\theta}(x)-\beta F^{\theta^{(t)}}(x)\right)\sum_{r=1}^{R}\gamma_{xr}\left(\Sigma_r^{(t)}\right)^{-1}}{\sum_{x=1}^{W}\sum_{r=1}^{R}\gamma_{xr}\left(\Sigma_r^{(t)}\right)^{-1}} \tag{17}$$

$$\beta_x^{(t)}=\frac{\sum_{x=1}^{W}F^{\theta^{(t)}}(x)\sum_{r=1}^{R}\gamma_{xr}\left(\Sigma_r^{(t)}\right)^{-1}\left(I^{\theta}(x)-\alpha_x^{(t)}\right)}{\sum_{x=1}^{W}\left(F^{\theta^{(t)}}(x)\right)^{2}\sum_{r=1}^{R}\gamma_{xr}\left(\Sigma_r^{(t)}\right)^{-1}} \tag{18}$$

The motion parameters are estimated by computing a motion increment which moves all the pixels so that $Q\left(\rho,\rho^{(t)}\right)$ is increased. To optimize $Q\left(\rho,\rho^{(t)}\right)$ with respect to the parameters θ as follow

$$\frac{\partial Q\left(\rho,\rho^{(t)}\right)}{\partial\theta}=0 \tag{19}$$

In order to find motion parameters θ which satisfy the above equation, we introduce a small motion increment $\tilde{\theta}$ and denote $\theta+\tilde{\theta}$ as the estimated parameters. Improved motion parameters that set (19) to a zero vector can be found by using a linear approximation of the spatial transformation.

$$L^{\theta+\tilde{\theta}}(x)=L^{\theta}(x)+\frac{\partial L^{\theta^T}(x)}{\partial\theta}\tilde{\theta},\ A^{\theta+\tilde{\theta}}(x)=A^{\theta}(x)+\frac{\partial A^{\theta^T}(x)}{\partial\theta}\tilde{\theta} \tag{20}$$

where $\tilde{\theta}$ is a small motion increment in the motion parameters. Putting (19) and (20) together yield a linear system as (21). This linear system can be easily solved to obtain the motion increment $\tilde{\theta}$, which is used to update the current motion parameters θ.

$$\begin{aligned}&\left(\sum_{x=1}^{X}\left(\sum_{k=1}^{K}\gamma_{xk}\frac{\partial L^{\theta}(x)}{\partial\theta}\left(\Sigma_k^{(t)}\right)^{-1}\frac{\partial L^{\theta^T}(x)}{\partial\theta}+\sum_{r=1}^{R}\gamma_{xr}\frac{\partial A^{\theta}(x)}{\partial\theta}\left(\Sigma_r^{(t)}\right)^{-1}\frac{\partial A^{\theta^T}(x)}{\partial\theta}\right)\right)\tilde{\theta}\\&=\sum_{x=1}^{X}\left(\sum_{k=1}^{K}\gamma_{xk}\frac{\partial L^{\theta}(x)}{\partial\theta}\left(\Sigma_k^{(t)}\right)^{-1}\left(L^{\theta}(x)-u_k^{(t)}\right)+\sum_{r=1}^{R}\gamma_{xr}\frac{\partial A^{\theta}(x)}{\partial\theta}\left(\Sigma_r^{(t)}\right)^{-1}A^{\theta}(x)\right)\end{aligned} \tag{21}$$

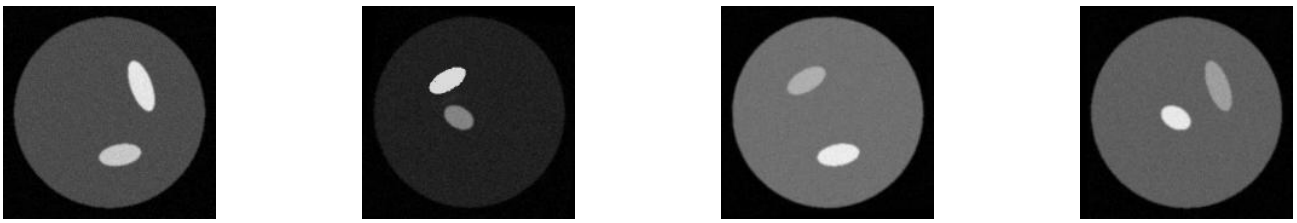
IV. Experimental Evaluation

In this section, the proposed joint image registration and fusion are compared with many representative methods. For registration performance, the mutual information-based method (MI) [14], the pattern intensity-based method (PI) [15], and the groupwise registration method (ER) [16] are compared. In order to demonstrate the performance of fusion, the Laplacian method (LP) [8] and the statistical signal process method (SS) [17] are compared. The joint groupwise image registration and fusion using the GMM method [14] is also compared.

The performance of each registration method is gauged by comparing the estimated transformations to the gold standard transformations. This difference is quantified using the average pixel displacement (APD), defined as the distance of each pixel from its true, registered position, averaged over all pixels used in the registration. Hence, a perfect registration means that its APD is 0, and a large APD indicates poor registration. If the APD is greater than 3 pixels, then the registration is considered a failure [18].

For the fusion evaluation, three fusion quality metrics are employed. The first is the $Q^{ab/f}$, which measures the degree of retention of edge information in a fused image [19]. The second is mutual information measure (MIM), which measures the extent to which the content of the source image is preserved in the fused image [20]. The third is the average gradient (AG), which is used to measure the spatial resolution of the fused image [20]. A high quality of fused image indicates that the value of these metrics is closer to 1.

Some publicly available multimodal image data sets are used to test the performances of the registration and fusion methods. Each image data set is registered by a supplied gold standard registration. Then, we generated trial image data sets by applying known displacements to the initially registered images. Details about image data sets and the experiments are given in the following subsection.



(a) (b) (c) (d)

Fig. 1. Multisensor phantom dataset

TABLE I. THE RESULT OF IMAGE REGISTRATION ON FOUR DATASETS. VALUES IN BOLD INDICATE BEST RESULTS FOR EACH CRITERION AND DATA SET.

Method	Datasets			
	Multisensor Phantom	*Medical Images*	*Face Images*	*Satellite Images*
MI	3.634	5.262	22.332	15.970
PI	4.671	13.794	17.948	22.952
ER	1.189	1.356	1.402	0.267
GMM	1.086	0.987	1.098	0.253
The proposed method	**1.068**	**0.922**	**1.022**	**0.251**

A. Multisensor Phantom

The multisensor phantom dataset is shown in Fig. 1. It is shown that the true object consists of a large circle encapsulating four ellipses. However, only two of the four ellipses are visible in each image. Fifty trail groups are generated by using randomly generated translation and rotation which are uniformly chosen from the range [-10, 10] pixels or angles. The proposed method is initialized with four Student-*t* components. The registration results are given in TABLE I. The proposed method has the lowest APD, and demonstrates a better result compared with others.

To evaluate the fusion performance of the proposed method under different levels of noise. Different levels of non-Gaussian noise and salt-and-pepper noise are added to each image. The variance of non-Gaussian noise and the density of salt-and-pepper noise increase from 0 to 0.04. Fig. 2. plots the average $Q^{ab/f}$ values of different methods. It is shown that the ER-SS method is slightly better than the proposed method when the noise is 0. When the noise is increased, the proposed method outperforms other methods. The MIM measure results are given in Fig. 3. The proposed method has the best performance compared with the other three methods. Fig. 4. plots the AG values of different methods. It is observed that the proposed method has the best performance among all.

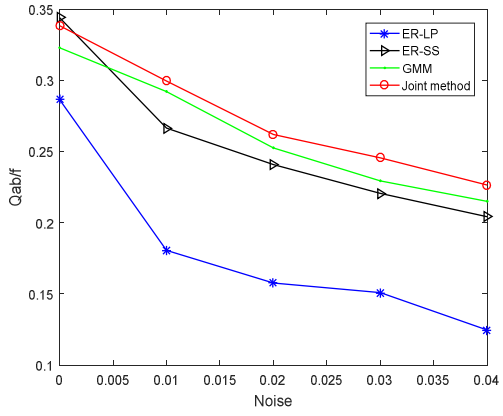


Fig. 2. The $Q^{ab/f}$ values of the ER-LP method, the ER-SS method, the GMM method and the proposed method with different levels of noise.

"ER-LP" indicates that the source images are registered by using the ER based method and fused by using LP method.

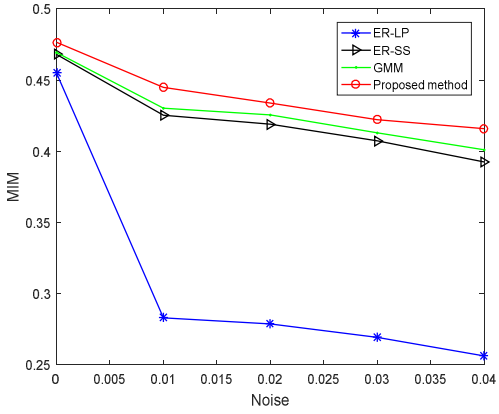


Fig. 3. The MIM values of the ER-LP method, the ER-SS method, the GMM method and the proposed method with different levels of noise

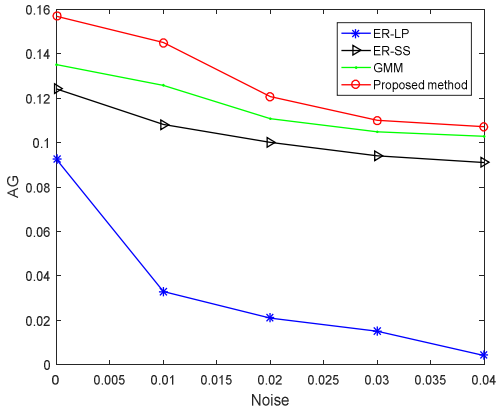


Fig. 4. The AG values of the ER-LP method, the ER-SS method, the GMM method and the proposed method with different levels of noise

B. Medical Images

The medical images come from the Retrospective image Registration Evaluation project (one slice is shown in Fig. 5). The group contains a CT image, a PET image and three MR images (T1-weighted, T2-weighted and PD weighted). Fifty sets of source images are generated using random translation and rotation from the range of [-5, 5] pixels or angles. The number of Student-*t* components is set as four. According to TABLE I., the proposed method has the best

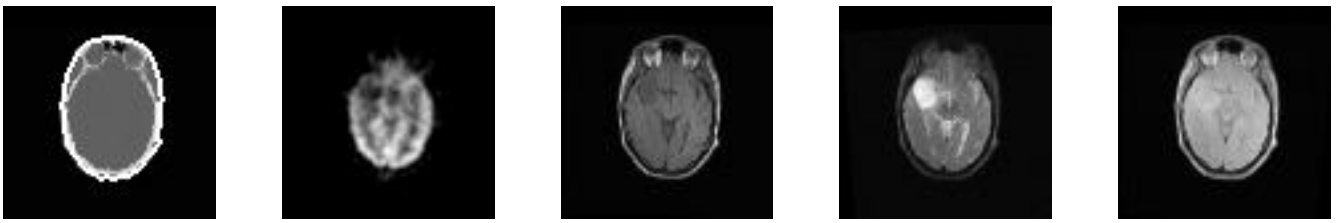
registration performance. TABLE II. shows fusion evaluation results, while the proposed method outperforms other methods on three evaluation measures. It is also observed that the proposed method is better than the GMM method.

C. Face Images

The face images with five different light positions are shown in Fig. 6. The light positions ranging from far left to far right. The dataset is obtained from the Extend Yale Face Database B. Fifty sets of source images are generated by applying rigid-body displacements. Parameters are chosen uniformly from the range [-10, 10] pixels or angles. The proposed method is initialized with five Student-*t* components. The registration results are shown in TABLE I. It is observed that the proposed method gets a lower APD than other methods. From TABLE II., it is shown that the fusion performance of the proposed method is better than those of the ER-LP method, ER-SS method, and the GMM method.

D. Satellite Images

The last dataset contains six images from different bands of Landsat 7 satellites. The Landsat 7 satellite images, each with 761 × 748 pixels in size, as shown in Fig. 7. A total of fifty trial images groups are generated by applying different sets of 6 degrees of freedom affine displacements. Parameter values are uniformly sampled from the following ranges: translation and rotation [-5, 5], scale [0.95,1.05], and shear [-0.2,0.2]. The number of Student-*t* distribution is chosen as nine. TABLE I. shows the registration results. The proposed method outperforms other methods. According to TABLE II. It is found that the proposed method has better fusion performance than the ER-LP and ER-SS method. The proposed method has almost the same fusion performance as that of the GMM method.

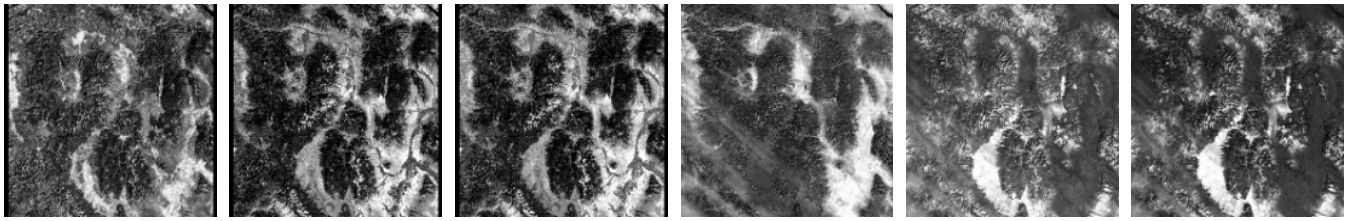


(a) (b) (c) (d) (e)

Fig. 5. Medical image dataset

(a) (b) (c) (d) (e)

Fig. 6. Face image dataset



(a) (b) (c) (d) (e) (f)

Fig. 7. Satellite image dataset

TABLE II. THE RESULT OF IMAGE FUSION ON FOUR DATASETS. VALUES IN BOLD INDICATE BEST RESULTS FOR EACH CRITERION AND DATA SET.

Method	Multisensor Phantoms			Medical Images			Face Images			Satellite Images		
	$Q^{ab/f}$	*MIM*	*AG*	$Q^{ab/f}$	*MIM*	*AG*	$Q^{ab/f}$	*MIM*	*AG*	$Q^{ab/f}$	*MIM*	*AG*
ER-LP	0.286	0.455	0.092	0.362	0.622	0.476	0.312	0.686	0.271	0.448	0.656	0.439
ER-SS	**0.344**	0.468	0.124	0.444	0.717	0.642	0.358	0.705	0.307	0.536	0.706	0.619
GMM	0.323	0.469	0.135	0.521	0.729	0.734	0.370	0.716	0.363	0.555	**0.730**	0.651
The proposed method	0.338	**0.476**	**0.156**	**0.567**	**0.731**	**0.735**	**0.373**	**0.724**	**0.379**	**0.574**	0.727	**0.697**

V. CONCLUSION

In this paper, a method of joint groupwise image registration and fusion is proposed. The SMM is applied to model groupwise image registration and fusion at the same time and combines these models into an ML formulation. The EM algorithm is then proposed to estimate the relevant parameters. Using the experiments of the multi-sensor phantom, medical image, face image, and satellite image, the proposed method has the better groupwise image registration and fusion performance than other conventional methods.

REFERENCES

[1] R.S. Blum, and Z. Liu, Multi-sensor image fusion and its applications, CRC Press, 2005.
[2] H. Zhu, K. Yuen, L. Mihaylova, and H. Leung, "Overview of environment perception for intelligent vehicles," IEEE Transactions on Intelligent Transportation Systems, vol, 18, no, 10, 2017, pp. 2584-2601.
[3] B. Zitova, and J. Flusser, "Image registration methods: A survey," Image and Vision. Computing, vol. 21, no. 11, 2003, pp. 977-1000.
[4] M. Gong, S. Zhao, L. Jiao, D. Tian, and S. Wang, "A novel coarse-to-fine scheme for automatic image registration based on SIFT and mutual information," IEEE Transactions on Geoscience and Remote Sensing, vol. 52, no. 7, 2014, pp. 4328-4338.
[5] E.G. Learned-Miller, "Data driven image models through continuous joint alignment," IEEE Transactions Pattern Analysis and Machine Intelligence, vol. 28, no. 2, 2006, pp. 236-250.
[6] Q. Wang, G. Wu, P.T. Yap, and D. Shen, "Attribute vector guided groupwise registration," NeuroImage, vol. 50, no. 4, 2010, pp. 1485-1496.
[7] H. Neemuchwala, A. Hero, and P. Carson, "Image matching using alphaentropy measures and entropic graphs," Signal Processing, vol. 85, no. 2, 2005, pp. 277-296.
[8] J. Orchard, and R. Mann, "Registering a multisensor ensemble of images," IEEE Transactions on Image Process, vol. 19, no. 5, 2010, pp. 1236-1247.
[9] H. Zhu, Y. Li, J. Yu, H. Leung, and Y. Li, "Ensemble registration of multisensory images by a variational Bayesian approach," IEEE Sensors Journal, vol. 14, no.8, 2014, pp. 2698-2705.
[10] H. Zhu, Q. Shi, Y. Li, and Q. Wu, "Ensemble image registration by a spatially constrained clustering approach," International Journal of Advanced Robotic Systems, 2016, pp. 1-7.
[11] Y. Zhu, and S.M. Cochoff, "Likelihood maximization approach to image registration," IEEE Transactions on Image Processing, vol. 11, no. 12, 2002, pp. 1417-1426.
[12] S. Chen, Q. Guo, H. Leung, and É. Bossé, "A maximum likelihood approach to joint image registration and fusion," IEEE Transactions on Image Processing, vol. 20, no. 5, 2011, pp. 1363-1372.
[13] Q. Zhang, Z. Cao, Z. Hu, Y. Jia, and X. Wu, "Joint image registration and fusion for panchromatic and multispectral images," IEEE Geoscience and Remote Sensing Letters, vol. 12, no. 3, 2015, pp. 467-471.
[14] Y. Li, Z. He, H. Zhu, W. Zhang, and Y. Wu, "Jointly registering and fusing images from multiple sensors," Information Fusion, vol. 27, no. C, 2016, pp. 85-94.
[15] Sharma, K. Ravi, Leen, K. Todd, Pavel, and Misha, "Bayesian sensor image fusion using local linear generative models," Optical Engineering, vol. 40, no. 7, 2001, pp. 1364-1376.
[16] P. Viola, and W. M. Wells, "Alignment by maximization of mutual information," International Journal of Computer Vision, vol. 24, no. 2, 1997, pp. 137-154.
[17] S. Arivazhagan, V.R.S. Mani, and G. Prema, "Nonrigid image registration using spatial and multiresolution techniques," in the International Conference on Signal Processing, Communication, Computing and Networking Technologies, 2011, pp. 818–823.
[18] Z. Zhang, and R.S. Blum, "A categorization of multiscale decomposition based image fusion schemes with a performance study for a digital camera application," Proceeding of IEEE, vol. 87, no. 8, 1999, pp. 1315-1326.
[19] G. Qu, D. Zhang, and P. Yan, "Information measure for performance of image fusion," Electronics Letters, vol. 38, no. 7, 2002, pp. 313-315.
[20] Q. Guo, S. Chen, and H. Leung, "Covariance intersection based image fusion technique with application to pansharpening in remote sensing," Information Sciences, vol. 180, no. 18, 2010, pp. 3434-3443.