

A Novel Markov Random Field Based Deformable Model for Face Recognition

Shu Liao and Albert C.S. Chung
Lo Kwee-Seong Medical Image Analysis Laboratory,
Department of Computer Science and Engineering,
The Hong Kong University of Science and Technology, Hong Kong

liaoshu@ust.hk, achung@cse.ust.hk

Abstract

In this paper, a new scheme to address the face recognition problem is proposed. Different from traditional face recognition approaches which represent each facial image by a single feature vector as the classification problem, the proposed method establishes a new way to formulate the face recognition problem as a deformable image registration problem. The main contributions of the paper lie in the following aspects: (i) Each pixel is represented by an anatomical feature signature calculated from its corresponding best scale salient region by using a new salient region detector based on the survival exponential entropy (SEE); (ii) The face recognition problem is formulated as a deformable image registration problem, the deformation model is represented by a Markov random field (MRF) labeling framework. Explicit pixel correspondence is established by the deformation framework. (iii) The survival exponential entropy based normalized mutual information (SEE-NMI) is proposed and integrated with the MRF based deformation model as the similarity measure to reflect the similarity between two facial images. The proposed method is evaluated on the FERET and FRGC version 2 databases and compared with several state-of-the-art face recognition approaches. Experimental results show that the proposed method achieves the highest recognition rate among all the compared approaches.

1. Introduction

Automatic face recognition (AFR) is an active research topic in computer vision. Its applications include biometrics, law enforcement, and surveillance development. It has also gained an increasing interest in building natural human-computer interaction (HCI) systems. Face recognition is a challenging task because factors such as pose [6], illumination [17, 12], facial expression [3] make it difficult to achieve high face recognition rate. As the demand

of such applications increases, many novel face recognition methods have been proposed in the last several decades.

Face recognition can be factorized into two essential parts for conventional face recognition approaches: (1) feature extraction from facial images; and (2) similarity measure and classifier design. Many feature extraction methods have been proposed regarding to the first part. For example, Wiskott *et al.* applied the Gabor wavelet features in the elastic bunch graph matching algorithm (EBGM) for face recognition [16]. Turk *et al.* used the principle component analysis (PCA) to compute facial image features [14]. This is commonly called the "eigenface". Belhumeur *et al.* proposed the linear discriminant analysis (LDA) method [2], which finds a linear transformation maximizing the inter-class distance while minimizing the intra-class distance. Manli and Martinez [20] selected the principal components in a two-stage LDA. Hwang *et al.* proposed the hybrid Fourier feature to model facial images [4]. In recent years, a new feature extraction method, called local binary patterns (LBP) was proposed, which was originally used in texture classification [5]. It has also been applied in face recognition [1] and facial expression recognition [11]. For the second part, the Kullback-Leibler distance [13], χ^2 statistics, and etc. are commonly used to measure the similarity between two facial images in classification.

In this paper, we propose a new way to formulate the face recognition problem. More specifically, the face recognition problem is converted to a deformable image registration problem. For each testing facial image, it is deformed to the fixed training facial image space by a pre-defined deformation model. A similarity measure function is defined between two facial images to guide the registration process. The goal of the registration process is to find a deformation field which can minimize/maximize the similarity measure function between the deformed testing facial image and the fixed training image. Finally, the testing facial image is classified to the subject of the training image which is most similar to the deformed testing image according to the similarity measure function. The main contributions of this pa-

per are listed as follows: (1) This paper proposes a new class of method (i.e. deformable image registration) to model the face recognition problem; (2) A feature driven Markov random field (MRF) deformation model is proposed for the registration process, a new salient region detector based on the survival exponential entropy (SEE) is proposed to calculate the best scale anatomical feature signatures for each pixel; (3) A new information-theoretic measure function, namely the survival exponential entropy based normalized mutual information (SEE-NMI) is proposed and integrated with the MRF based deformation model as the similarity measure function to guide the registration process. The proposed method is evaluated on both the FERET and FRGC version 2 databases. It is shown that the proposed method achieves the highest recognition rate among all the compared state-of-the-art approaches.

The paper is organized as follows. In Section 2, we describe the details about how to formulate the face recognition problem as a deformable image registration problem. Section 3 introduces the feature signature calculation process for each pixel and the feature based MRF deformation model. Section 4 gives the definition of the survival exponential entropy based normalized mutual information (SEE-NMI) and describes how to integrate it with the MRF deformation model. Section 5 analyzes the experimental results of the proposed method and the comparisons with other widely used methods on the FERET and FRGC databases. Section 6 concludes the paper.

2. Formulate Face Recognition as a Deformable Image Registration Problem

In this section, we describe the motivation and details of formulating face recognition as a deformable image registration problem.

2.1. Background Knowledge of Deformable Image Registration

In computer vision, groups of images obtained by sampling the same subject at different times, or from different subjects, will be represented in different coordinate systems. The task of image registration is to transform different groups of images into a common coordinate system for comparison purpose.

Figure 1 shows the general flow chart of image registration.

As shown in Figure 1, The basic input data for the registration process are two images: one is the fixed image, denoted as X , the other is the floating image, denoted as Y . The goal of registration is to find an optimal transformation T_{opt} based on a pre-defined transformation model such that the transformed floating image is aligned to the fixed image. The optimal transformation T_{opt} is defined according

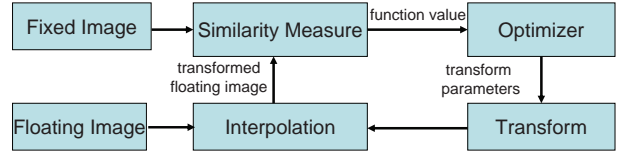


Figure 1. The flow chart of image registration.

to the similarity measure function E , the role of E is to evaluate the matching degree between the current transformed floating image $T(Y)$ and the fixed image X , where T is the transformation found in the current iteration. Therefore, by minimizing/maximizing the similarity measure function E , the corresponding transformation is said to be able to precisely map the floating image to the fixed image. The registration process thus can be illustrated by Equation 1:

$$T_{opt} = \arg \max_T E(X, T(Y)), \quad (1)$$

where we assume that we need to maximize the target function E . After obtaining T_{opt} , the final transformed floating image can be calculated as $T_{opt}(Y)$.

In order to minimize/maximize E , an optimizer is also needed. Moreover, interpolation is also required in case some of the pixels of the transformed floating image do not fall exactly on the image grid of the fixed image. Some commonly used similarity measure functions E in the literatures are: sum of squared difference (SSD), mutual information (MI) and *et al.* Transformation models can also be various, simple transformation models include rigid transform and affine transform [15], which has small degrees of freedom. In order to transform the floating image more accurately to the fixed image, more complicated transformation models are required, such as the B-spline based transformation and the free-form deformation [10].

2.2. Face Recognition as a Deformable Image Registration Process

Face recognition has close connection with the deformable image registration process. For instance, one of the major challenges for face recognition is the pose and facial expression variations. Such variations are actually caused by the deformable process of the muscles and features of faces. Therefore, intuitively, by establishing reliable correspondence between two facial images via deformable image registration, such variations and interference factors can be eliminated. Moreover, if a robust similarity function E is adopted, illumination variation effects can also be reduced. For example, the mutual information (MI) is proved to be robust against global illumination changes, bias fields and image modalities [15].

Therefore, we are motivated to formulate the face recognition problem as a deformable image registration process.

Algorithm 1 Classification of an input testing facial image

Input: testing image Y , n training images X_i ($i=1,\dots,n$). The deformation model T , and the similarity measure function E .

Output: A class label assigned to Y .

1. FOR $i = 1$ to n
2. Register Y to X_i by finding the optimal transformation T_{opt} to minimizing/maximizing E . Record the value of E with respect to T_{opt} , denote as $E_{opt}(i)$.
3. IF $E_{opt}(i)$ is the smallest/largest among $E_{opt}(1)$ to $E_{opt}(i)$
4. Update $E_{global} = E_{opt}(i)$.
5. Update $Label(Y) = \text{Class label of } X_i$.
6. END IF
7. END FOR
8. Return $Label(Y)$

The framework of classifying an input testing image can be summarized by Algorithm 1.

In the following sections, we will explain each component of Algorithm 1 in detail: The deformation model T and the similarity measure function E .

3. The Feature based MRF Deformation Model

In this section, a feature based Markov random field (MRF) deformation model is proposed and adopted as the transformation model in this paper. There are two stages of this deformation model: First, each pixel is represented by a feature vector signature calculated from an anatomical region around it. The scale of the anatomical region is determined by a new salient region detector based on the survival exponential entropy (SEE). Then, the deformation field is represented by a MRF labeling framework.

3.1. Feature Signature for Each Pixel

In this section, we describe the feature signature for each pixel, which will be used to establish geometric correspondence between two facial images by the MRF deformation model.

Since faces are topological objects, different features of faces have different sizes and scales. Therefore, in order to accurately represent the anatomical properties around each pixel p , the most salient scales with respect to the anatomical region R_p centered at p should be first determined. Then anatomical descriptors are calculated from R_p as the signatures for pixel p .

In this paper, we measure the saliency of anatomical regions based on local image complexity. The survival exponential entropy [21] (SEE) is used to statistically measure the local image complexity. SEE is defined as follows [21]:

Definition 1: For a random vector X in R^m , the survival

exponential entropy of order α is [21]:

$$M_\alpha(X) = \left(\int_{R_+^m} \bar{F}_{|X|}^\alpha(x) dx \right)^{\frac{1}{1-\alpha}} \quad (2)$$

for $\alpha \geq 0$, where m defines the number of dimensions for X . $|X|$ denotes the random vector with components $|X_1|, \dots, |X_m|$. The notation $|X| > x$ means that $|X_i| > x_i$ for $x_i \geq 0, i = 1, \dots, m$. The multivariate survival function $\bar{F}_{|X|}(x)$ of the random vector $|X|$ is defined as:

$$\bar{F}_{|X|}(x) = P(|X| > x) = P(|X_1| > x_1, \dots, |X_m| > x_m) \quad (3)$$

for $x \in R_+^m$ with R_+^m defined by Equation 4:

$$R_+^m = \{x \in R^m : x = (x_1, \dots, x_m), x_i \geq 0, i = 1, \dots, m\}. \quad (4)$$

As compared with the conventional Shannon's entropy, SEE has several advantages: (1) SEE is always nonnegative; (2) SEE have consistent definitions in both the continuous and discrete domains; (3) The Shannon's entropy is calculated based on the density of the random variable $p(X)$. However, $p(X)$ may not exist as stated in [13]. SEE is calculated based on the survival function which always exists.

In this paper, the random vector X is the local intensity probability distribution histogram of a squared region with scale s (i.e. the side length of the square) centered at each pixel p . Let $H(s, p)$ denote the local intensity probability distribution calculated from the squared region centered at p with side length s . It is obvious that the larger of the value of $M_\alpha(H(s, p))$, the more image complexity contained in the squared region. However, if that region exhibits high degrees of self-similarity over a wide range of scales, it should also be considered as non-salient (e.g. textures). Therefore, the scale saliency measure function of each pixel p with scale s , denoted as $A(s, p)$, is defined by the SEE value of $H(s, p)$ multiplied by the Jensen Shannon divergence between $H(s, p)$ and $H(s - \Delta s, p)$ as in Equation 5:

$$A(s, p) = M_\alpha(H(s, p)) \cdot JSD(H(s, p), H(s - \Delta s, p)), \quad (5)$$

where $JSD(\cdot)$ denotes the Jensen Shannon divergence, $s - \Delta s$ denotes the predecessor scale with respect to scale s .

For each pixel p , the best scale S_p associated with it is defined as:

$$S_p = \arg \max_s A(s, p). \quad (6)$$

Based on the experimental results, it is found that $\alpha = 3$ is a good choice as the order parameter of SEE. The scale space is defined from 4 to 20 pixels, that is, the best scale of each pixel is selected from the set of candidate scales $\{s | s = 4, 5, \dots, 20\}$, the discrete scale difference parameter Δs is set to 1. Such settings are used throughout the whole paper.

Figure 2 shows the top 10 salient regions calculated by Equation 5 with their corresponding center pixel p located at the center of the region and the best scale S_p as the side length of the squares. Figures 2 (a) and (b) are facial images taken from the same subject but with different facial expressions in the FERET database, while Figure (c) is a facial image taken from a different subject in the FERET database.

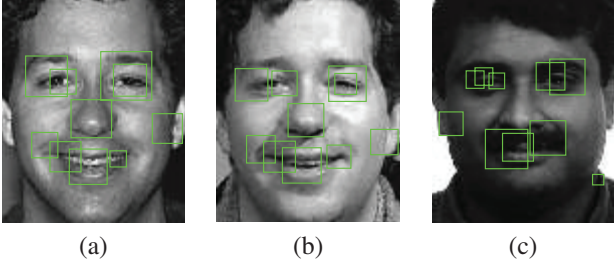


Figure 2. The top 10 most salient regions detected by using Equation 5, with the referencing pixel at the center of each squared region. The best scale is represented by the side length of each squared region. (a) and (b) are two facial images of the same subject taken from FERET, (c) is a facial image of a different subject taken also from FERET.

It is observed that the locations of the most salient pixels and their corresponding best scales are very similar with each other in Figures (a) and (b), but are significantly different from each other compared to Figure (c). It demonstrates the effectiveness of the proposed best scale salient measure expressed by Equation 5, as indeed Figures 2 (a) and (b) are from the same subject and it exhibits high degrees of repeatability, which can also clearly distinguish the one from Figure 2 (c).

One point should be emphasized is that, determining the best scale of each pixel is of essential importance. For example, the salient region detected around the nose region in Figure 2 (a), the region will become non-salient if the scale of the squared region becomes smaller. Therefore, saliency of each pixel is closely related to the scales under consideration.

After determining the best scale S_p of each pixel p , we then extract anatomical features from the squared regions V_p centered at p with side length S_p as signatures for pixel p . In this paper, The histogram of the uniform local binary patterns (LBP) [1] calculated from V_p is used as signature for each pixel p , denoted as \vec{f}_p . The radius of LBP is set to 2 and the number of neighboring samples is set to 8 following the same line in [1].

It is worth pointing out that the proposed method is different from the conventional LBP [1] for face recognition. The conventional LBP method extracts a single feature vector of each pre-defined partitioned region of facial images, combining the feature vectors extracted from each parti-

tioned region together, and pass them to the classifier. While our approach used LBP features as signatures for each pixel, which is used to establish correspondence between different pixels in the deformation model to be discussed in the next section.

3.2. The Markov Random Field Deformation Model

After calculating the feature signatures \vec{f}_p for each pixel p with respect to its best scale S_p , it is served as the input for a Markov random field (MRF) deformation model.

The general form of the MRF energy function can be expressed as:

$$\begin{aligned} E &= E_{data} + E_{smoothness} \\ &= \sum_{p \in \Omega} D_p(l_p) + \sum_{(p,q) \in N} V_{p,q}(l_p, l_q), \end{aligned} \quad (7)$$

where Ω is the set of pixels, N is the neighborhood system defined in Ω . In this paper, the 4-connected neighborhood system is used. $D_p(l_p)$ is the energy function for the data term, which penalizes the cost of assigning label l_p to pixel p , and $V_{p,q}(l_p, l_q)$ is the energy function related to the smoothness term penalizes the cost of label discrepancy between two neighboring pixels.

The registration problem is converted to the MRF labeling problem by quantizing the deformation space. A discrete set of labels $L \in \{l^1, l^2, \dots, l^n\}$ is defined. Each label l^i ($1 \leq i \leq n$) corresponds to a displacement vector \vec{d}_i . Each label assignment l_p to pixel p denotes moving p to a new position according to the corresponding displacement vector \vec{d}_{l_p} . In this paper, each pixel can be displaced off the original position bounded by a discretized window $\Psi = \{0, \pm s, \pm 2s, \dots, \pm ws\}^d$ of dimension d . Since we are dealing with 2D facial images, the dimension d is 2. The 2D displacement window for the MRF deformation field in this paper is set to: $\Psi = \{0, \pm 1, \pm 2, \dots, \pm 12\}^2$.

The data term energy function $D_p(l_p)$ is defined based on the feature signature for each pixel as:

$$\begin{aligned} D_p(l_p) &= D_p(G_{fixed}(p), G_{float}(p + \vec{d}_{l_p})) \\ &= D_p(\vec{f}_p^{fixed}, \vec{f}_{p+\vec{d}_{l_p}}^{float}) \\ &= JSD(\vec{f}_p^{fixed} || \vec{f}_{p+\vec{d}_{l_p}}^{float}), \end{aligned} \quad (8)$$

where G_{fixed} is the fixed image, G_{float} is the floating image, \vec{f}_p^{fixed} denotes the feature vector signature for pixel p of the fixed image, $\vec{f}_{p+\vec{d}_{l_p}}^{float}$ denotes the feature vector signature of p after p is displaced off the original position according to the displacement vector \vec{d}_{l_p} of the floating image. $JSD(\cdot)$ denotes the Jensen Shannon divergence. Therefore, we can observe that the data term energy function actually

is establishing pixel correspondence between two facial images based on comparing the feature signature similarities between two pixels.

The piece-wise truncated absolute distance is used as the energy function related to the smoothness term:

$$V_{p,q}(l_p, l_q) = \min(\lambda, |\vec{d}_{l_p} - \vec{d}_{l_q}|), \quad (9)$$

where λ is a constant representing the maximum penalty. λ is set to 15 in this paper. The truncated absolute distance is a metric as stated in [18].

The energy function defined in Equations 7 is optimized by the α -expansion algorithm [18].

It should be noted that the energy function expressed by Equation 7 is defined based on the local pixel information. As both the data term and smoothness term are defined based on comparing the individual pixel-wised feature signature similarity and label similarity. Such local information may become unstable in the presence of the global illumination changes. Therefore, in the next section we will introduce the survival exponential entropy based normalized mutual information (SEE-NMI), which is a global measure metric and can serve as the complement for the original MRF energy term defined by Equation 7. SEE-NMI is a generalized information-theoretic metric which shares the robustness property of the conventional mutual information (MI) such as the robustness against bias fields and illumination changes.

4. The Survival Exponential Entropy based Normalized Mutual Information

In this section, we propose a global similarity measure metric, namely the survival exponential entropy based normalized mutual information (SEE-NMI) to integrate with the MRF model proposed in Section 3. SEE-NMI is defined based on SEE and shares similar properties of the conventional MI [15], such as its robustness against image bias fields and illumination changes.

The SEE-NMI is defined by Equation 10:

$$R(G_{fixed}, T(G_{float})) = \frac{M_\alpha(G_{fixed}) + M_\alpha(T(G_{float}))}{M_\alpha(G_{fixed}, T(G_{float}))}, \quad (10)$$

where G_{fixed} denotes the fixed image, $T(G_{float})$ denotes the transformed floating image under the transformation model T . $M_\alpha(G_{fixed})$ and $M_\alpha(T(G_{float}))$ denote the survival exponential entropy of the intensity probability distribution function of the fixed image and transformed floating image respectively. $M_\alpha(G_{fixed}, T(G_{float}))$ denotes the joint survival exponential entropy of the joint intensity probability distribution function of the fixed and transformed floating images. α is the order of the survival exponential entropy, as discussed in Section 3.1, $\alpha = 3$ is used in

this paper. SEE-NMI not only shares the robustness properties of the conventional MI [15], but also has the advantages of the survival exponential entropy defined in Section 3.1.

The physical meaning of SEE-NMI is the amount by which the uncertainty about the intensity distribution of $T(G_{float})$ decreases when the intensity distribution of G_{fixed} is provided. In other words, it reflects the amount of the information of the intensity distribution of G_{fixed} contains about the intensity distribution of $T(G_{float})$. The optimal transformation T_{opt} is assumed to be achieved by maximizing SEE-NMI, where the images are aligned such that the amount of intensity distribution information they contain about each other is maximized.

It should be noted that SEE-NMI is calculated based on the global intensity distribution information of the fixed and floating images. Therefore, it is more robust against the regional illumination changes than the local information based metric defined in Equation 7 for the original MRF model. Thus, SEE-NMI is integrated with the original MRF model by including it as the third energy term in the MRF model. Now the overall energy function is defined as:

$$\begin{aligned} E &= E_{data} + E_{smoothness} - E_{global} \\ &= \sum_{p \in \Omega} D_p(l_p) + \sum_{(p,q) \in N} V_{p,q}(l_p, l_q) \\ &\quad - R(G_{fixed}, T(G_{float})), \end{aligned} \quad (11)$$

where $R(G_{fixed}, T(G_{float}))$ denotes the SEE-NMI function defined in Equation 10, the data term energy E_{data} and the smoothness term energy $E_{smoothness}$ are with the same definition as the ones defined in Equation 7. Therefore, the registration process is now formulated as an optimization problem to minimize the energy function defined in Equation 11. Again, the α -expansion algorithm [18] is used to optimize the energy function.

Figure 3 demonstrates the effectiveness and robustness of the energy function defined in 11 under different interference conditions such as different pose and facial expressions and illumination changes. Figure 3 (a) is a facial image obtained from the FERET database with neutral facial expression, which is served as the fixed image. Figure 3 (b) is the facial image of the same subject as in Figure 3 (a) but with different facial expressions. Figure 3 (c) is the facial image of the same subject as in Figure 3 (a) but is taken under different illumination conditions. Figures 3 (d) and (e) are the results of registering the images in Figures 3 (b) and (c) to the fixed image in Figure 3 (a), Figures 3 (f) and (g) are their corresponding deformation fields.

It is observed that after the deformable image registration process, both the facial expression variations and illumination changes can be satisfactorily recovered with respect to the fixed image in Figure 3 (a), which is evidence from the high degrees of visual similarity between the registered images in Figures 3 (d) and (e) and the fixed image in Figure 3

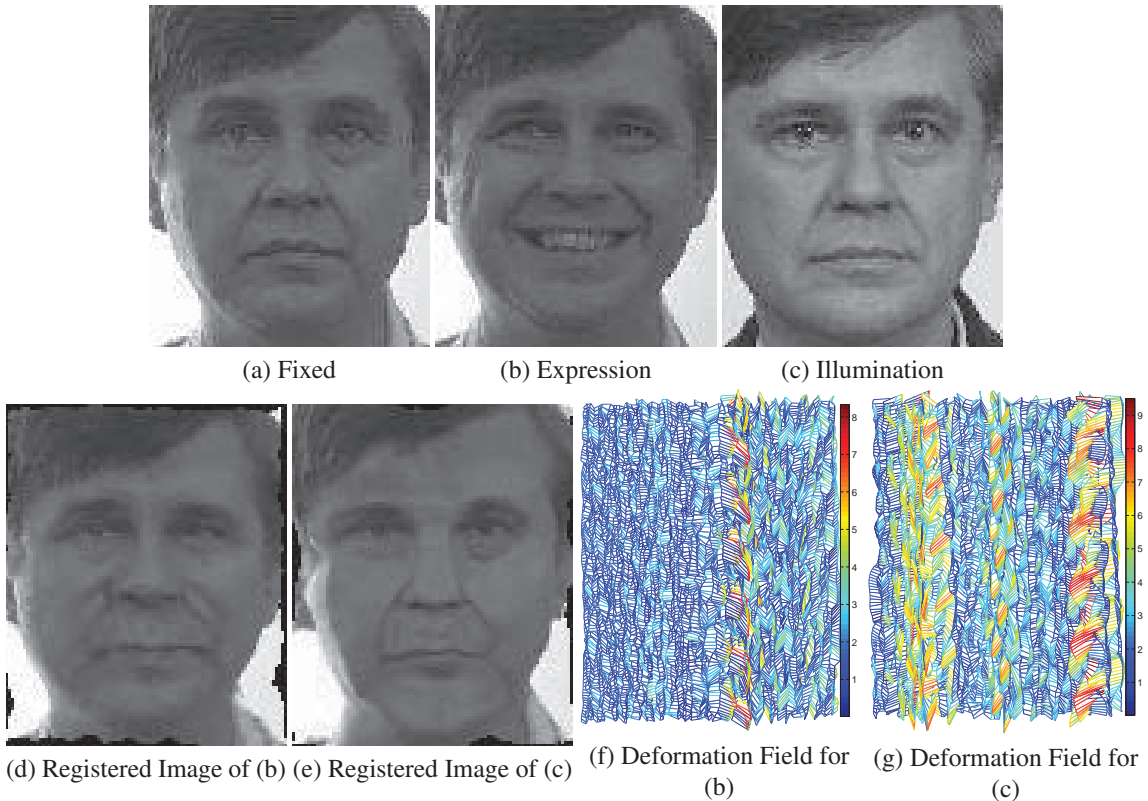


Figure 3. (a) The fixed image; (b) Image of the same subject as (a) but with different facial expression, (c) Image of the same subject as (a) but with different illumination condition, (d) Resulting image after registering image (b) to the fixed image (a); (e) Resulting image after registering image (c) to the fixed image (a); (f) The corresponding deformation field for transforming (b) to (d); (g) The corresponding deformation field for transforming (c) to (e).

(a). Therefore, the effectiveness and robustness of the proposed method is strongly implied.

5. Experimental Results

In this section, systematic experiments were conducted to evaluate the proposed method on both the full FERET database and the FRGC version 2 database. The proposed method is also compared with several state-of-the-art approaches. In all experiments, the nearest neighbor classifier was used in this paper.

5.1. Experiments on the FERET Database

To perform face recognition experiment on the FERET database, we used the same Gallery and Probe sets as the standard FERET evaluation protocol. In the FERET database, Fa is used as Gallery, which contains 1196 frontal images of 1196 subjects. Fb contains 1195 images of expression variations, Fc contains 194 images taken under different illumination conditions, Dup I has 722 images taken later in time and Dup II is a subset of Dup I which consists of 234 images that were taken at least one year af-

ter the corresponding Gallery images. Figure 4 shows some sample images from the FERET database.



Figure 4. Samples in the FERET Database.

In this paper, all facial images were cropped according to the manually located eyes positions supplied with the FERET data and normalized to the resolution of 128×128 pixels. The rank-1 recognition rates (in %) of different approaches are listed in Table 1. The first two rows are the

results obtained by the proposed method with and without adding the SEE-NMI energy function to the MRF deformation model.

Methods	Fb	Fc	Dup I	Dup II
1. <i>MRF</i> + SEE-NMI	98.2	98.8	83.2	79.4
2. <i>MRF</i>	96.4	91.6	79.2	74.7
3. <i>LGBPHS</i> [19]	98.0	97.0	74.0	71.0
4. <i>LBP</i> [1]	97.0	79.0	66.0	64.0
5. <i>PCA</i> [14]	85.0	65.0	44.0	22.0
6. <i>UMD LDA</i> [9]	96.2	58.8	47.2	20.9
7. <i>USC EBGM</i> [9]	95.0	82.0	59.1	52.1
8. <i>Bayesian, MAP</i> [8]	82.0	37.0	52.0	32.0

Table 1. The rank-1 recognition rates of different approaches according to the same FERET evaluation protocol, where *MRF* denotes the proposed method without using the SEE-NMI metric in the MRF deformation model, *MRF* + SEE-NMI denotes the proposed method with the SEE-NMI metric in the MRF deformation model

From Table 1, we can observe that even without integrating the SEE-NMI metric into the MRF deformation model (Row 2), the proposed method can achieve high recognition rates among all the compared methods. Especially for the probe set Dup I and Dup II, where the variations caused by aging are satisfactorily captured by the MRF deformation model. It is also observed that the recognition rate of the proposed method without SEE-NMI for probe set Fc is lower than the *LGBPHS* [19] approach. The reason is analyzed in Section 3.2, where the MRF deformation model without integrating the global similarity metric SEE-NMI only considers the local pixel information, therefore it is not very robust against the illumination change interference. However, after integrating the SEE-NMI into the MRF deformation model, the proposed method has the highest recognition rates among all the compared methods for all the probe sets. Moreover, the recognition rate of each probe set with the SEE-NMI metric is also higher than the one obtained without using the SEE-NMI metric. Therefore, the complementary property of the SEE-NMI metric for the original MRF deformation model defined by Equation 7 is also illustrated.

5.2. Experiment on the FRGC Database

To further evaluate the performance of the proposed method, we also evaluate the proposed method on the FRGC version 2 database [7] which is known as the largest face image data sets available. Before we performed the experiment, the facial images were normalized and cropped to the size of 120×120 to extract the facial region based on the eye positions provided by the FRGC. Figure 5 shows some example facial images from the FRGC database after the cropping pre-processing step.



Figure 5. Sample facial images in the FRGC Database. Images in the FRGC database are taken under both the controlled and uncontrolled illumination conditions.

In the FRGC version 2 database, there are 12776 images taken from 222 subjects in the training set, 16028 target images taken under the controlled illumination condition. We conduct experiment 1 and experiment 4 protocols to evaluate the performance of the proposed method. In experiment 1, there are 16028 query images taken under the controlled illumination condition, the goal to perform experiment 1 is to test the basic recognition abilities of different approaches. In experiment 4, there are 8014 query images taken under the uncontrolled illumination condition. Experiment 4 is the most challenging protocol in FRGC as the uncontrolled large illumination variation brings significant difficulties to achieve high recognition rate.

The experimental results for the FRGC version 2 database in experiment 1 and experiment 4 are evaluated by using the Receiving Operator Characteristics (ROC), which is the face verification rate (FVR) versus the false accept rate (FAR). There are three ROC values, ROC 1 corresponding to images collected within semester, ROC 2 corresponding to images collected within year and ROC 3 corresponding to images collected between semesters. Tables 2 and 3 list the performance of different approaches on the face verification rate (FVR) at the false accept rate (FAR) of 0.1% in experiment 1 and experiment 4:

Methods	FVR at FAR = 0.1% (in %)		
	ROC 1	ROC 2	ROC 3
1. BEE Baseline [14]	77.61	75.17	70.63
2. <i>LBP</i> [1]	86.10	83.27	79.23
3. <i>LGBPHS</i> [19]	92.75	91.16	87.85
4. <i>MRF</i>	95.37	93.25	89.79
5. <i>MRF</i> + SEE-NMI	97.51	95.87	92.60

Table 2. The FVR value of different approaches at FAR = 0.1% in experiment 1 of the FRGC version 2 database.

It is observed from Tables 2 and 3 that the proposed method significantly outperforms the BEE baseline algo-

Methods	FVR at FAR = 0.1% (in %)		
	ROC 1	ROC 2	ROC 3
1. BEE Baseline [14]	17.08	15.14	13.52
2. LBP [1]	26.83	22.64	19.08
3. LGBPHS [19]	31.07	28.84	23.92
4. MRF	63.21	60.07	57.40
5. MRF + SEE-NMI	74.16	73.38	71.77

Table 3. The FVR value of different approaches at FAR = 0.1% in experiment 4 of the FRGC version 2 database

rithm (PCA), the conventional LBP method [1] and the LGBPHS method [19] under both the controlled and uncontrolled conditions. Especially for experiment 4, which is the most challenging protocol in FRGC version 2 database, the proposed method achieves the FVR rate of 71.77% in ROC 3 when the SEE-NMI metric is integrated with the MRF deformation model, which is significantly higher than the recognition rates obtained by the other compared methods.

6. Conclusion

In this paper, a new way of formulating the face recognition problem is proposed. The face recognition problem is formulated as a deformable image registration problem. A feature based Markov random field (MRF) deformation model is proposed in this paper to drive the registration process. The feature signature for each pixel is represented by the histogram of the uniform local binary pattern distribution calculated from the best scale salient region detected by a new best scale salient region detector. A new information-theoretic similarity measure named the survival exponential entropy based normalized mutual information (SEE-NMI) is proposed and integrated with the MRF deformation model. It is shown that the SEE-NMI metric brings complementary information to the MRF deformation model which is based on the local pixel information. The proposed method is evaluated on the FERET and FRGC databases. It is shown that the proposed method outperforms all the compared state-of-the-art and baseline algorithms, which illustrates the robustness of the proposed method against the appearance variations of expression, lighting and aging. The proposed method hopefully can inspire a new thinking and new way to tackle the face recognition problem.

References

- [1] T. Ahonen, A. Hadid, and M. Pietikainen. Face recognition with local binary patterns. In *ECCV*, volume 1, pages 469–481, 2004.
- [2] P. Belhumeur, J. Hespanha, and D.J. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE PAMI*, 19(7):711–720, 1997.
- [3] B. Fasel and J. Luetttin. Automatic facial expression analysis: A survey. *Pattern Recognit.*, 36(1):259–275, 2003.
- [4] W. Hwang, G. Park, and J. Lee. Multiple face model of hybrid fourier feature for large face image set. In *CVPR*, pages 1574–1581, 2006.
- [5] T. Ojala, M. Pietikainen, and T. Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE PAMI*, 24(7):971–987, 2002.
- [6] Z. Pan, G. Healey, and etal. Face recognition in hyperspectral images. *IEEE PAMI*, 25(12):1552–1560, 2003.
- [7] P. Phillips, P. Flynn, and etal. Overview of the face recognition grand challenge. In *CVPR*, pages 947–954, 2005.
- [8] P. Phillips, H. Syed, and etal. The feret evaluation methodology for face-recognition algorithms. *IEEE PAMI*, 22(10):1090–1104, 2000.
- [9] S. Ravela and R. Manmatha. Retrieving images by appearance. In *ICCV*, pages 608–613, 1998.
- [10] D. Rueckert et al. Nonrigid registration using free-form deformations: application to breast mr images. *IEEE Trans. Med. Imaging*, 18:712–721, 1999.
- [11] C. Shan, S. Gong, and P. McOwan. Robust facial expression recognition using local binary patterns. In *ICIP*, pages 370–373, 2005.
- [12] Z. Taiping, T. Yan, F. Bin, S. Zhaowei, and L. Xiaoyu. Face recognition under varying illumination using gradientfaces. *IEEE TIP*, 18(11):2599–2606, 2009.
- [13] M. C. Thomas and J. A. Thomas. *Elements of Information Theory*. Wiley-Interscience, 1991.
- [14] M. Turk and A. Pentland. Eigenfaces for recognition. *J. Cogn. Neurosci*, 3:71–86, 1991.
- [15] W. Wells, P. Viola, and et al. Multi-Modal Volume Registration by Maximization of Mutual Information. *MedIA*, 1(1):35–51, 1996.
- [16] L. Wiskott, J.-M. Fellous, and etal. Face recognition by elastic bunch graph matching. *IEEE PAMI*, 19(7):775–779, 1997.
- [17] X. Xie and K. M. Lam. Face recognition under varying illumination based on a 2d face shape model. *Pattern Recognit.*, 38(2):221–230, 2005.
- [18] B. Yuri, V. Olga, and Z. Ramin. Fast approximate energy minimization via graph cuts. *IEEE PAMI*, 23:1222–1239, 2001.
- [19] W. Zhang, S. Shan, and etal. Local gabor binary pattern histogram sequence(lgbphs): A novel non-statistical model for face representation and recognition. In *ICCV*, pages 786–791, 2005.
- [20] M. Zhu and A. Martinez. Selecting principal components in a two-stage lda algorithm. In *CVPR*, volume 1, pages 132–137, 2006.
- [21] K. Zografos and S. Nadarajah. Survival exponential entropies. *IEEE Trans. on Information Theory*, 51(3):1239–1246, 2004.