ISSN (ONLINE):2395-695X
ISSN (PRINT):2395-695X

International Journal of Advanced Research in Basic Engineering Sciences and Technology (IJARBEST)
Vol.2, Issue.12, December 2016

Atlas Construction and Sparse Representation for the Recognition of Facial Expression

T.Jeslin[1], R. Ravi[2]
Assistant Professor, Department of ECE, Universal College of Engineering and Technology, Vallioor, Tirunelveli[1]
Professor & Research Centre Head, Francis Xavier Engineering College, Tirunelveli[2]

Abstract— **Automatic facial expression recognition has essential real world applications. A new dynamic facial expression recognition method is proposed. This dynamic facial expression recognition is formulated as a longitudinal GroupWise registration problem. The main contributions of this method lie in the following aspects: (1) Subject-specific facial feature movements of different expressions are described by a diffeomorphic growth model; (2) Salient longitudinal facial expression atlas is built for each expression by a sparse group wise image registration method, which consists of two stages, Atlas construction stage and Recognition stage. (3) Both the image appearance information in spatial domain and topological evolution information in temporal domain are used to guide recognition by a sparse representation method. This framework is also compared with several state-of-the-art dynamic facial expression recognition methods. The experimental results demonstrate that the recognition rates of the new method are consistently higher than other methods under comparison.**

I. Introduction

Digital Image Processing

Digital image processing is the use of computer algorithms to perform image processing on digital images. As a subcategory or field of digital signal processing, digital image processing has many advantages over analog image processing. It allows a much wider range of algorithms to be applied to the input data and can avoid problems such as the build-up of noise and signal distortion during processing. Considerable advances have been made over the past 30 years resulting in routine application of image processing to problems in medicine, manufacturing, entertainment, law enforcement, and many others. Examples include mapping internal organs in medicine using various scanning technologies (image reconstruction from projections), automatic fingerprint recognition (pattern recognition and image coding), and HDTV (video coding).

Automatic facial expression recognition (AFER) has essential real world applications. Its applications include, but are not limited to, human computer interaction (HCI), psychology and telecommunications. It remains a challenging problem and active research topic in computer vision, and many novel methods have been proposed to tackle the automatic facial expression recognition problem.

Intensive studies have been carried out on AFER problem in static images during the last decade: Given a query facial image, estimate the correct facial expression type, such as anger, disgust, happiness, sadness, fear or surprise. It mainly consists of two steps: feature extraction and classifier design. For feature extraction, Gabor wavelet, local binary pattern (LBP), and geometric features such as active appearance model (AAM) are in common use. For classifier, support vector machine is frequently used. Joint alignment of facial images under unconstrained condition has also become an active research topic in AFER.

In recent years, dynamic facial expression recognition has become a new research topic and receives more and more attention. Different from there cognition problem in static images, the aim of dynamic facial expression recognition is to estimate facial expression type from an image sequence captured during physical facial expression process of a subject. The facial expression image sequence contains not only image appearance information in the spatial domain, but also evolution details in the temporal domain. The image appearance information together with the expression evolution information can further enhance recognition performance. Although the dynamic information provided is useful, there are challenges regarding how to capture this information reliably and robustly. For instance, a facial expression sequence normally constitutes of one or more onset, apex and offset phases. In order to capture temporal information and make temporal information of training and query sequences comparable, correspondences between different temporal phases need to be established. As facial actions over time are different across subjects, it remains an open issue how a common temporal feature for each expression among the population can be effectively encoded while suppressing subject-specific facial shape variations.

In this paper, a new dynamic facial expression recognition method is presented. It is motivated by the fact that facial expression can be described by diffeomorphic motions of muscles beneath the face. Intuitively, 'diffeomorphic' means the motion is topologically preserved and reversible. The formal

ISSN (ONLINE):2395-695X
ISSN (PRINT):2395-695X

International Journal of Advanced Research in Basic Engineering Sciences and Technology (IJARBEST)
Vol.2, Issue.12, December 2016

definition of ‘diffeomorphic’ transformation is given in Section II. Different from previous works by using pairwise registration to capture the temporal motion, this method considers both the subject-specific and population information by a GroupWise diffeomorphic registration scheme. Moreover, both the spatial and temporal information are captured with a unified sparse representation framework. Our method consists of two stages: atlas construction stage and recognition stage. Atlases, which are unbiased images, are estimated from all the training images belonging to the same expression type with GroupWise registration.

Atlases capture general features of each expression across the population and can suppress differences due to inter-subject facial shape variations. In the atlas construction stage, a diffeomorphic growth model is estimated for each image sequence to capture subject specific facial expression characteristics. To reflect the overall evolution process of each expression among the population, longitudinal atlases are then constructed for each expression with group wise registration and sparse representation. In the recognition stage, we first register the query image sequence to atlas of each expression. Then, the comparison is conducted from two aspects: image appearance information and temporal evolution information. The preliminary work has been reported in .For the proposed method, there are three main contributions and differences compared to the preliminary work in: A more advanced atlas construction scheme is used. The atlases are constructed using the conventional GroupWise registration method, thus lots of subtle and important anatomical details are lost due to the naive mean operation.

To overcome this shortage, a sparse representation based atlas construction method is proposed in this paper. It is capable of capturing subtle and salient image appearance details to guide recognition, and preserving common expression characteristics. In the recognition stage, the previous method in compared image differences between the warped query sequence and atlas sequence, which is based on image appearance information only. In this paper, the temporal evolution information is also taken into account to drive the recognition process. It has shown to provide complementary information to image appearance information and can significantly improve the recognition performance. The proposed method has been evaluated in a systematic manner on five databases whose applications vary from posed dynamic facial expression recognition to spontaneous pain expression monitoring. Moreover, possible alternatives have been carefully analyzed and studied with different experimental settings.

The rest of the paper is organized as follows: Section II describes the proposed method. Section III analyzes experimental results. Section IV concludes the paper.

I. PROPOSED METHOD

A new dynamic facial expression recognition method is proposed. It is motivated by the fact that facial expression can be described by diffeomorphic motions of muscles beneath the face. The 'diffeomorphic' means the motion is topologically preserved and reversible. This method is different from previous works by using pairwise registration to capture the temporal motion. This method considers both the subject-specific and population information by a GroupWise diffeomorphic registration scheme. Both the spatial and temporal information are captured with a unified sparse representation framework. This method consists of two stages: atlas construction stage and recognition stage. Atlases, which are unbiased images, are estimated from all the training images belonging to the same expression type with GroupWise registration. Atlases capture general features of each expression across the population and can suppress differences due to inter-subject facial shape variations. In the atlas construction stage, a diffeomorphic growth model is estimated for each image sequence to capture subject- specific facial expression characteristics. To reflect the overall evolution process of each expression among the population, longitudinal atlases are then constructed for each expression with GroupWise registration and sparse representation. In the recognition stage, first register the query image sequence to atlas of each expression. Then, the comparison is conducted from two aspects: image appearance information and temporal evolution information.

A. GroupWise Registration

As facial expression process is topologically preserved and reversible, as illustrated in Figure 1, it can be considered as a diffeomorphic transformation of facial muscles. Therefore, the diffeomorphic transformation during the evolution process of facial expression can be used to reconstruct facial feature movements and further guide the recognition task. Given P facial expression images $I_1 \dots, I_p$, a straightforward solution to transform them to a common space is to select one image as the template, then register the remaining P−1 images to the template by applying P −1 pairwise registration. The registration quality is sensitive to the selection of template, where the template is estimated to be the Frechet mean on

ISSN (ONLINE):2395-695X
ISSN (PRINT):2395-695X

International Journal of Advanced Research in Basic Engineering Sciences and Technology (IJARBEST)
Vol.2, Issue.12, December 2016

the Riemannian manifold whose geodesic distances are measured based on diffeomorphisms. The diffeomorphic GroupWise registration problem can be formulated as the optimization problem by minimizing:

$$\hat{I}^{opt}, \psi_1^{opt}, \ldots, \psi_P^{opt} = \arg min \sum_{i=1}^{P} (d(\hat{I}, \psi_i\,(I_i))^2 + \lambda R(\psi_i)) \qquad (1)$$

Where, both the template $\hat{I}^{opt}$ and the optimal diffeomorphic transformation ψ_i^{opt} (i = 1, P) that transforms I_i to $\hat{I}^{opt}$ are variables to be estimated.

Fig.1 Diffeomorphic transformation process

D (·) is the similarity function that measures the matching degree between two images, R (·) denotes the regularization term to control the smoothness of transformation, and λ is a parameter to control the weight of R (·). $\hat{I}^{opt}$ And ψ_i^{opt} can be estimated by a greedy iterative estimation strategy. First, initialize $\hat{I}$ as the mean image of I_i.

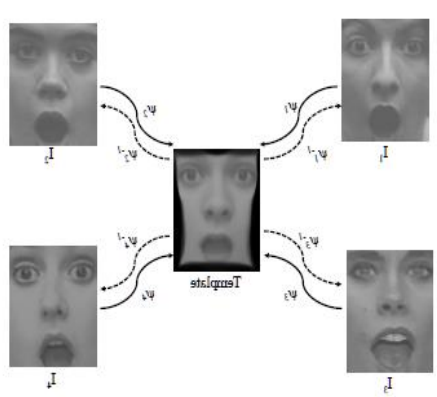

Fig.2 Diffeomorphic GroupWise image registration

ISSN (ONLINE):2395-695X
ISSN (PRINT):2395-695X

International Journal of Advanced Research in Basic Engineering Sciences and Technology (IJARBEST)
Vol.2, Issue.12, December 2016

Fig.2 illustrates the diffeomorphic GroupWise image registration. The estimated template, which is named as atlas, represents overall facial feature changes of a specific expression among the population. The atlas is unbiased to any individual subject and reflects the general expression information.

In this atlas construction stage, the longitudinal facial expression atlases are constructed to obtain salient facial feature changes during an expression process. Given K types of facial expressions of interest, and C different subject image sequences for each expression, denote the image at the *j*th time point of the *i*th subject (i = 1,...,C) as $I_{t_j^i}$. Assume each image sequence begins at time point 0 and ends at time point 1. For each expression, to construct N atlases at given time points T = {t_1, TN}, where $t_k \in$ [0, 1] (k = 1, N), this formulate it as an energy minimization problem by minimizing:

$$M_{t,}\ \phi^i = \arg\min \sum_{t\in T} \sum_{i=1}^{C} \{(d(\widetilde{M}_t, \tilde{\phi}^i_{(t_0^i \to t)}\ (I_{t_0^i}))^2\ +\ \lambda_{\phi^i}\ R(\tilde{\phi}^i)\} \tag{2}$$

Equation 2 can be interpreted as following. First, the subject-specific growth model ϕ^i is estimated for each subject *i*. Then, propagate the subject-specific information to each time point t∈T and construct atlas. Given a subject *i*, there are n_i images in facial expression image sequence. The growth model ϕ^i of subject *i* can be estimated by minimizing the energy function:

$$J(\phi^i) = \int_0^1 \| v_s^i\ \|_U^2\ ds + \frac{1}{\sigma^2} \sum_{j=0}^{n_i-1} \|\ \phi^i_{\left(t_0^i \to t_j^i\right)} \left(I_{t_0^i}\right) - I_{t_j^i}\ \|_2^2 \tag{3}$$

The first term of Equation 3 controls the smoothness of growth model. In the second term, the growth model is applied to $I_{t_0^i}$ and warped to other time points t_j^i, then, the results are compared with existing observations $I_{t_j^i}$ at time points t_j^i. A smaller difference between the warped result and the observation indicates that the growth model can describe the expression more accurately.

Equation 3 estimates the growth model by considering differences at all available time points, which is reflected by the summation in the second term. The least number of images n_i in the subject-specific facial expression sequence used to estimate the growth model.

Given the estimate dϕ_i, is able to construct facial expression atlas at any time point of interest. Assume there are N time points of interest T = {t_1, TN} to construct facial expression atlas. Based on the

ISSN (ONLINE):2395-695X
ISSN (PRINT):2395-695X

International Journal of Advanced Research in Basic Engineering Sciences and Technology (IJARBEST)
Vol.2, Issue.12, December 2016

estimated growth model ϕ_i, subject i's facial expression image can be interpolated at time point t ∈ T with operation$\phi^{i}_{(t^i_0 \rightarrow t)}(I_{t^i_0})$. Moreover, the optimization of Equation 2 with respect to variable M_t becomes:

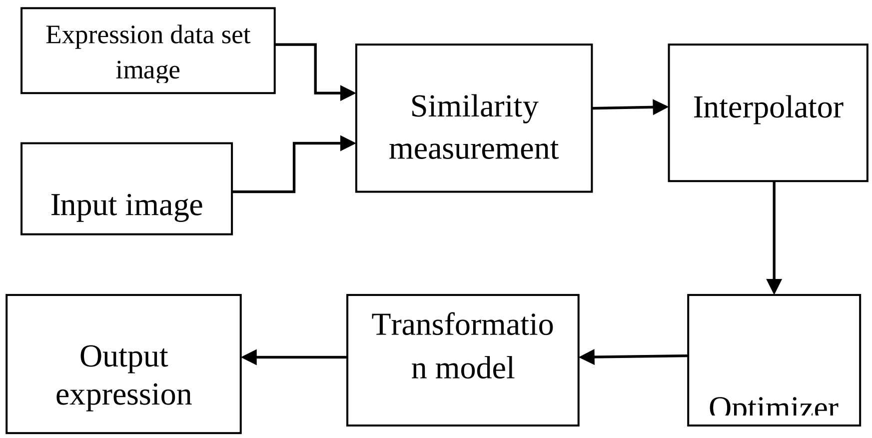

Fig.3 Flow chart for GroupWise image registration of facial expression recognition system.

$$J(M_t) = \sum_{t \in T} \sum_{i=1}^{C} \{(d(M_{t,} \phi^{i}_{(t^i_0 \rightarrow t)}(I_{t^i_0}))^2\} \quad (4)$$

The optimization of Equation 4 can be formulated as a GroupWise image registration problem.

B. The Importance of Keeping Sharp Mean Image in Registration

GroupWise registration method seeks to alternatively estimate the group mean and register each subject to the tentative group mean. However, the initial group mean image M^0 generated right after the linear alignment.

C. Objective Function of GroupWise Registration

An objective functions in GroupWise registration with sharp mean, which generalizes the conventional unbiased GroupWise registration algorithm.

Each anatomical region may be aligned differently with the mean image, therefore the use of the same weight (generally obtained from the entire subject) for each anatomical region may lead to different amounts of fuzziness across different regions of the mean image.

ISSN (ONLINE):2395-695X
ISSN (PRINT):2395-695X

International Journal of Advanced Research in Basic Engineering Sciences and Technology (IJARBEST)
Vol.2, Issue.12, December 2016

D. Solutions to GroupWise Registration

- It can have a sharp group mean image to allow for better registration with all subjects.
- It is closer to the final population center than any other subjects.

II. EXPERIMENTAL RESULTS

Fig.4 is taken as input image for facial expression recognition. The input image is read by using the function image read and shows the input image by using image show function.

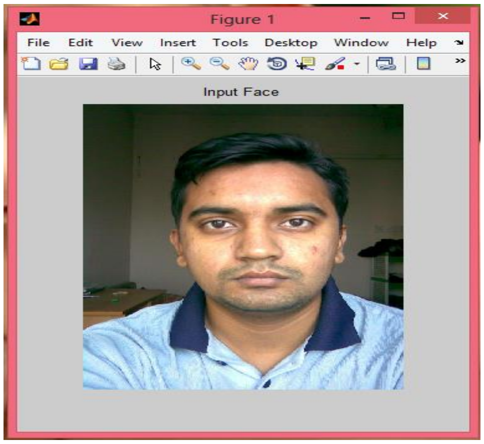

Fig.4 Input image

This Fig.5 shows the similarity measurement between input image and the images in the database.

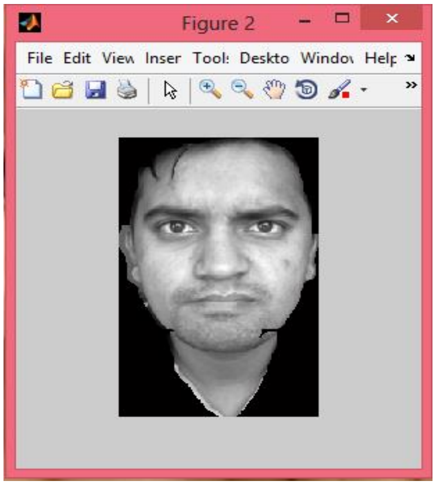

Fig .5 Similarity measurement output

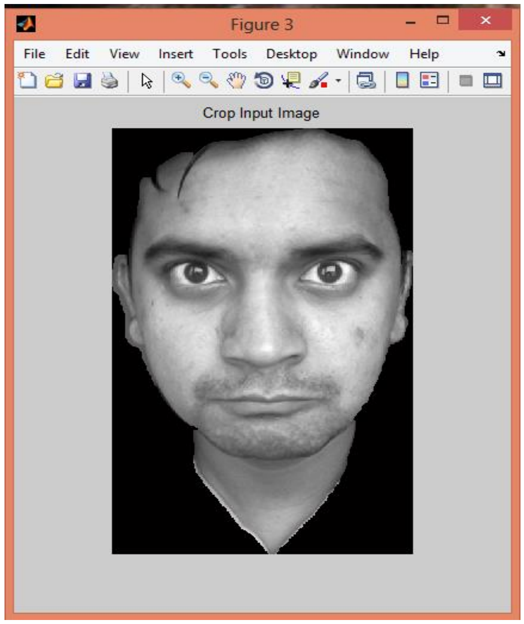

Fig.6 crop input image

This Fig.6 image shows the optimized output image. The optimized output can be obtained by comparing each similarities between the images.

The Fig.7 shows the transformation model output image. This output is produced by using 2D affine geometric transformation.

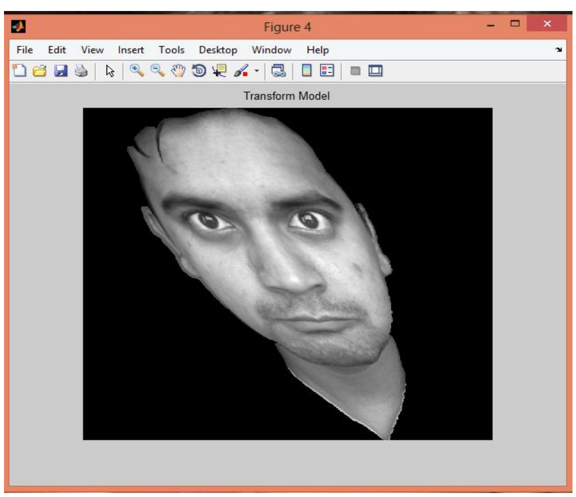

Fig.7 Transformation model output

ISSN (ONLINE):2395-695X
ISSN (PRINT):2395-695X

International Journal of Advanced Research in Basic Engineering Sciences and Technology (IJARBEST)
Vol.2, Issue.12, December 2016

The correct recognition result is shown in Fig.8 Thus, it expresses what type of expression the person delivered.

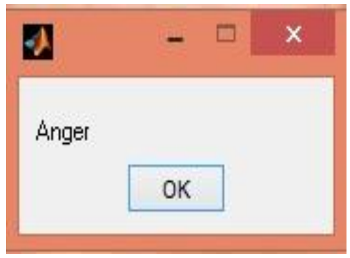

Fig.8 Recognized output

III. CONCLUSION

A new way to tackle the dynamic facial expression recognition problem was proposed. The proposed method mainly consists of two stages, namely atlas construction stage and recognition stage. In the atlas construction stage, longitudinal atlas of different facial expressions is constructed based on sparse representation GroupWise registration. The constructed atlas can capture overall facial appearance movements for a certain expression among the population. In the recognition stage, both the image appearance and temporal information are considered and integrated by diffeomorphic registration and sparse representation. The experimental results of this method consistently achieve higher recognition rates than other compared methods. 3D shape facial expression images can be set and thus it gives more accurate expression. The SIFT or SURF algorithm can be used. This is one possible future direction for this study.

REFERENCES

[1] Bartlett. M, Littlewort. G, Frank. M, Lainscsek. C, Fasel. I, and Movellan. J, (2005) “Recognizing facial expression: machine learning and application to spontaneous behavior,” in IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 568–573.

[2] Cao. X, Wei. Y, Wen. F, and Sun. J, (2012) “Face alignment by explicit shape regression,” in IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 2887–2894.

[3] Cootes. T, Twining. C, Petrovic. V, Babalola. K, and Taylor, C, (2010) “Computing accurate correspondences across groups of images,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, no. 11, pp. 1994–2005.

[4] De la Torre. F and Nguyen. M, (2008) "Parameterized kernel principal component analysis: Theory and applications to supervised and unsupervised image alignment," in IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 1–8.

[5] Edwards. G, Taylor. C, and Cootes. T, (1998) "Interpreting face images using active appearance models," in IEEE FG, pp. 300–305.

[6] Fasel. B and Luettin. J, (2003) "Automatic facial expression analysis: a survey," Pattern Recognition, vol. 36, pp. 259–275.

[7] Kemelmacher-Shlizerman. I, Suwajanakorn. S, and Seitz. S, (2014) "Illumination-aware age progression," in IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 3334–3341.

[8] Peng. Y, Ganesh. A, Wright. J, Xu. W, and Ma. Y, "Rasl: Robust alignment by sparse and low-rank decomposition for linearly correlated images," in IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 763–770.

[9] R.Ravi, Muthukumaran.N," Hardware Implementation of Architecture Techniques for Fast Efficient Lossless Image Compression System" Wireless Personal Communications (Annexure- I), ISSN: 1572-834X, September 2016.

[10] R.Ravi, Raja Ratna, "Scrutiny of unruly and abuse in wireless networks to mitigate physical layer threats using discriminate- based misbehavior prevention" Cluster Computing- The journal of networks software tools and applications (Annexure- I)", Vol. 19, Issue. 1, 2016

[11] R.Ravi, Raja Ratna, "An intelligent approach based on neuro-fuzzy detachment scheme for preventing jamming attack in wireless networks" Journal of Intelligent and Fuzzy Logic (Annexure- I), pp. 801-820, Vol. 28, No. 2, ISSN. 1064-1246, January 2015.

[12] R.Ravi, Muthukumaran.N, N," The Performance Analysis of Fast Efficient Lossless Satellite Image Compression and Decompression for Wavelet Based Algorithm" Wireless Personal Communications (Annexure-I), Page No. 839-859 Volume. 81, No. 2, March 2015.

[13] R.Ravi, Muthukumaran.N, "Hardware Implementation of Architecture Techniques for Fast Efficient loss less Image Compression System" Wireless Personal Communications (Annexure-I), ISSN: 0929-6212, July 2014.