Recognition of Dynamic Facial Expression Using Atlas Construction and Sparsity Orthogonal Matching

[1]Ashwini Sugave, [2]Ashwini Gaikwad

Department of Computer Science and Engineering
Marathwada Shikshan Prasark Mandal's Deogiri Institute of Engineering & Management Studies,
Aurangabad Maharashtra state, India 2018-2019

***Abstract*: Human face can exhibits complex and strong changes that are both unpredictable and varying in time Now a days, facial expression recognition has become a emerging and research topic due to advancement in this field. To overcome the disadvantages of conventional group-wise registration methods Spatio-temporal information and SOM algorithm used in recognition stage.This paper presents sparse representation with sparsity estimation and variable step size to improve recognition efficiency and accuracy. We have adopted SOM algorithm for better performance with reference to sparse representation classification algorithm. We also explored Sparsity Orthogonal matching algorithm is used in atlas construction to reach to true sparsity and lead to improved sparse representation as well as to resolve sparse coefficient and larger step size to shorten the time of iterations. The recognition rate is higher than other method shows by experimental result.**

Index Terms: Atlas construction, Diffeomorphic image registration, sparse representation classification, Sparsity Orthogonal Matching Spatial Domain, Temporal Domain.

1 INTRODUCTION

Automatic facial expression recognition has many applications include computer vision , medical image analysis, financial security, law enforcement etc. It is most powerful way that people coordinate conversion and communicate emotions and other mental cues. Facial expression analysis aims to analyze human facial expressions from videos and classify them into correct facial type, such as anger, disgust, happiness, fear, sadness and surprise. It mainly depends on two stages: 1) feature extraction ,these features are often represented in different form such as novel approaches like motion based method, motion based method and geometric features like active appearance model (AAM), 2) classifier design, support vector machine is mostly used.The aim of facial expression recognition is to estimate facial expression type from an image sequence.

It is most powerful way that people coordinate conversion and communicate emotions and other mental cues. This paper works on basic human emotions like angry, fear, happiness, sad, neutral and surprise. To estimate facial expression type from an image sequence is the aim of facial expression recognition. Here, we use spatial domain as well as temporal domain are used to guide recognition of image sequence. In spatial domain, image appearance information used to enhance recognition performance and temporal domain contains evaluation details.

Dynamic facial expression recognition using atlas construction and sparse representation with sparsity estimation and variable step size is done in this paper. Atlas of facial expression are constructed based on group-wise registration and to slove the sparse coefficient for sparse representation theory we use SOM algorithm. Beacause of robustness and saliency features we use sparse representation with sparseness parameter which affects the quality of atlases. Atlases is nothing but the group of images belonging to specific expression type [1]. In recognition stage, expression type is determined by comparing the corresponding query sequence with each atlas sequence, comparison conducted based on image appearance information and temporal evaluation information. In atlas construction stage, diffeomorphic growth model is estimated to get image sequence. Diffeomorphic means the motion is preserved. SRC framework is used to minimization of norm and redundant dictionary construction.

2 LITERATURE SURVEY

2.1 Markov Random Field

The entire system concentrates on group-wise deformable image registration and MRF problem [2]. There are two main states present in this paper. 1) from each pixel position anatomical features are selected from the facial image and they reflect the structural property and structures such as nose, eyes and lips etc.2) to perform group-wise image registration, the deformable model is converted into markov Radom field labeling problem.

2.2 Partical Filtering

The method is used here not only to recognize facial action units but also the temporal model from face image sequence [3]. The algorithm perform partical filtering to extract 15 facial points in an facial image. It also perform automatic segmentation and recognition of temporal segment that takes input from video or images and that convert into facial expression. These both segmentation perform on 27AUs.

2.3 Action Unit

This paper works on automatic detection of AUs and classification of 6 basic facial expressions..It mainly focuses on three stages; Facial tracking and feature extraction, extract dynamic signals to parametric face, Machine learning methods is used for better understanding changes of facial expression. The algorithm perform integration of facial action units AU1 to AU15 and geometric features M1 to M14 to recognize facial expression Dynamic responses of these texture feature gives the specific expression.

2.4 Lipachitz Embedding and Expression Manifold

System flow works in both embedded space and image space [5]. In the image space, track some feature points from input video and also perform manifold Lipschitz. In this paper, similar expressions are combined together in neighborhood on manifold and it became a path of emotions on manifold. Using a mixed model facial features are clustered. For cluster each, some ASM techniques are propagated in low dimensional space due to which avoids incorrect matching.

2.5 Local binary Pattern

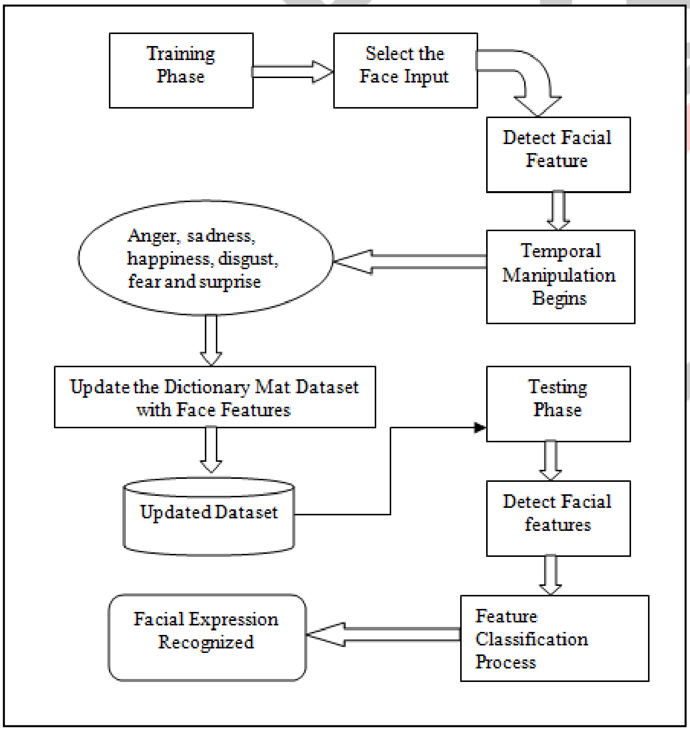
This technique is much effective for because of local binary pattern features.[6] Vector machine classifier is used to improve the boosted LBP feature, which gives best expression recognition and describe appearance informati. A basic LBP operatator select 3×3 neghiourhood of each pixel and 256bin histogram.. Disadvantages of this paper are it works only with static images.and they do not consider temporal information and head pose variation.

2.6 Diffeomorphic Matching

The method works with diffeomorphic matching [7]. The basic step of facial expression recognition is to select landmark points from different images by using some automatic methods. Here, the author use the rigid registration algorithm. By using diffeomorphic matching, the distance is calculated between all landmark sets. Multidimensional scaling is used to recognize the structure of data as well as to find configuration of points.

3 SYSTEM DEVELOPMENT

System architecture 3.1 shows all subprocess that are Feture extraction, testing, classification etc. Here, given a query facial image, estimate the correct facial expression type, such as anger, disgust, happiness, sadness, fear or surprise. The image appearance information together with the expression evolution information can further enhance recognition performance. 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. Finally in recognition stage, expression type is determined by comparing the corresponding query sequence with each atlas sequence.



3.1 Atlas Construction by Groupwise Registration

In input image sequence three phases are there. The phase starts with neutral expression followed by offset and ends with apex. Atlases are constructed for each emotion by using sparse registration method. Image registration means to transfer the set of images to the common i.e. called template space. Input for this stage is video or set of images. From that video it take some image sequence and divide time interval between that set. We formulate the following equation no. 1 to construct atlases by using sparse group-wise registration:

$M_t, \emptyset^i = \arg\min \sum_{t\in T}\sum_{i=1}^{C}\{ (d(\widetilde{M}_t, \widetilde{\emptyset}^i_{(r,t1\to r)}(I_{r,t}))^2 + \gamma_{\emptyset^i} R(\widetilde{\emptyset}^i)\}$ (1)

The algorithm [8] of groupwise registration summarized by greedy iterative algorithm.. M_t is estimated based on the sparse representation of R by minimizing:

$$E(\delta) = \frac{1}{2}\|R\delta - \overrightarrow{m_t}\|_2^2 + \lambda s\|\delta\|_1 \quad (2)$$

where R =[r1,...,rC], ri1(i = 1,...,C) is a column vector corresponding to the vectorization of Ri ,and m is the vectorization of M_t . ‖ · ‖1 is L1 norm and λ sti is the parameter that controls sparseness degree of reconstruction coefficient vector. Therefore, to slove sparse coefficient in sparse solution we proposed SOM algorithm that minimizes the disadvantages of greedy algorithm

3.2 SRC Framework

First Sparse representation based classification (SRC) framework has been proposed by Jhon[9] .In this paper, we use SRC framework based on aspect: construction of redundant dictionary using training sample and sparse coefficient solution.

The basic idea of SRC framework: 1) test sample can be represented by training samples linear combination and we added noise constraint to improve robustness because environmental factors affect on recognition rate.2) minimization of norm by using greedy algorithm that belonging to sparse computation.3) Minimum residual error is calculated by comparing reconstructed sample and test samples of each class (eg. sad, happiness).According to Minimum residual error finalize the category.

The steps of sparse representation classification algorithms are summarize below:

1. Input: training sample $M_t \in R^{m\times n}$, test sample $y \in R^{m\times n}$,constant δ_s and step coefficient ß
2. Each column of the *M* and *y* are normalized by the *l*
3. Solve the l_0 norm of the minimization:

$\hat{x} = arg \min_x \|x\|_0$ s.t. $\|y - Mx\| \le \varepsilon$

Or translate to solve the $l_1$norm of theminimization:

$\hat{x} = arg \min_x \|x\|_1$ s.t. $\|y - Mx\| \le \varepsilon$

4. Compute the residuals for the each class $j \in [1, C]$.

$.y_j(y) = \|y - M\delta_j(\hat{x})\|_2$

5. *Output:* $I(y)$ =arg $min_j r_j(y)$

3.3 Recognition of Query Sequence

For recognition of facial expression atlas sequence is used. In this stage, a query image sequence is used as input for evaluation of emotion. Image appearance information, evaluation information and temporal information are used to perform recognition. Recognition process is described by the following eq no. 2:

3.4 SOM Algorithm

Sparsity Orthogonal matching algorithm works on two important parameters: sparsity estimation and variable step size[10]. Aim of that algorithm is to overcome the disadvantages of greedy algorithm in sparse solution.SOM algorithm not only improve recognition rate but also increase recognition accuracy and efficiency as compare to other algorithm. The method of step size approximation is divided into many stages in iteration. The algorithm will decrease the number of subsequent iterations and larger stepsize is adpoted to reduce the time of iteartion at the begining. And then gradually reduce the step size and improve the accuracy. Each stage increases the fixed step size to meet the requirement. The choice of step size will also affect the performance of the algorithm.

In SOM algorithm, there is no need to set the input to true Sparsity and the initial value of the sparsity is estimated by matching test. Then, the atom is selected from the projection set, and the index set and the support set are updated. The original signal is estimated by least square method, and update residuals. Finally approach to true sparsity and lead to better sparse representation.Steps of algorithm summarized below:

SOM Algorithm

1. Input: training sample A $\in R^{m\times n}$, test sample y $\in R^{m\times n}$,constant δ_s and step coefficient ß
2. Initialization : sparity $S_0 = 1$, step size step=m / log_2n, residual error $\gamma_0 = y$, index set $\wedge_0 = \emptyset$ and the support set $A_0 = \emptyset$.
3. Compute the projection set u = $\{u_j | u_j = |\langle \gamma_0, A_l\rangle|$,l=1,2….,n}, and select the S_0 larger value in u.The index value and the support set are stored $\wedge_0$ *and* A_0 seperately.
4. If $\|A_0^T y\|_2 < \frac{1+\delta_s}{\sqrt{1+\delta_s}}\|y\|_2$, the S0 = S0 + 1 is confirmed. And then turn on step 3. Or compute the residual error γ0 = y – A0 A0+y, and set L= S0, stage = 1 and the times of iteration t = 1. Turn on step 5.

5. Compute the projective set $u = \{u_j \mid u_j = |\langle \gamma_{t-1}, A_l \rangle|, l = 1,2,\ldots,n\}$, and select the L larger value in u. The corresponding index values form a set J0.

6. Update index sets and support sets: $\wedge_t = \wedge_{t-1} \cup J_0, A_t = A_{t-1} \cup A_l (l \in J_0)$

7. The equation y=$A_t x_t$ is computed and least square solution is given:

$$\hat{x}_t = arg \min_{x_t} \|y - A_t x_t\|_2$$

8. Select the largest L of the absolute value marked as $\hat{x}_t L$, and then the L column in corresponding to A_t is markes as $\wedge_t L$

9. Update the residual error $\gamma_t = y - A_t L \hat{x}_t L$

10. If residual error $\|\gamma_t\|_2 < \varepsilon$, turn on the step 11. If $\|\gamma_{new}\|_2 \geq \|\gamma_{t-1}\|_2$, stage = stage+1, step=$\lceil \beta \times step \rceil$ $L = L + step$ and $t = t + 1$,turn on the step 5. If the two conditions is not satisfied,$\wedge_t = \wedge_t L$, t = t+1,turn on the step 5.

11. Output: sparse coefficient$\hat{x}$, and the nonzero items in corresponding to the index set $\wedge_t$are the final iteration$\hat{x}_t$.

4 PERFORMANCE ANALYSIS

4.1 Figures and Tables

To verify the performance of proposed method, face images are collected as training data. We peform facial expression recognition expriement on test image. We evaluated that, recognition rate has been increased as well as improve performance accuracy and efficiency as compared to other facial expression methods.

TABLE 1
AVERAGE RECOGNITION RATE OF ALL EMOTIONS

Methods	Recognition accuracies (%)
Conventional	94.13
Sparse + Appearance info	95.40
Sparse+ Appearance+ temporal	96.48
Sparsity + Orthogonal matching	96.78

Figure 2 shows the average recognition rates of different expressions obtained with and without using sparse representation in atlas construction.

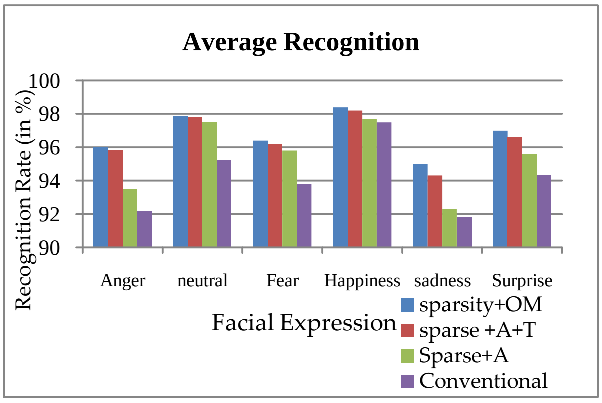


Fig. 2. Average recognition rate of different facial expression.

As show in figure 3, our method compare and some dynamic facial expression methods. Although experimental protocols of compared methods are not exactly the same Recognition rate is higher than other methods.

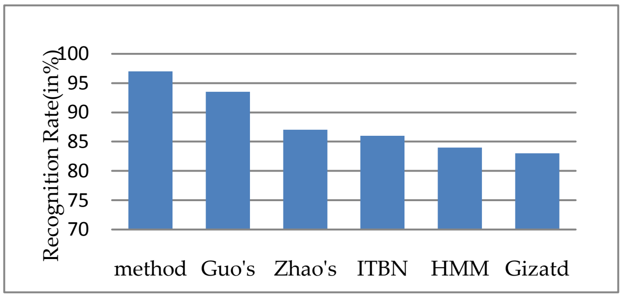


Fig. 3. Recognition rate with different methods..

Figure 4 shows that average recognition accuracies obtained by the proposed method with different number of time points N to construct atlas sequences. The larger value of N more accurately atlas sequence can present facial feature and vice versa. For ex. N=4 the recognition rate is 93.6 , atlas sequence cannot describe evaluation process sufficiently. When N=14 satisfactory recognition rate obtained.

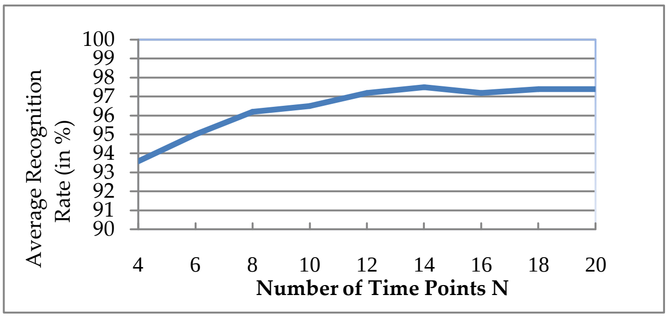


Fig. 4. Recognition rate with no.of time points.

TABLE 2

CONFUSION MATRIX OF PROPOSED METHOD

	Angry(%)	Neutral(%)	Fear (%)	happy(%)	Sad (%)	Surprise(%)
Angry	96	0	0	0	2.0	0
Neutral	0	97.9	0	0.6	1.5	0
Fear	0	0	96.4	0	0.5	1.5
happy	0	0.5	0	98.4	0	0
Sad	2.4	0.8	0.5	0	95	0.8
Surprise	0	0	2.5	0.4	0	97

Following figure5 shows that recognition accuracies using different values of β. For image appearance and temporal evaluation information β controls the weight in recognition step. When β=0.5, recognition accuracy increases. Both are important temporal domain as well as spatial domain for recognition.

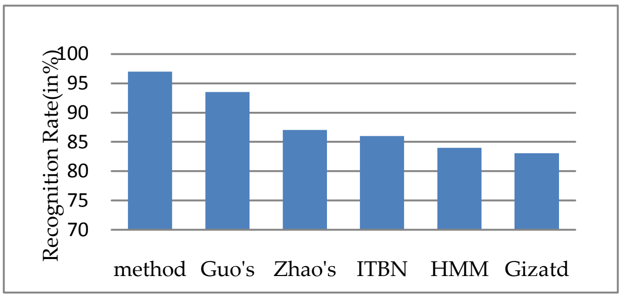


Fig. 5. Average recognition accuracies obtained by the proposed method with different values of ß .

As shown in figure 6, We take 50 trainig samples from database .It is important to chose correct value of sparsity because it reduces the time of iteration. For example, δ_s=0.9 then the sparsity estimation value is maintained at 1 then time of iterations are enlarged. Best value of δ_s is 0.2, it reduces the time of iteartions.

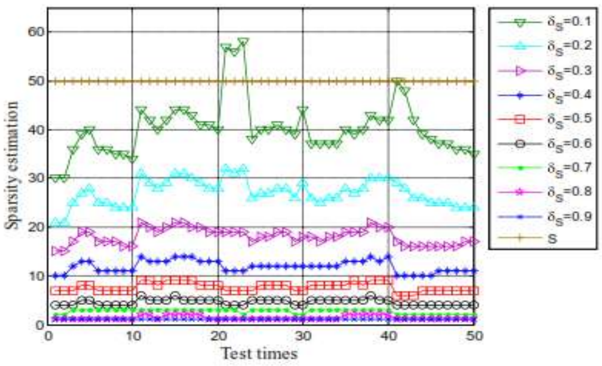


Fig. 6. Sparsity Estimatione.

Conclusion

Based on atlas construction and sparse representation, an improved sparsity orthogonal matching (SOM) algorithm is proposed to better recognition as well as to improve accuracy and efficiency. Sparse representation classification algorithm is used in atlas construction stage to improve the quality of atlases and to solve the sparse coefficient. Atlases can capture overall facial features for certain expression and that further used for fast recognition. In recognition stage, SOM algorithm is proposed to solve problem of greedy algorithm and to decrease the number of iteration for expression recognition. SOM algorithm depends on sparsity estimation to estimate initial value of sparsity by matching test and variable step size is added to approaches the true sparsity. The experimental result shows that gives better performance than other facial expression technique and compared many facial expression methods to get ride over which is more efficient technique .

References

[1] Yimo Guo,Guoying Zhao, Senior Member, IEEE, and Matti Pietikäinen, Fellow, IEEE “Dynamic Facial Expression Recogniton with Atlas Construction and Sparse Representation” IEEE Transaction on Image Processing, vol. 25, no. 5, May 2016

[2] Shu Liao, Dinggang Shen, and Albert C.S. Chung, “A Markov Random Field Groupwise Registration Framework for Face Recognition” IEEE Transactions on Pattern Analysis and Machine Intellence, vol. 36, no. 4, April 2014 .

[3] Maja Pantic, Member, IEEE, and Ioannis Patras, Member, IEEE, “Dynamics of Facial Expression:Recognition of Facial Actions and Their Temporal Segments From Face Profile Image Sequences” IEEE Transactions on Systemvol. 36, no. 2, April 2006.

[4] Hui Fang,Neil Mac Parthalian Andrew J. Andrew J. Aubrey, Gray K.L. Tam, Rita Borgo Paul L. “ Facial Expression Recognition in Dynamic Sequences: an Integrated Approach” Cronfa - Swansea University Open Access vol. 47, no.3, 2014.

[5] Ya Chang, Changbo Hu , Rogerio Feris , Matthew Turk, “Manifold based analysis of facial expression” Science Direct Computer Science Department, University of California, Santa Barbara, CA 93106, USA, vol. 21, no. 6, 2006.

[6] Caifeng Shan, Shaogang Gong, Peter W. McOwan “Facial expression recognition based on Local Binary Patterns: A comprehensive study” Philips Research AE Eindhoven, The Netherlands Dept of Computer Science, Queen Mary, London E1, UK ,vol. 27, 2009.

[7] Siamak Yousefi, Minh, Phuoc Nguyen, Nasser Kehtarnavaz “ Facial Expression Recognition Based on Diffeomorphic Matching” Proceedings of 2010 IEEE 17th International Conference on Image Processing. University of Texas at Dallas, sep. 2010.

[8] Y. Guo, G. Zhao, and M. Pietikäinen, “Dynamic facial expression recognition using longitudinal facial expression atlases,” in Proc. 12th Eur. Conf. Comput. Vis., 2012.

[9] Wright, J., Yang, A.Y., Ganesh, A., et al.: Robust face recognition via sparse representation [J]. IEEE Trans. Pattern Anal. Mach. Intell. **31**(2), 210–227 (2009).

[10] Cheng, Y., Feng, W., Feng, H., et al.: A sparsity adaptive subspace pursuitalgorithmforcompressive sampling[J].ActaElectron.Sin. 38(8), 1914–1917 (2010)