

Dynamic Facial Expression Recognition with Extreme Sparse Learning

R. Gowsalya¹, C. Rajeshkannan²

M.E. Computer Science and Engineering, Mount Zion College of Engineering and Technology, Lena Vilakku,
Thirumayam Tk, Pudukkottai, Tamil Nadu 622507, India.

Assistant Professor, Department of Computer Science and Engineering, Mount Zion College of Engineering and
Technology, Lena Vilakku, Thirumayam Tk, Pudukkottai, Tamil Nadu 622507, India.

ABSTRACT: Recollection of natural emotions from make human faces is an attractive topic with a broad range of market potential applications like human-computer interaction, automated tutoring systems, image and broadcast retrieval, smart atmosphere, and chauffeur warning systems. Traditionally, facial emotion recognition systems have been evaluated on laboratory controlled data, which is not representative of the environment faced in real-world applications. A new dynamic facial expression recognition method is proposed. Dynamic facial expression recognition is formulated as a longitudinal group wise registration problem. The main contributions of this method lie in the following aspects: [a] subject-specific facial feature movements of different expressions are described by a diffeomorphic growth model; [b] salient longitudinal facial expression atlas is built for each expression by a sparse groupwise image registration method, which can describe the whole face feature changes among the whole population and can suppress the bias due to large inter-subject facial variations; [c] both the profile appearance information in spatial domain and topological evolution data in secular domain are used to guide recognition by a sparse representation method.

KEYWORDS: Dynamic Expression Identification, Facial Emotion Notification, Sparse Expression Representation, Feature Analysis, Facial Features Identification.

I. INTRODUCTION

Programmed Facial Expression Identification has fundamental certifiable applications. Its applications incorporate, yet are not constrained to, Human PC communication, brain science and broadcast communications. It remains a testing issue and dynamic research point in PC vision, and numerous novel techniques have been proposed to handle the programmed outward appearance acknowledgment issue.

Concentrated reviews have been done on Automatic Facial Expression Recognition [AFER] issue in static pictures amid the most recent decade: Given a question facial picture, assess the right outward appearance sort, for example, outrage, appall, joy, pity, dread or astonishment. It primarily comprises of two stages: include extraction and classifier plan. For highlight extraction, Gabor wavelet, Local Binary Pattern [LBP], and geometric components, for example, dynamic appearance demonstrate are in like manner utilize. For classifier, bolster vector machine is every now and again utilized. Joint arrangement of facial pictures under unconstrained condition has additionally turned into a dynamic research theme in AFER.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 2, February 2017

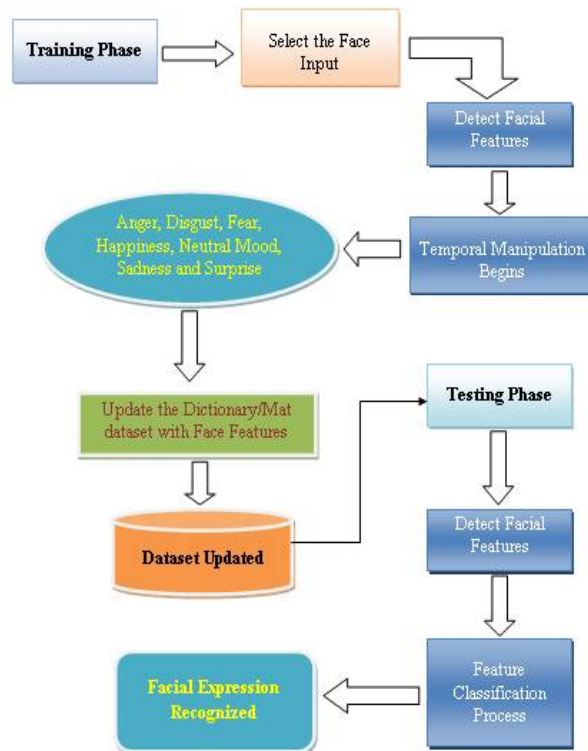


Fig.1 System Architecture Design

Lately, dynamic outward appearance acknowledgment has turned into another exploration subject and gets increasingly consideration. Not quite the same as the acknowledgment issue in static pictures, the point of element outward appearance acknowledgment is to gauge outward appearance sort from a picture succession caught amid physical outward appearance procedure of a subject. The outward appearance picture arrangement contains not just picture appearance data in the spatial area, additionally development points of interest in the fleeting space. The picture appearance data together with the expression development data can additionally improve acknowledgment execution. Despite the fact that the dynamic data gave is helpful, there are difficulties with respect to how to catch this data dependably and powerfully.

For example, an outward appearance arrangement regularly constitutes of at least one onset, summit and counterbalance stages. With a specific end goal to catch fleeting data and make transient data of preparing and inquiry successions tantamount, correspondences between various worldly stages should be built up. As facial activities after some time are distinctive crosswise over subjects, it remains an open issue how a typical fleeting element for every expression among the populace can be adequately encoded while stifling subject-particular facial shape varieties.

II. FEATURE EXTRACTION

After frame conversion new face coordinate system on the image Plane is developed. Using Delaunay triangulation facial features like nose tip, left eye, right eye and mouth are detected. While the nose tip is considered as the origin of the face coordinate system, the reference vector connecting the nose tip to the midpoint between the centers of the two eyes is considered as the positive y-axis.

III. OPTICAL FLOW

In order to compute the dynamic features, we start by computing the Optical Flow of a given video based on both brightness and gradient constancy assumption. The local motion of facial components resulting from the act of expressing an emotion, global motion of the head is subtracted from the flow vector.

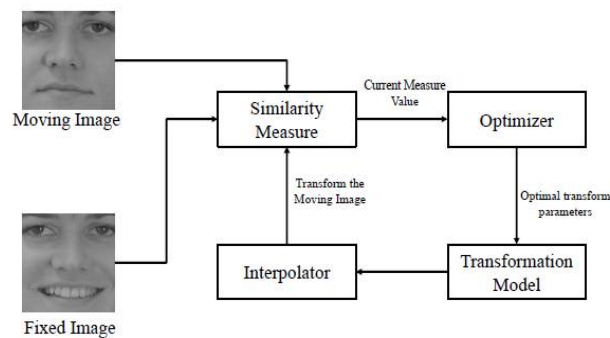


Fig.2. A typical flow chart for pair-wise image registration: the moving image is transformed to fixed image space.

Spatio-Temporal Features

Four pose-level features are proposed for encoding the motion information of facial components. The first feature is the divergence of the flow field, which measures the amount of local expansion or contraction of the facial muscles. The second feature captures the local spin around the axis perpendicular to the OF plane and is referred to as Curl. It is useful to measure the dynamics of the local circular motion of the facial components.

The third feature is the scalar projection. This Project feature captures the amount of expansion or contraction of each point with respect to the nose point. For example, the “convivial” and “sad” emotions can be distinguished by this feature clearly. The fourth feature is the Rot feature. The Rot feature measures the amount of clockwise or anti-clockwise rotation of each facial point operation with respect to the position vector.

This feature helps in distinguishing between “happy” and “anger” emotions. The sign and magnitude of Rot feature are different for a sample lip point (the magnitude is exaggerated for better illustration) depending on the facial emotion. A spatio-temporal descriptor is obtained by concatenating the spatio-temporal features extracted at each local area in the video.

Extreme Sparse Learning

Sparse representation is that the images can be efficiently approximated by linear combination of a few elements called atoms. The dictionary can be obtained by either applying predefined transforms to the data or directly learning from training data. Extreme Learning Machine (ELM) is the classification method used. Supervised sparse coding is performed on the dictionary D and the classification error based on the given ELM output weight vector β . Algorithm used in Supervised sparse coding is Class specific Matching pursuit. Supervised sparse coding, estimates a sparse code matrix X that simultaneously minimizes the reconstruction error based on the given dictionary D and the classification error based on the given ELM output weight vector β .

Class Specific Matching Pursuit (CSMP)

The basic idea underlying the matching pursuit process is to sequentially find atoms in the dictionary and the corresponding sparse coefficients that minimize the objective function ϵ t..An updated sparse code matrix is obtained.

IV. EXISTING SYSTEM

The existing method uses a longitudinal atlas construction and diffeomorphic image registration problem. The system 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.

Disadvantages

✚ It is not robust enough to overcome challenges of strong illumination changes, occlusion and movements of head when using Atlas construction.



Fig.3. Illustration of the whole facial expression process.

IV. PROPOSED METHODOLOGY

To identify the emotions in the presence of self-occlusion and illumination variations, we combine the idea of sparse representation with Extreme Learning Machine (ELM) to learn a powerful classifier that can handle noisy and imperfect data. Sparse representation is a powerful tool for reconstruction, representation, and compression of high-dimensional noisy data (such as images/videos and features derived from them) due to its ability to uncover major information about signals from the base elements or dictionary atoms. While the sparse representation approach has the ability to enhance noisy data using a dictionary learned from clean information, it is not sufficient because our end goal is to correctly recognize the facial emotion.

Algorithm 1: Estimate Sparse Learning

Input: Images $-i(t_0:t)(I_i(t_0))$ of each subject i ($i = 1; \dots; C$) that are interpolated at time point t with the growth model $-i$.

Output: Atlas M_t constructed at time point t .

1. Initialize $M_t = \frac{1}{C} \sum_{i=1}^C I_i(t_0)$.

2. Initialize $\hat{I}_i = -i(t_0:t) \wedge (I_i(t_0))$.

3. FOR $i = 1$ to C

Perform diffeomorphic image registration: register I_i to M_t to minimize the image metric defined in Equation 4 between \hat{I}_i and M_t . Denote the registered images as R_i .

END FOR

4. Update $M_t = \frac{1}{C} \sum_{i=1}^C R_i$.

5. Repeat Steps 3 and 4 until M_t converges.

6. Return M_t .

In a sparse-representation-based classification task, the desired dictionary should have both representational ability and discriminative power. Since separating the classifier training from dictionary learning may cause the learned dictionary to be sub-optimal for the classification task, we propose to jointly learn a dictionary (which may not be necessarily

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 2, February 2017

over-completed) and a classification model. To the best of our knowledge, this is the first attempt in the literature to simultaneously learn the sparse representation of the signal and train a non-linear classifier based on sparse codes.

Advantages

- ✚ A pose-balanced OF-based spatio-temporal descriptor, which is able to robustly represent facial emotions.
- ✚ Sparse Representation based Classification can improve the discriminative ability of the method and lead to better classification results.

TABLE I
CONFUSION MATRIX OF THE PROPOSED METHOD FOR CK+ DATABASE.

	Anger (%)	Contempt (%)	Disgust (%)	Fear (%)	Happiness (%)	Sadness (%)	Surprise (%)
Anger	96.1	0	0	1.5	0	2.4	0
Contempt	0	91.8	7.3	0	0	0.9	0
Disgust	0	0.8	98.8	0	0.4	0	0
Fear	0	0	0	95.5	3.4	1.1	0
Happiness	0	0	0	0.8	99.2	0	0
Sadness	2.2	0	0	1.0	0	96.8	0
Surprise	0	0	0	0	0	0.7	99.3

V. LITERATURE SURVEY

In this paper title is "A Survey on Visual Surveillance of Object Motion and Behaviors", the author describe the main concept behind this system, which is based on visual surveillance. The entire system concentrates highly in the area to analyze possible research directions, e.g., occlusion handling, a combination of two and three-dimensional tracking, a combination of motion analysis and biometrics, anomaly detection and behavior prognosis, content-based retrieval of surveillance videos, behavior understanding and natural language description, fusion of information from multiple sensors, and remote surveillance.

In this paper title is "A Markov Random Field Groupwise Registration Framework for Face Recognition", the authors proposed a new framework for tackling face recognition problem.

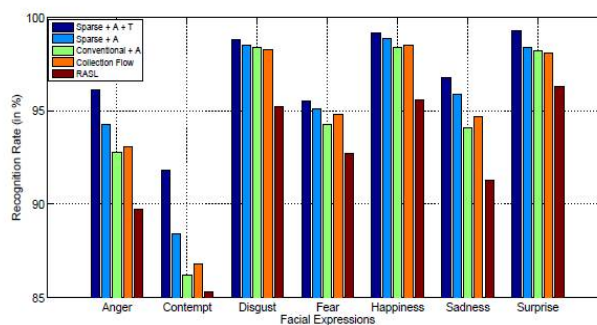


Fig.4. The average recognition rates of seven different facial expressions on the CK+ database by using different schemes. "RASL" is the standard RASL algorithm with affine transformation model, "Collection Flow" is the collection flow algorithm, "Conventional" is the conventional groupwise registration method used to construct atlas, and "Sparse" is the proposed sparse representation scheme used to construct atlas. "A" denotes image appearance information, and "T" denotes temporal evolution information.

The face recognition problem is draw up as groupwise deformable image registration and feature matching problem. The main contributions of the proposed method lie in the following aspects: [a] each pixel in a facial image is represented by an anatomical signature obtained from its corresponding most salient scale local region determined by the Survival Exponential Entropy [SEE] information theoretic measure. [b] Based on the anatomical signature calculated from each pixel, a novel Markov random field based groupwise registration framework is proposed to formulate the face recognition difficult situation as a feature guided deformable image registration problem. [c] The

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 2, February 2017

proposed method has been extensively evaluated on four publicly available databases: FERET, CAS-PEAL-R1, FRGC ver 2.0, and the LFW.

In this paper title is "Facial expression recognition: A survey", the authors proposed Automatic facial expression recognition system, which has many applications including, but not limited to, human attitude understanding, detection of mental disorders, and synthetic human expressions. Two popular methods utilized mostly in the literature for the automatic FER systems are based on geometry and get up. Even though there is lots of research using static images, the research is still going on for the development of new methods which would be quiet easy in computation and would have less memory usage as compared to previous process. This paper presents a quick survey of facial expression recognition. A comparative study is also carried out using various feature extraction techniques on JAFFE dataset.

VI. EXPERIMENTAL RESULTS

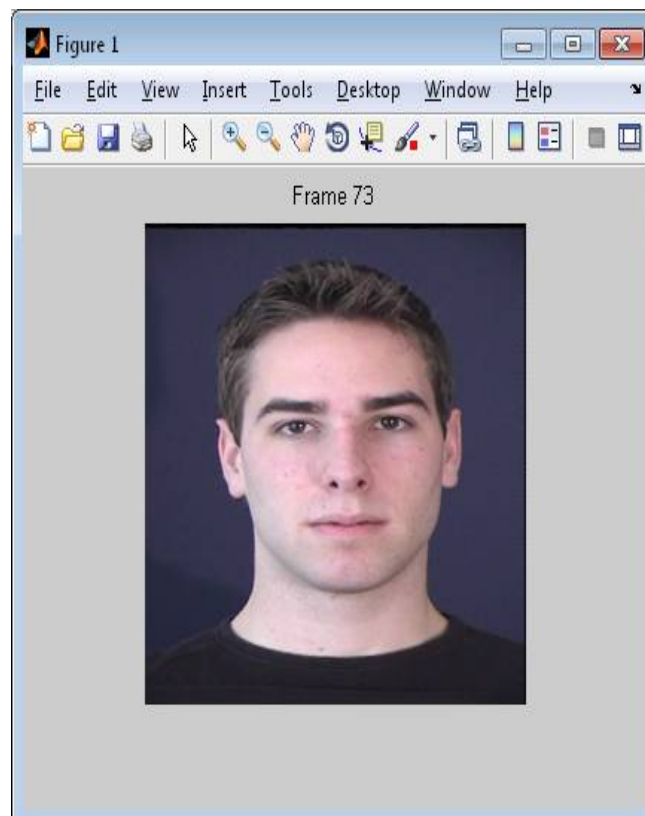


Fig.5 Sample Input Frame

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 2, February 2017

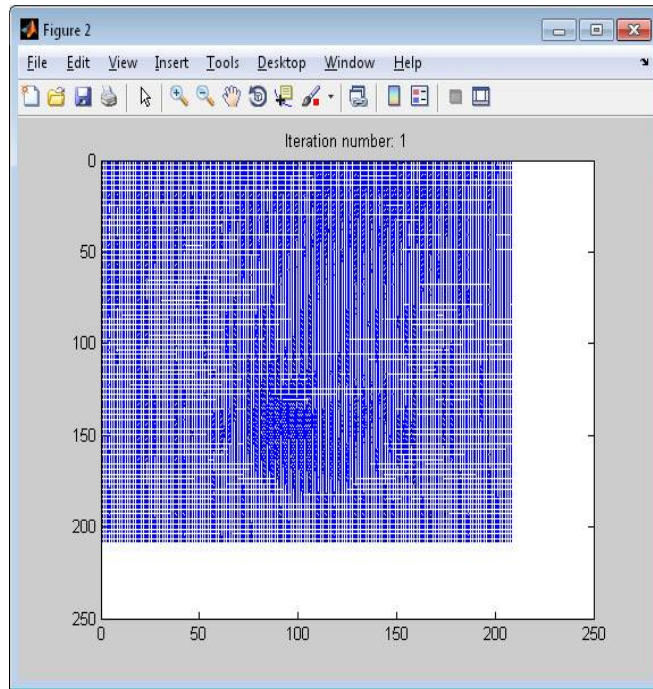


Fig.6 Iteration Process

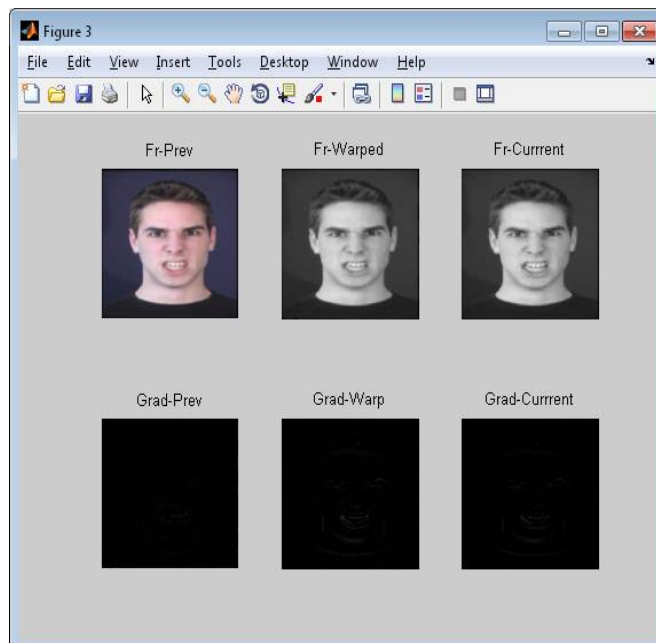


Fig.7 Analyzing Input Frames

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 2, February 2017

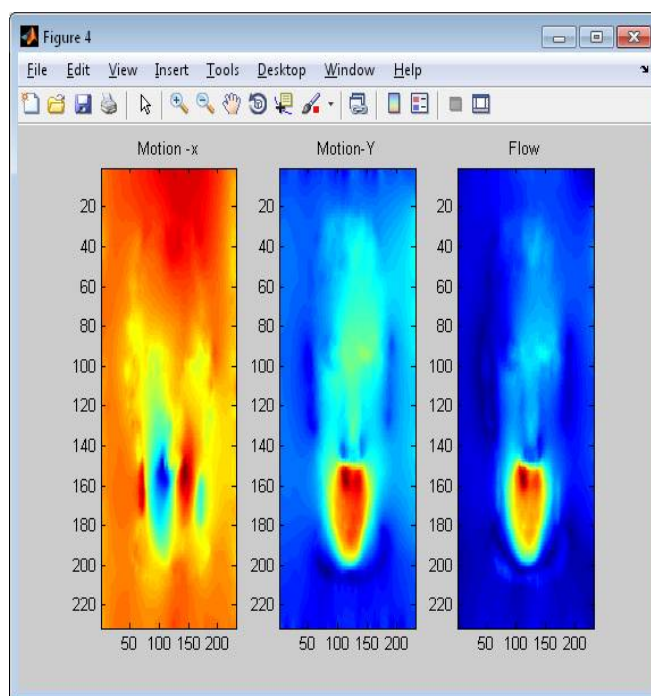


Fig.8 Estimating Motions

VII. CONCLUSION AND FUTURE SCOPE

In this system, we propose a new way to tackle the dynamic facial expression recognition problem. It is formulated as a longitudinal atlas construction and diffeomorphic image registration problem. Our 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. Our method has been extensively evaluated on five dynamic facial expression recognition databases. The experimental results show that this method consistently achieves higher recognition rates than other compared methods. One limitation of the proposed method is that it is still not robust enough to overcome challenges of strong illumination changes. The main reason is that the LDDMM registration algorithm used in this paper may not compensate strong illumination changes. One possible solution is to use complex image matching metrics in the LDDMM framework, such as localized correlation coefficient and localized mutual information which have some degrees of robustness against illumination changes. This is one possible future direction for this study.

REFERENCES

- [1] B. Fasel and J. Luetttin, "Automatic facial expression analysis: a survey," *Pattern Recognition*, vol. 36, pp. 259–275, 2003.
- [2] Z. Zeng, M. Pantic, G. Roisman, and T. Huang, "A survey of affect recognition methods: Audio, visual, and spontaneous expressions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, pp. 39–58, 2009.
- [3] M. Bartlett, G. Littlewort, M. Frank, C. Lainscsek, I. Fasel, and J. Movellan, "Recognizing facial expression: machine learning and application to spontaneous behavior," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2005, pp. 568–573.
- [4] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 971–987, 2002.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 2, February 2017

- [5] S. Lucey, A. Ashraf, and J. Cohn, "Investigating spontaneous facial action recognition through aam representations of the face," in Face Recognition Book, 2007, pp. 275–286.
- [6] E. Learned-Miller, "Data driven image models through continuous joint alignment," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, no. 2, pp. 236–250, 2006.
- [7] G. Zhao and M. Pietikainen, "Dynamic texture recognition using local binary patterns with an application to facial expressions," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 29, pp. 915–928, 2007.
- [8] L. Cohen, N. Sebe, A. Garg, L. Chen, and T. Huang, "Facial expression recognition from video sequences: temporal and static modeling," Computer Vision and Image Understanding, vol. 91, pp. 160–187, 2003.
- [9] Y. Zhang and Q. Ji, "Active and dynamic information fusion for facial expression understanding from image sequences," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, pp. 699–714, 2005.
- [10] S. Koelstra, M. Pantic, and I. Patras, "A dynamic texture-based approach to recognition of facial actions and their temporal models," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, pp. 1940–1954, 2010.
- [11] A. Ramirez and O. Chae, "Spatiotemporal directional number transitional graph for dynamic texture recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence, p. In Press, 2015.
- [12] H. Fang, N. Parthala, A. Aubrey, G. Tam, R. Borgo, P. Rosin, P. Grant, D. Marshall, and M. Chen, "Facial expression recognition in dynamic sequences: An integrated approach," Pattern Recognition, vol. 47, no. 3, pp. 1271–1281, 2014.
- [13] L. Zhong, Q. Liu, P. Yang, B. Liu, J. Huang, and D. Metaxas, "Learning active facial patches for expression analysis," in IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2012, pp. 2562–2569.
- [14] Z. Wang, S. Wang, and Q. Ji, "Capturing complex spatio-temporal relations among facial muscles for facial expression recognition," in IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2013, pp. 3422–3429.
- [15] M. Beg, M. Miller, A. Trouve, and L. Younes, "Computing large deformation metric mappings via geodesic flows of diffeomorphisms," International Journal of Computer Vision, vol. 61, pp. 139–157, 2005.
- [16] S. Yousefi, P. Minh, N. Kehtarnavaz, and C. Yan, "Facial expression recognition based on diffeomorphic matching," in International Conference on Image Processing, 2010, pp. 4549–4552.
- [17] Y. Guo, G. Zhao, and M. Pietikainen, "Dynamic facial expression recognition using longitudinal facial expression atlases," in European Conference on Computer Vision, 2012, pp. 631–644.
- [18] Y. Chang, C. Hu, R. Feris, and M. Turk, "Manifold based analysis of facial expression," Image and Vision Computing, vol. 24, no. 6, pp. 605–614, 2006.
- [19] M. Pantic and I. Patras, "Dynamics of facial expression: Recognition of facial actions and their temporal segments from face profile image sequences," IEEE Transactions on Systems, Man, and Cybernetics Part B, vol. 36, no. 2, pp. 433–449, 2006.
- [20] P. Yang, Q. Liu, X. Cui, and D. Metaxas, "Facial expression recognition using encoded dynamic features," in IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2008, pp. 1–8.
- [21] M. Yeasin, B. Bullot, and R. Sharma, "From facial expression to level of interests: A spatio-temporal approach," in IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004, pp. 922–927.
- [22] G. Edwards, C. Taylor, and T. Cootes, "Interpreting face images using active appearance models," in IEEE FG, 1998, pp. 300–305.
- [23] J. Saragih, S. Lucey, and J. Cohn, "Deformable model fitting by regularized landmark mean-shift," International Journal of Computer Vision, vol. 91, no. 2, pp. 200–215, 2011.
- [24] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005, pp. 886–893.
- [25] G. David, "Distinctive image features from scale-invariant keypoints," International Journal of Computer Vision, vol. 60, no. 2, pp. 91–110, 2004.

BIOGRAPHY



Ms. Gowsalya. R. Studying M.E. Computer Science and Engineering in Mount Zion College of Engineering and Technology, which is located in Lena Vilakku, Pudukkottai, TamilNadu, India.