

The 3D Emotion Recognition Using SVM and HoG Features

Dayanand G. Savakar

*Department of Computer Science
Dr. P. G. Halakatti Post Graduate Centre
“Vachana Sangama”, Rani Channamma University
Toravi, Vijayapur 586108, Karnataka, India
dgsavakar@gmail.com*

Ravi Hosur*

*Department of Computer Science & Engineering
BLDEA's V. P. Dr. P. G.
Halakatti College of Engineering & Technology
Vijayapur 586103, Karnataka, India
mca.hosur@bldeacet.ac.in*

Received 11 March 2019

Accepted 8 July 2019

Published 31 July 2020

Emotion recognition is becoming commercially popular due to the major role of analytics in various aspects of marketing and strategy management. Several papers have been proposed in emotion recognition. They are mainly classified in the past under 2D and 3D emotion recognition, out of which 2D emotion recognition has been more popular. Various aspects like facial posture, light intensity variations and sensor-independent recognition have been studied by different authors in the past. However, in reality, 3D emotion recognition has been found to be more efficient which has a broader area of use. In this paper, a 3D tracking plane with 2D feature points has enabled us to recognize emotions by statistical voting method from all planes having over threshold number of points in their respective contour area. The proposed technique's results are comparable to existing methods in terms of time, space complexity and accuracy improvement.

Keywords: Emotion; features; classification; 3D face; expression recognition.

1. Introduction

Emotion analysis is becoming an integral part of video analytics system day by day. Last decade was dedicated to growth in image sharing, while the current decade belongs to videos. The amounts of videos that are being shared in social media have increased by several folds. Mental states of a human being are represented by emotions which are related to thought process, feelings, pleasure, sadness, etc.

*Corresponding author.



Fig. 1. Classification of emotions based on likelihoods.

Though there is no hard and fast definition of emotion, it mainly reflects different states of human mind. Emotions of a human being are identified by the expressions on the face of a person which are actually movements or locations of the muscles under the skin of the human face. In other words, facial expressions of a person express his/her mental states, i.e. emotions. Figure 1 presents how the facial expressions are related to emotions.

In the images shown in Fig. 1, various facial expressions reflecting the different mental states are identified using a common facial emotion recognition system. These identified emotions depending on facial expressions are categorized using the features of human face. Usually, the emotion of anger is defined when the upper and lower lips are tightly pushed against each other with eyebrows being lowered.

As an example, in the process of identification, if the smile is partial to predict where one side of the mouth is raised a bit, then that facial expression denotes the emotion of contempt. The emotion disgust is denoted by the face expressions that enable a person to raise the upper lip and cheeks. The emotion of fear makes a person to raise the eyebrows a bit and open the mouth marginally. If a person raises the corners of the mouth upwards denoting 100% pure smile, then this facial expression denotes the emotion of joy. On the other hand, dropped jaws and cheeks and curved brows with a faint smile are recognized as the emotion of sadness. If the brows are arched and the eyes and mouth are widely opened, then that emotion is surprise.

The derived emotions using facial expressions have been listed in Table 1. These emotions have been categorized depending on the increasing and decreasing likelihoods of the features of the human face projected on a 3D plane.

There are many application areas for facial expression analysis. Computer-assisted analysis of facial expression can help in the process of identification, verification and authentication. Automatic feedback can be acquired from the customers on reacting to a specific item by examining their facial expressions. As 2D emotion recognition suffers from various drawbacks due to illumination change and head movements, 3D emotion recognition is being used widely which is more effective and

Table 1. Classification of emotions based on likelihoods.

Emotion	Increasing likelihood	Decreasing likelihood
Joy	Smile	Forehead enhancement, forehead furrow
Anger	Forehead furrow, lid tightening, eye widening, chin enhancement, mouth opening, lip suction	Inner brow raising, forehead boost, smile
Disgust	Nose wrinkling, upper lips enhancement	Lip suction, smile
Surprise	Inner forehead enhancement, brow increment, eye widening, jaw dropping	Brow furrow
Fear	Inner brow increment, brow furrow, eye widening, lip stretching	Brow improvement, lip corner depression, jaw dropping, smile
Sadness	Internal forehead enhancement, forehead furrow, lip corner depression	Brow improvement, lip corner depression, jaw dropping, smile
Contempt	Brow furrow, smirking	Smirking

overcomes the drawbacks of 2D emotion recognition system. This research work uses a 3D facial model, further extracts the HoG features from a 3D facial model and trains a convolutional neural network with the normalized facial features

2. Literature Review

Although most of the research works studied particularly one or two modes, but frequently, more than two modes are employed for expressing emotions.¹ The results are related to facial expressions, and the attributes are obtained from 2D, 3D and IR (infrared) sensors.

Classification of data for vision-related functions is established² to achieve enhancement of feature space. The task is carried using feature learning with deep model and the concurrent transfer learning and feature learning using constrained deep transfer feature learning methods.

The solution³ covers the finest areas of the human faces that can be applied for evaluation. It is conveyed that the method presented fairly outruns advanced HR evaluation processes in natural scenarios.

In another paper,⁴ an innovative approach has been employed for identifying and following facial landmark attributes based on 3D static- and dynamic-range statistics. The productiveness of the identified landmarks is justified by applying it for geometrically founded categorization of facial expression for both formal and casual expressions and head posture evaluation.

Liu and Yin⁵ present a unique infrared thermal video descriptor for the purpose of better recognition of casual emotions. At first, each thermal video is portrayed as a sequence of video clips. The face areas in the clips are bound to the front look depending on the scale-invariant feature transform (SIFT) flow. Every video is described by a histogram of the bag of SIFT flow and facial temperature changes the video words. The histogram obtained is applied as a descriptor for categorization by using support vector machine (SVM).

A unique⁶ 3D binary edge (3D-BE) feature has been taken into account for symbolizing high-resolution dynamic facial expressions in three dimensions. For achieving time-related facts, authors have applied a method which uses a latent-dynamic conditional random field along with the 3D-BE. The obtained pain expression identification system confirms that 3D-BE expresses the facial features related to pain very well, and demonstrates the ability of noncontact pain identification from 3D facial expression images.

Authors study⁷ the parts of the face that reveal facial expressions by employing a reverse correlation method, and additionally build up a unique 3D local normal component characteristic expression that depends on human understandings. An innovative 3D normal component-related characteristic (3D-NLBP) is suggested to describe positive and negative expressions (e.g. happiness and sadness). This method obtains an acceptable efficiency and has been justified by implementing on both high-resolution datasets and real-time low-resolution depth map videos. The work in Ref. 8 explores the common idea of applying a single, unchanged, 3D surface as an estimate to the shape of all input faces. Authors elaborate that this directs to a genuine, effective and simple-to-execute method for frontalization. Significantly, in addition to this, it generates aesthetic new frontal views and is amazingly efficient when applied for face recognition and gender evaluation.

An automated method for identifying considerable collection of anthropometric landmarks on 3D face images has been presented in Ref. 9. This method depends entirely on shape without the use of any texture for identifying morphologically important landmarks. For relocation of agreements and automated identification of landmarks, a dense corresponding points-assisted morphable model has been fitted to an undiscovered query face. The approach presented here is able to identify all the collections of previously described landmarks containing subtle ones that are otherwise very complex to identify manually.

For continuing to provide standardization¹ and to support the domain, to make further advancement overcoming the present drawbacks, the FG 2015 Facial Expression Recognition and Analysis challenge (FERA 2015) will provoke contestants to evaluate FACS Action Unit (AU) intensity and AU occurrence on a familiar standard dataset having dependable human interpretations. Three sub-contests have been explained: the AU occurrence identification, the AU intensity evaluation for pre-separated data and completely automated AU intensity evaluation.

The research work in Ref. 10 attempts to tackle this difficulty by suggesting a completely automated 3D facial expression identification model, which handles high-dimensionality complexity in a twofold solution. The average recognition correctness is 79% for this approach by employing the Bosphorus dataset and is 79.36% by employing the BU-3DFE dataset.

Azazi *et al.*¹¹ present a freshly prepared database of 3D videos for casual facial expressions from a heterogeneous collection of grown-up youth. Well-evaluated emotion inductions were employed to draw out emotional expressions and

paralinguistic communication. Facial Action Coding System has been applied for obtaining facial actions' frame-level ground truth.

Zhang *et al.*¹² present a characteristic — Nebula characteristic — that is latest, condensed and related to space and time in four dimensions for upgrading expression and facial movement estimation efficiency. Given a volume related to space and time, the voxelized data is fit to a cubic polynomial. For formal expressions identification, the Nebula characteristic method demonstrates enhancement over LBPTOP on the depth images and remarkable betterment over the nondynamic 3D-only method.

A unique approach¹³ for identifying and following landmark facial characteristics on completely geometric range models in three and four dimensions also includes fitting of a latest multi-frame restricted 3D temporal deformable shape model (TDSM) to a series of data. The effectiveness of the 3D feature identification and following over the range model series has also been evaluated using an application which depends on a 3D geometry-assisted face and expression evaluation and expression collection separation. Authors tried out their approach on the public databases, namely BU-3DFE, BU-4DFE and FRGC 2.0, and also applied their process later on their prepared dataset of 3D dynamic instantaneous expressions.

Authors in Ref. 14 created prominent 3D head posture assessment on 3D facial expression models by handling the complexity of head pose assessment using a generic model and relative importance-based separation on a Laplacian fairing model. The processes for pose assessment are validated using both static and dynamic 3D facial datasets.

An interestingly new¹⁵ dynamic curvature-dependent method (dynamic shape-index-related process) has been presented for 3D face recognition. This approach is induced by the thought of dynamic texture in two dimensions and surface descriptors in three dimensions. The 3D dynamic surface is represented by its surface bending-dependent shape-index knowledge. The surface characteristics are attributed to local zones in the direction of the temporal axis.

A unique head angle assessment system has been elaborated in Ref. 16 which performs identification of region by employing Kinect, followed by face recognition, feature tracing and ultimately head angle analysis utilizing an active camera. Analysis is performed on a common head model in three dimensions generated using the previous information about shape of the head and the geometry-related association between the images in two dimensions and a general model in three dimensions.

A research work proposes¹⁷ the current progresses made in 3D and 4D facial expression analyses with the evolution in 3D facial data capture and tracing, and extends recently accessible 3D/4D face datasets appropriate for 3D/4D facial expressions recognition and also the existent facial expression detection systems that take into account either 3D or 4D data in a detailed manner.

Automatic facial expression recognition¹⁸ recognizes a collection of facial key points by calculating SIFT characteristic descriptions. The evaluations are related to depth images associated with human face applicability. An average identification accuracy of 78.43% can be found by training an SVM for every facial expression to be

identified, and fusing them to generate a multi-class classifier by applying on the BU-3DFE dataset.

For identifying expressions on the face, a method focused¹⁹ on the modes of dynamic data in three dimensions. Authors present a recently developed high-resolution 3D dynamic dataset associated with facial expression, which has been made accessible to the scientific researchers. The dataset has been evaluated by employing authors' facial expression identification investigation applying an HMM-related 3D facial descriptor related to time and space.

The contribution of our work is designing a system that can efficiently map the 2D plane video sequences into a 3D homomorphic plane and transform 2D face tracking algorithm to 3D one, followed by a 2D rendering of the mapped vector on the video in real time.

The proposed algorithm manages to outperform the existing video-based facial tracking and emotion tracking in terms of both accuracy and computation constraints through nonlinear optimization of the projected feature vectors.

We first adopt a facial tracking technique and extend it to multiple facial tracking on image with face and lighting invariance. Once such a robust system is designed that could detect multiple faces at different postures and with different emotions, we extended the algorithm to work on sequence of images. We followed this by optimizing the correlation vectors between the sequences.

3. Methodology

The process of following of facial expression-induced emotions in the consecutive video frames is implemented with the help of a facial prototype of the candidate person with an approximated initialization. The prototype consisting of 70 points has been displayed in Fig. 2.

For each single point one classifier is employed and altogether 70 lightweight classifiers are employed for training the prototype following this algorithm. By locating a basic rough region, the classifiers inspect for a tiny zone (which gives rise to the name "local") about every point for a strengthened fit, and the prototype is then activated in incremental manner in the trail obtaining the finest fit, gradually giving

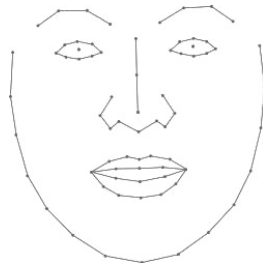


Fig. 2. Facial prototype containing 70 points.

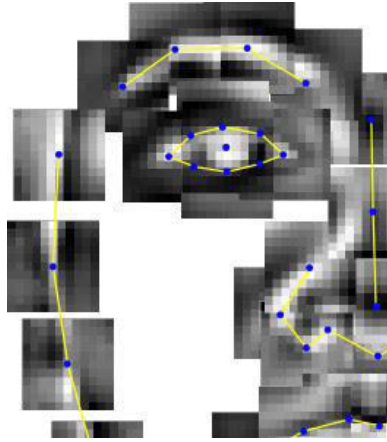


Fig. 3. The process of fitting the prototype data points applied on an image.

rise to the best fit. The method of fitting on a gray-scale face image has been demonstrated in Fig. 3.

A prototype of a face is actually the interpreted data of a human face. A facial prototype can be simply formed as the geometrical characteristics of faces do not vary from person to person to a great extent. The tagged face dataset employed to form a face prototype used in this research work has been illustrated in Fig. 4.

The process of PCA (Principal Component Analysis) (Fig. 5) has been employed to form a prototype from the collected observations. At first, the average observation points are obtained and then the average points of whole observations are determined. PCA is employed to extract forcefully the discrepancy as a linear vector set. The process of classifying emotions has been performed depending on the increasing and decreasing likelihood characteristics reflected from the human face by projecting the face image on a 3D plane. Each of the specific emotions is attributed to the changes in the characteristics. Features like eyes, lips, mouth, chin and cheeks have been taken into account.

In the proposed system, all of the possible facial feature points are taken to be part of multiple planes and have been processed in parallel fashion to determine the maximum likelihood of the point. The proposed system eliminates the drawbacks of



Fig. 4. An imprecise face interpretation.

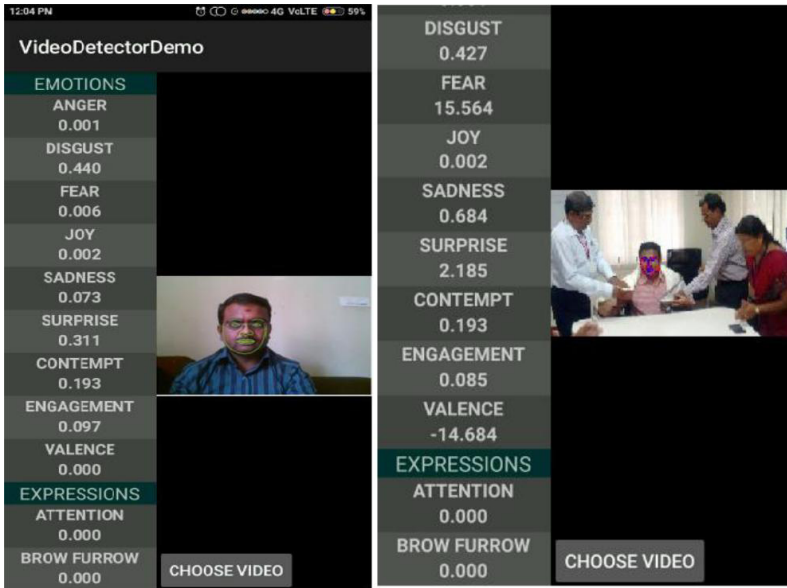


Fig. 5. Feature breakdown.

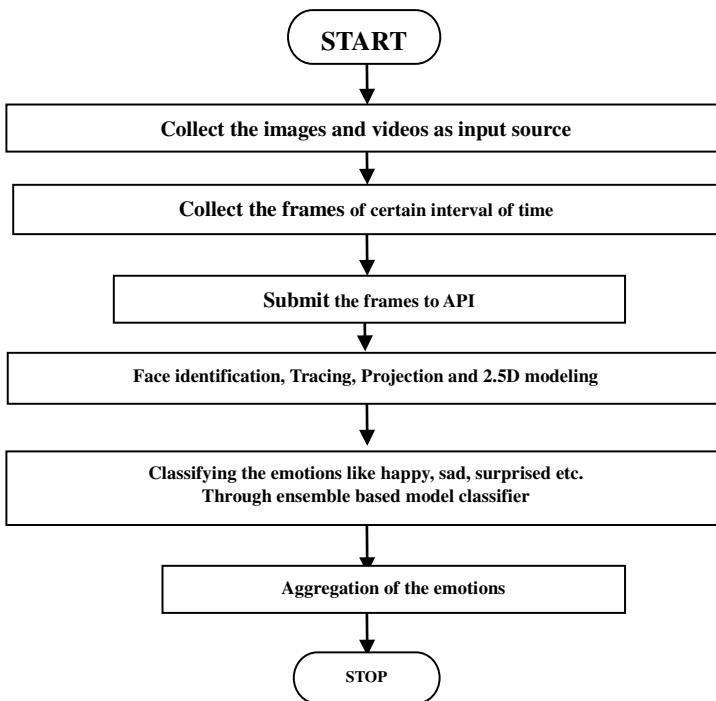


Fig. 6. Flowchart of the proposed work.

2D systems and is very much able to detect the facial expression using complete geometrical model of the face. Multi-planes and multiple classifiers provide immense amount of granularity to the proposed system. The flowchart of the proposed work is shown in Fig. 6.

4. Results and Discussion

Various video frames with multiple faces demonstrating different facial expressions as well as emotions have been used for the purpose of experiment. The detailed result has been tabulated in Table 2.

The accuracy percentage of the proposed method in comparison to other state-of-the-art methods has been presented in Table 3.

Table 2. Different test cases for emotion recognition.

Image	Dominant emotion	Detected average emotion (s)	Image	Dominant emotion	Detected average emotion(s)
	Edited	Happiness		Natural	Happiness
	Natural	Anger		Natural	Happiness
	Natural	Joy & Surprise		Natural	Surprise
	Natural	Anger & Surprise		Natural	Joy
	Natural	Anger		Edited	Happiness & Joy
	Natural	Anger		Natural	Joy
	Natural	Happiness & Joy		Edited	Disguise & Joy
	Natural	Anger		Natural	Contempt
	Natural	Anger & Fear		Natural	Joy

Table 2. (Continued)



























Image	Dominant emotion	Detected average emotion (s)	Image	Dominant emotion	Detected average emotion(s)
	Natural	Joy		Natural	Sadness & Anger
	Natural	Contempt		Natural	Joy
	Natural	Disgust & Fear		Natural	Disgust
	Natural	Disgust		Natural	Sadness
	Natural	Joy		Natural	Anger
	Natural	Anger		Edited	Joy
	Natural	Sadness		Natural	Contempt
	Natural	Joy		Natural	Contempt
	Natural	Sadness		Natural	Joy
	Edited	Joy		Natural	Anger
	Edited	Fear		Natural	Fear
	Natural	Sadness & Joy		Natural	Disgust
	Natural	Joy		Natural	Joy

Table 2. (Continued)







Image	Dominant emotion	Detected average emotion (s)	Image	Dominant emotion	Detected average emotion(s)
	Natural	Joy		Natural	Contempt
	Natural	Disgust & Fear		Natural	Anger
	Natural	Sadness		Natural	Fear

Table 3. Comparison of emotion recognition accuracy.

S. no.	Method name	Accuracy (%)
1	AUC ¹	58.00
2	Linear SVM ²	81.60
3	Matrix completion theory (SAMC) ³	83.00
4	SVM (texture-based) ⁶	73.00
5	Proposed method	74.347

It can be concluded from Table 3 that the proposed system is robust in nature for accurately recognizing the facial expressions having a competitive accuracy with respect to other existing systems.

5. Conclusion

In this paper, a novel 3D face emotion recognition method has been presented. A 3D facial model has been constructed from 2D facial images. We then extracted the HoG features from the 3D facial model and trained a convolutional neural network with the normalized facial features. We presented the experimental results on a set of images captured in real time. The accuracy performance of the proposed method has been compared with other state-of-the-art emotion recognition algorithms and it is found that the proposed method achieves a competitive accuracy percentage of 74.347% with respect to others. The performance of the proposed method has the attributes of effectiveness and robustness. In future, our main objective is to increase the performance accuracy so that it can compete its human counterpart.

References

1. Z. Zhang, J. Girard, Y. Wu, X. Zhang, P. Liu, U. Ciftci, S. Canavan, M. Reale, A. Horowitz, H. Yang, J. Cohn, Q. Ji and L. Yin, "Multimodal spontaneous emotion corpus for human behavior analysis," in *Proc. 2016 IEEE Int. Conf. Computer Vision and Pattern Recognition* (2016), pp. 3438–3446.

2. Y. Wu and Q. Ji, "Constrained deep transfer feature learning and its applications," in *Proc. 2016 IEEE Conf. Computer Vision and Pattern Recognition* (2016), pp. 5101–5109.
3. S. Tulyakov, X. Alameda-Pineda, E. Ricci, L. Yin, J. Cohn and N. Sebe, "Self-adaptive matrix completion for heart rate estimation from face videos under realistic conditions," in *Proc. 2016 IEEE Int. Conf. Computer Vision and Pattern Recognition* (2016), pp. 2396–2404.
4. S. Canavan, P. Liu, X. Zhang and L. Yin, "Landmark localization on 3D/4D range data using a shape index-based statistical shape model with global and local constraints," *Comput. Vis. Image Understand.* **139**, 136–148 (2015).
5. P. Liu and L. Yin, "Spontaneous facial expression analysis based on temperature changes and head motions," in *Proc. 2015 11th IEEE Int. Conf. and Workshops Automatic Face and Gesture Recognition* (2015).
6. X. Zhang, L. Yin and J. Cohn, "Three dimensional binary edge feature representation for pain expression analysis," in *Proc. 2015 11th IEEE Int. Conf. and Workshops Automatic Face and Gesture Recognition* (2015), pp. 1–7.
7. X. Zhang, Z. Zhang, D. Hipp, L. Yin and P. Gerhardstein, "Perception driven 3D facial expression analysis based on reverse correlation and normal component," in *Proc. 2015 Int. Conf. Affective Computing and Intelligent Interaction* (2015), pp. 616–622.
8. T. Hassner, S. Harel, E. Paz and R. Enbar, "Effective face frontalization in unconstrained image," in *Proc. 2015 IEEE Conf. Computer Vision and Pattern Recognition* (2015), pp. 4295–4304.
9. S. Z. Gilani, F. Shafait and A. Mian, "Shape-based automatic detection of a large number of 3D facial landmarks," in *Proc. 2015 IEEE Conf. Computer Vision and Pattern Recognition* (2015), pp. 4639–4648.
10. M. F. Valstar, T. Almaev, J. M. Girard, G. McKeown, M. Mehu, L. Yin, M. Pantic and J. F. Cohn, "FERA 2015: Second facial expression recognition and analysis challenge," in *Proc. 2015 11th Int. Conf. and Workshops Automatic Face and Gesture Recognition* (IEEE, 2015).
11. A. Azazi, S. L. Lutfi, I. Venkat and F. Fernández-Martínez, "Towards a robust affect recognition: Automatic facial expression recognition in 3D faces," *Expert Syst. Appl.* **42**(6), 3056–3066 (2015).
12. X. Zhang, L. Yin, J. Cohn, S. Canavan, M. Reale, A. Horowitz, P. Liu and J. Girard, "BP4D-Spontaneous: A high resolution spontaneous 3D dynamic facial expression database," *Image Vis. Comput.* **32**, 692–706 (2014).
13. M. Reale, X. Zhang and L. Yin, "Nebula feature: A space-time feature for posed and spontaneous 4D facial behavior analysis," in *Proc. 10th IEEE Int. Conf. Automatic Face and Gesture Recognition* (2013).
14. S. Canavan, X. Zhang and L. Yin, "Fitting and tracking 3D/4D facial data using a temporal deformable shape model," in *Proc. 2013 IEEE Int. Conf. Multimedia and Expo* (2013).
15. P. Liu, M. Reale and L. Yin, "Saliency-guided 3D head pose estimation on 3D expression models," in *Proc. 15th ACM Int. Conf. Multimodal Interaction* (2013).
16. S. Canavan, Y. Sun and L. Yin, "A dynamic curvature based approach for facial activity analysis in 3D space," in *Proc. 2012 IEEE CVPR Workshop on Socially Intelligent Surveillance and Monitoring* (2012).
17. P. Liu, M. Reale and L. Yin, "3D head pose estimation based on scene flow and a 3D generic head model," in *Proc. 2012 IEEE Int. Conf. Multimedia and Expo* (2012), doi:10.1109/ICME.2012.61.
18. G. Sandbach, S. Zafeiriou, M. Pantic and L. Yin, "Static and dynamic 3D facial expression recognition: A comprehensive survey," *Image Vis. Comput.* **30**(10), 683–697 (2012).

19. S. Berretti, B. Ben Amor, M. Daoudi and A. del Bimbo, "3D facial expression recognition using SIFT descriptors of automatically detected key point," *Vis. Comput.* **27**, 1021 (2011).



Dayanand G. Savakar has 28 years of teaching experience and 14 years of research experience. He has completed his B.E. degree from Karnataka University Dharwad, Post-graduation from Birla Institute of Technology and Science, Rajasthan, and Ph.D. from Visvesvaraya Technological University, Belgaum, India. He is now working as a Professor in the Department of Computer Science, Post Graduate Centre, Rani Channamma University, Vijayapur, Karnataka, India. He has published more than 60 research articles

in international journals/conferences. Currently, he is guiding eight Ph.D. candidates. He has served/been serving as a member of various boards of several universities. His areas of interests are image processing, pattern recognition and information security.



Ravi Hosur has 13 years of teaching experience and six years of research experience. He has completed his B.E. degree and M.Tech. degree both from Visvesvaraya Technological University, Belgaum, Karnataka India. He is now pursuing his Ph.D. degree at the above university. Currently, he is working as an Assistant Professor in the Department of Computer Science and Engineering, BLDEA's Vachana Pitamaha Dr. P. G. Halakatti College of Engineering & Technology, Vijayapur, Karnataka, India. He has

published more than eight research articles in international journals/conferences. His areas of interests are digital image modeling and processing, pattern recognition and information security.