Analysis of Human-Machine Interaction Through Facial Expression and Hand-Gesture Recognition

Rajeshree Rokade, Ketki Kshirsagar, Jayashree Sonawane, Sunita Munde

Abstract*:*This paper focuses on a review of recent work on facial expression and hand gesture recognitions. Facial expressions and hand gestures are used to express emotions without oral communication. The human brain has the ability to identify the emotions of persons using expressions or hand gestures within a fraction of a second. Research has been conducted on human–machine interactions (HMIs), and the expectation is that systems based on such HMI algorithms should respond similarly. Furthermore, when a person intends to express emotions orally, he or she automatically uses complementary facial expressions and hand gestures. Extant systems are designed to express these emotions through HMIs without oral communication. Other systems have added various combinations of hand gestures and facial expressions as videos or images. The meaning or emotions conveyed by particular hand gestures and expressions are predefined in these cases. Accordingly, the systems were trained and tested. Further, certain extant systems have separately defined the meanings of such hand gestures and facial expressions.

Keywords : Facial gesture, Hand gesture, Segmentation, Feature extraction, Classification, Experimentation

I. INTRODUCTION

Gesture or sign recognition is a way for computers to begin to understand human body language, thus building richer bridges between machines and humans. The most expressive way humans display emotions is through facial expressions. Humans detect and interpret facial expression from a scene with little or no effort. Facial gestures (facial muscle actions) regulate our social interactions. Most approaches to automatic facial gesture analysis in facial image sequences attempt to recognize a set of prototypical emotional facial expressions (e.g., "happy," sad," "fear," "surprise," "despair," "interest," "pleasure," "irritation," "pride," "anger," and "disgust"). Research on facial gesture analyses from frontal-view face images investigated whether and to what extent human facial and hand gestures are recognized automatically. Indian classical dance-step recognition is one of the applications of combined facial and hand-gesture recognition systems.

Hand gestures are not universal. Every culture has its own interpretation. However, some of signs or gestures are more common than others. Manual communications have been developed in places or situations where speech is neither practical nor permitted, such as SCUBA diving, television recording, loud workspaces, stock exchanges, and baseball. Recently, there has been a movement to teach and encourage the use of sign language with toddlers before they learn to talk. Sign language is an effective communication technique for non-deaf persons who are speech impaired. Technology can also be used to generate speech or text for teleconferencing and for security purposes. It can also be applied to sign interpretation and learning. In the field of bedridden patient monitoring, one can combine facial and hand gestures to provide instructions or to verify the status of the patient without verbal communications.

Chen et. al. proposed human–computer interaction (HCI) modalities [1] based on hand gestures and facial expression in elderly care, smart-home applications, and intelligent space applications. Filipino students are often expressive [2] during programming sessions and typing. Thus, as the frequency of hand signing increases, confusion and frustration in programming increases. This helps the development of culturally sensitive facial expression and gesture tutoring systems. Hand gestures and facial expressions improve performance using the spatiotemporal pyramid representation proposed by Zhipeng Zhao and Ahmed Elgammal [3]. The spatiotemporal pyramid was built using a weighting histogram from different layers of subdivision.

Mihai Gavrilescu [4] recognized emotions from facial expressions and body postures based on stochastic context-free grammar (SCFG) using eight hand gestures and body postures. Anger, sadness, fear, happiness, surprise, and disgust were accounted. Montgomery and Haxby [5] developed a mirror neuron system (MNS), which mapped actions based on motor representations. Experimentation of functional magnetic resonance imaging was used where participants imitated, viewed, and produced facial expressions and social hand gestures. There are distinct representations of different types of social nonverbal communications in an MNS. A media-player system controlled by facial expression and hand gestures [6] was proposed by Agarwal and Umer. To find movement, one landmark point for a finger and 18 landmark points for lips were captured using support vector machines. Especially for the hearing impaired [7], the system was applied based on computed trajectory using the Cam-shift algorithm for face and hand motions. HMM was applied for hand tracking, finger tracking, face detection, and feature extraction. Efficiency was calculated by the number

Revised Manuscript Received on July 22, 2019.

Dr. Rajeshree Rokade, EXTC department, Lokmanya Tilak College of Engineering, Navi Mumbai, India Email: Rajeshree.rokade@gmail.com

Dr. Ketki Kshirsagar, EXTC department, VIIT College, Pune, India. Email: ketkikshirsagar@viit.ac.in

Ms. Jayashree Sonawane, EXTC department, Lokmanya Tilak College of Engineering, Navi Mumbai, India Email: jayashreesonawane@gmail.com

Ms. Sunita Munde, EXTC department, Lokmanya Tilak College of Engineering , Navi Mumbai, India Email: sunitamunde@gmail.com

of objects presented in the video or image, such as one or two hands.

Frameworks for a vision-based multimodal analyzer combining face and body gestures were discussed by Gunes et al. [8].A systematic approach [9] was applied for analyzing emotionsfrom hand signs and facial expressions. “Affectiveanalysis in MMI” described an overview of affective analyses of facial signs and expressions, supported by psychological studies explaining emotionsas discrete points of an emotional space. “Facialexpression analysis” and “gesture analysis” included the algorithms and experimental results. The motions of tracked feature points were translated to MPEG-4 to provide facial animation parameters (FAP). It described motion at a high-level manner. Hand gestures and hand segments were located in a video sequence via color segmentation and motion estimation algorithms. The position of segments were tracked to find the hand’s position over time and fed into a HMM architecture to obtain the affective gesture estimation.

Prabhu and Jayagopi [10] developed a real-time system based on multimodal emotion recognition using signs and facial gestures. Four emotional facial expressions were studied and more descriptors of signing and mental states were added for accuracy. Audio and video of both parameters are required for joint learning.

Ju and Kang [11] developed a robot interaction-based hand gesture and face pose system. Facial points were used to determine facial expressions. 3-dimensional (3D) face models were constructed for finding feature points using the variation of facial expressions. Based on expressions, background color and objects in displays were changed. Several emotional facial expressions were used by Pimpalkar et al. [12] for facial expressions. One multimodal approach is the facial expression recognition system (FERS), which recognizes facial expressions. Another approach is the hand gesture recognition system.These approaches are based on Gaussian mixture models (GMM) for face recognition. Viola–Jones and CamShift algorithms [26] were used for hand tracking and prediction.

Pateraki et al. [13] provided an integrated approach for tracking faces, facial features, and hands, applicable to the interaction of robots in public spaces. A blob tracker was used to track the skin of the right hand, the left hand, and a face. This hybrid approach was utilized by integrating an appearance-based detector and a feature-based tracker for the eyes, nose, and mouth. Figure 1 shows the system flow. The first challenge in any type of sign or gesture recognition system is segmentation. Segmentation includes clearly separating out the face and hand portions from the background. Feature extraction is the second challenge. The third challenge is classifying reference patterns to be matched with all possible segments of input signals.

This paper is divided in to five sections. Section 1 describes the database for training and testing. Section 2 describes the segmentation techniques applicable for faces and hands. Section 3 explains feature extraction from facial expressions and gestures. Section 4 describes classification and results. Section 5 presents the conclusion.

Hand gestures are not universal. Every culture has its own interpretation. However, some of signs or gestures are more common than others. Manual communications have been developed in places or situations where speech is neither practical nor permitted, such as SCUBA diving, television recording, loud workspaces, stock exchanges, and baseball. Recently, there has been a movement to teach and encourage the use of sign language with toddlers before they learn to talk. Sign language is an effective communication technique for non-deaf persons who are speech impaired. Technology can also be used to generate speech or text for teleconferencing and for security purposes. It can also be applied to sign interpretation and learning. In the field of bedridden patient monitoring, one can combine facial and hand gestures to provide instructions or to verify the status of the patient without verbal communications.

Chen et. al. proposed human–computer interaction (HCI) modalities [1] based on hand gestures and facial expression in elderly care, smart-home applications, and intelligent space applications. Filipino students are often expressive [2] during programming sessions and typing. Thus, as the frequency of hand signing increases, confusion and frustration in programming increases. This helps the development of culturally sensitive facial expression and gesture tutoring systems. Hand gestures and facial expressions improve performance using the spatiotemporal pyramid representation proposed by Zhipeng Zhao and Ahmed Elgammal [3]. The spatiotemporal pyramid was built using a weighting histogram from different layers of subdivision.

Mihai Gavrilescu [4] recognized emotions from facial expressions and body postures based on stochastic context-free grammar (SCFG) using eight hand gestures and body postures. Anger, sadness, fear, happiness, surprise, and disgust were accounted. Montgomery and Haxby [5] developed a mirror neuron system (MNS), which mapped actions based on motor representations. Experimentation of functional magnetic resonance imaging was used where participants imitated, viewed, and produced facial expressions and social hand gestures. There are distinct representations of different types of social nonverbal communications in an MNS. A media-player system controlled by facial expression and hand gestures [6] was proposed by Agarwal and Umer. To find movement, one landmark point for a finger and 18 landmark points for lips were captured using support vector machines. Especially for the hearing impaired [7], the system was applied based on computed trajectory using the Cam-shift algorithm for face and hand motions. HMM was applied for hand tracking, finger tracking, face detection, and feature extraction. Efficiency was calculated by the number of objects presented in the video or image, such as one or two hands.

Frameworks for a vision-based multimodal analyzer combining face and body gestures were discussed by Gunes et al. [8].A systematic approach [9] was applied for analyzing emotionsfrom hand signs and facial expressions. “Affectiveanalysis in MMI” described an overview of affective analyses of facial signs and expressions, supported by psychological studies explaining emotionsas discrete points of an emotional space. “Facialexpression analysis” and

"gesture analysis" included the algorithms and experimental results. The motions of tracked feature points were translated to MPEG-4 to provide facial animation parameters (FAP). It described motion at a high-level manner. Hand gestures and hand segments were located in a video sequence via color segmentation and motion estimation algorithms. The position of segments were tracked to find the hand's position over time and fed into a HMM architecture to obtain the affective gesture estimation.

Prabhu and Jayagopi [10] developed a real-time system based on multimodal emotion recognition using signs and facial gestures. Four emotional facial expressions were studied and more descriptors of signing and mental states were added for accuracy. Audio and video of both parameters are required for joint learning.

Ju and Kang [11] developed a robot interaction-based hand gesture and face pose system. Facial points were used to determine facial expressions. 3-dimensional (3D) face models were constructed for finding feature points using the variation of facial expressions. Based on expressions, background color and objects in displays were changed. Several emotional facial expressions were used by Pimpalkar et al. [12] for facial expressions. One multimodal approach is the facial expression recognition system (FERS), which recognizes facial expressions. Another approach is the hand gesture recognition system.These approaches are based on Gaussian mixture models (GMM) for face recognition. Viola–Jones and CamShift algorithms [26] were used for hand tracking and prediction.

challenge is classifying reference patterns to be matched with all possible segments of input signals.

This paper is divided in to five sections. Section 1 describes the database for training and testing. Section 2 describes the segmentation techniques applicable for faces and hands. Section 3 explains feature extraction from facial expressions and gestures. Section 4 describes classification and results. Section 5 presents the conclusion.

II. FACIAL AND HAND GESTURES

Gunes [8] presented a survey of research conducted on facial expression and body gesture recognition. The system, communicates different emotions through facial expression, such as with a lowered head position or a lethargic movement. Head and face gestures, hand and arm gestures, and body gestures become the combinations. When a person communicates, he or she generally use head gestures and macro-head movements. Facial segmentation utilizes steps, which include detection or tracking of the face, to identify facial regions, features, and movement [14].Obick [15] developed a taxonomy of motion in terms of movement, activity, and action

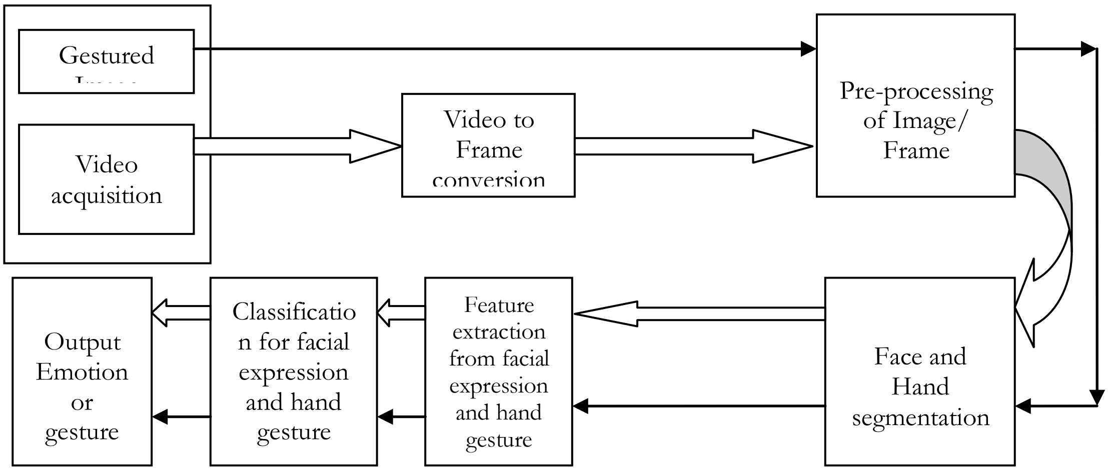

Figure 1. System Flow

Pateraki et al. [13] provided an integrated approach for tracking faces, facial features, and hands, applicable to the interaction of robots in public spaces. A blob tracker was used to track the skin of the right hand, the left hand, and a face. This hybrid approach was utilized by integrating an appearance-based detector and a feature-based tracker for the eyes, nose, and mouth. Figure 1 shows the system flow. The first challenge in any type of sign or gesture recognition system is segmentation. Segmentation includes clearly separating out the face and hand portions from the background. Feature extraction is the second challenge. The third

Chen et al. [1] presented two HCI modalities (i.e., hand gestures and facial expressions) for elderly care, smart home applications, and intelligent space applications. A 320×240 resolution web camera was used to collect 400 samples. To train the system, 500 images were applied. An architecture of two levels was developed for low-level and high-level hand gesture motion analysis and recognition.In the database of [2], many facial expressions from eight male and four females were cataloged.

The database used by Zhao [3] for facial expressions data set was introduced by Doller et al. [16], and the hand gesture data set was collected by Wong et al. [17]. The facial expression database [16] involved two databases, each expressing six different emotions under two lighting setups. Some emotional facial expressions and hand gestures are shown in Figure. 2

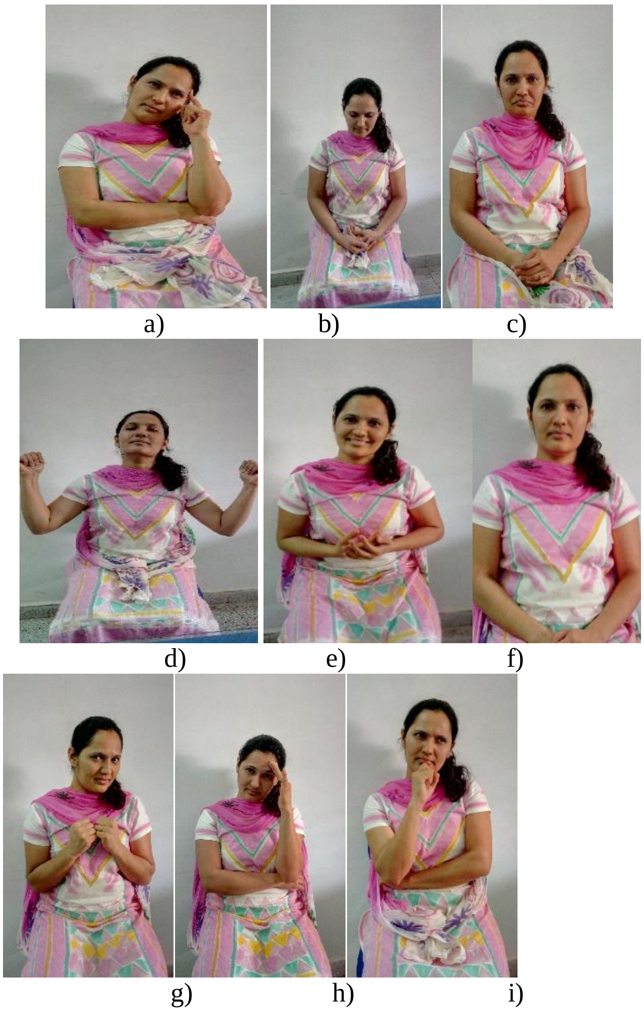

Figure 2: Emotions using facial expressions and hand gestures: a) "thinking;" b) "shame;" c) "sad;" d) "pride;" e) "neutral;" f) "happy;" g) "fear;" h) "disgust;" i) "unsure"

SCFG includes eight combinations of body postures and hand gestures for each emotion. A new database [4] was created to test the system. Databases of 64 signers (32 females and 32 males) were generated for the six basic emotions. The protocol allowed only the upper body to be recorded. Similar 8-to-12-s videos were combined. The LIRIS–ACCEDE standard database was also introduced. FERS classifiers were trained on Cohn–Kanade [18] and tested on the MMI database [19].

An new database was created [5] from 12 participants (7 women and 5 men), aged 22 to 31. All participants were right-handed. Five basic emotions and five hand signs were introduced. Research in psychology [8] has indicated that with distinct facial expressions [20, 21, 22] of at least six emotions are universally associated. Karpouzis et al. [9] created dyadic pairs of gestures and emotional states in their work. More details regarding these pairs can be found in their published paper.

III. SEGMENTATION

The segmentation of faces and hands from video or static imageswas a major challenge, especially when the background was non-uniform or non-static, the signer was old (>60), the distance between camera and person was large, signer had glasses, or signer had a beard. Many existing systems introduced uniform backgrounds or portraits to capture gestures without facial hair. Most videos are front views to reduce segmentation problems. Gunes et. al. [8] catalogued hand gestures [23], gesticulations, language-like gestures, pantomimes, emblems, and sign languages.

Two-level architectures have been developed for hand gesture recognition. At the high level there exists hand gesture recognition and motion analysis. At the low level, there exists hand posture detection and hand tracking [1]. Gestures are comprised of two or more postures. As the number of postures increases, complexity grows. Both wanted and unwanted postures were presented. "Wanted" postures indicate signs, whereas "unwanted" postures do not. Often, unwanted postures are transition motions between two wanted postures. The probability of a wanted posture must be higher than the probability of an unwanted posture. The system consists of a tracking module and an estimator module. The tracking module is developed to track the face. It consists of 2-dimensional (2D)-point measurements, $p_i(u_i,v_i)$, of the tracked features, where $i= 1,...,m$ andm is the number of measurement points. The set of 2D features to be tracked is obtained by projecting their corresponding 3D model points, $p_i\ (X_i,\ Y_i,\ Z_i)$, where $i= 1,...,m$, and m is the number of tracked features. The estimator module estimates 3D motions and expressions.

Facial expressions and hand gestures of the Filipino students were analyzed [2] during programming sessions. The duration of each session was 45 min per student for seven Java programs. A web camera was used to capture their movements. To extract facial information, the Affectiva SDK [22] was introduced. Each face was divided in to five components [4]: eyebrow, cheek, lips, wrinkles, and eye, segmented by the Haar cascade classifier [25], Viola–Jones face detector classifier [26], lip-feature point tracker [27], shortest distance classifier [28], and the Haar cascade classifier [29], respectively. The K-nearest neighbor algorithm [30] determined head orientation whereas Viola–Jones face detection algorithm [31] was used to detect and segment the face.

Gestures are combination of different postures. Hand gestures have two components: local finger motion and global hand motion. Global hand motion refers to changes in the position of the hand and local finger. It is analyzed in relation with global body motion, because they are related. Local finger motion refers to the movement of only the fingers, irrespective of the hand. The Ada-Boost filter is applied to find segmented areas of body or hand. Components of hand signs are shown in Figure. 3.

American Sign Language is composed of sequences of static hand gestures, Local finger motions (i.e., hand is not moving), and global hand motions.

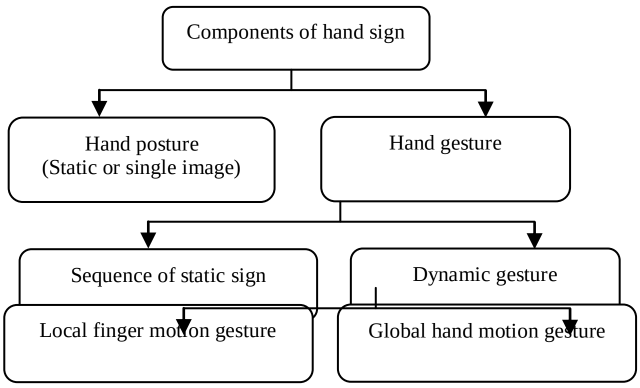

Figure. 3. Components of hand signs

For each signer [5], 10 time series were obtained (five for facial expressions and five for hand gestures). Facial expressions and social hand gestures in MNS were compared based on frontal operculum (FO) activity, superior temporal sulcus (STS), and inferior parietal lobule (IPL). As signs and facial expression change, changes in STS, IPL, and FO are observed using percent of signal changes. For each component, a single segmentation method cannot be applied. Segmentation or localization depends on which component is present in the database. The most popular segmentation methods are based on shape or color. In the color-based approach, the RGB color space is converted to other color spaces. Color-based segmentation is again divided into two approaches: parametric [32] and non-parametric [33]. To track or segment hands in 2D, deformable hand-shape templates [34] are applied, whereas for 3D tracking, model-based approach [34] are applied, which takes advantage of prior knowledge. Facial features can be viewed [9] as slowly varying (i.e., wrinkles), static (i.e., skin complexion), or rapidly varying (i.e., eyebrows rising). Facial expressions can be more accurately recognized [35] from image sequences than still images. The positions and shapes of the mouth, eyes, eyelids, etc., from still images, are accurately extracted by Bassil's experiment [5].

IV. FEATURE EXTRACTION

Feature extraction largely depends on the face and hand representations chosen. Features differ from signs or expressions. Extant systems have proposed various methods for automatic facial-data extraction, including motion energy maps[36]; facial motion and optical flow[36, 37, 38]; feature measurement[39]; Gabor wavelet representation[40];model-based techniques/holistic analysis[41]; principal component analysis (PCA); "Eigenactions;" local feature analysis (LFA); Fisheractions; and independent component analysis [42].

Rokade et al. proposed a thinning algorithm for feature extraction [43]. Kshirsagar et. al. proposed key-point-based feature extraction [44]. Bhuyan and Ghosh [45] proposed a system that tested a limited number of dynamic hand gestures. For dynamic gestures, video object plane was extracted. The Hausdorff tracker [46], [47], [45] was introduced for this extraction. MacLeant and Herperst [47] proposed a skeleton gesture recognition method. Features were extracted using Radon transform [48]. Here the features were rotation invariant. [49] End points of skeletons were used for feature extraction. Bhuyan et al. [50] proposed extracting certain features from the gesture trajectory to identify the form of the trajectory. Trajectory length, location feature, ending hand orientation, average velocity, minimum velocity, maximum velocity, etc., were used to enable a robot to recognize and respond to hand gestures from a human operator. The interface used predefined hand gestures, each representing an individual control command for the motion control of the remote robot. However, it required higher recognition speed at the expense of classification accuracy.

Ming-Hsuan Yang et al. [51] proposed an algorithm for extracting and classifying 2D motion in an image sequence based on motion trajectories. 40 limited gestures (i.e., training data) were used for recognition. For testing the data, large information storage was needed, because increasing training data also increased memory requirements. Kim, Lee, and Jin-Huiark [52] introduced very limited training data. Rokade and Dye [53] provided spelled sign-word recognition and a key frames system, developed with Haar-like features and a statistical approach [1] that concentrates on unique information of areas in low-level hand posture recognition with different backgrounds, under various light conditions. It is robust against image noises. Ada-Boost was introduced to obtain target images and to remove false images to improve detection accuracy. For facial feature recognition, a statistical mode- based real time system was developed [1], which separated rigid and non-rigid motion facial expressions. The system consisted of an extended Kalman filter for global motion of head and 3D anthropometric muscle-based active appearance model for local facial expression recognition. The two sets of controls are used anthropometrical control and muscle actuator.

For facial expression detection [2], the Affective SDK [24] was introduced. It tracks points based on facial action coding system to recognizing human emotions from facial expression [54]. Movement of the face is measured in the form of points. Scores [24] were assigned from the range of 0–100, indicating the likeliness of each expression. Each session was divided into 6,211 intervals of 3 s each. An interval is discarded if there is no facial expression observed. The score of each student is computed by observation (e.g., total number of character removals, total number of character insertions, number of times the code was compiled with errors, number of times the code was compiled without errors, number of times an incorrect code was submitted, and number of times a correct code was submitted). "Engagement over time" was calculated for each student (i.e., how much time the student was expressive during the session). The Spatiotemporal pyramid representation [3] is an extension of the "bag-of-words" model for capturing temporal and geometric arrangements of local motion features. Cuboids of spatiotemporally windowed data with interest points were extracted. The gradients of intensities in the cuboids' were converted into a vector, which were set to the low dimensional motion features space of the video sequence using PCA.
STS, IPL, and FO [5] regions were experimented for different hand gesture and facial expressions. The regions activated by the perception and production of actions were identified based on the response during

imitation, using a threshold and a cluster size. For feature extraction, graphs were observed, indicating changes in signal with respect to different hand gestures and facial expressions. The experiment found differences in peak location for expression features and hand gesture in IPL and FO, whereas no visible differences in peak features were found for STS.The facial features [5] define a mapping between FAPs and the movement of specific facial features. Definition parameters and FPs reflect the salient points on the human face. 15 feature points were labeled, tracked, and detected based on FPs. Distances between these points were specified to find facial expressions.The distance between the center of a segmented head, right hand, and left hand were considered as features. $C_f=(X_f, Y_f)$ are the coordinates of the head center. $C_{rh}=(X_{rh}, Y_{rh})$ and $C_{lh}=(X_{lh}, Y_{lh})$ are the coordinates of the right and left hand center, respectively.

V. CLASSIFICATION AND DISCUSSION

Hand gestures and facial expression detection can be used for intelligent space applications [1], such as elderly care and smart home applications. Seven expressions (20 testing samples each) were specified to find the recognition rate. The expressions are recognized for two situations: one person-dependent and another person-independent. The person-depended recognition observed rate was 86.1%, whereas the person independent rate was 76.2%.

Table I: Comparison of standard emotions

Sr. No.	Research	Method	Test sequences	Recognition rates [%]
1	Koelstra and Pantic [55]	automatic localization, Temporal segments	Cohn–Kanade [59] and MMI database	70.25
2	Valstar and Pantic [56]	*Gabor wavelet*	Cohn–Kanade [59] and MMI database, 487 gray scale recordings of 97 signers	72.0
3	Gavrilescu 2014 [57]	Haar-Cascade classifier, Facial Action Coding system (FACS)	-	83.4
4	Mihai Gavrilescu [4]	neural networks and classifiers	Cohn–Kanade [59] and MMI database [19]. Videos of length 8–12 s from 64 signers	86.4
5	Bartlett et al. [58]	holistic spatial analysis	Ekman–Hager facial action exemplars	53.44
6	Prabhu and Jayagopi [10]	Multimodal emotion recognition	MUG [60] database with 242 videos of 86 signers	83.83
8	Ju and Kang [11]	GMM and Viola and Jones [31]	10 students and 300 images	92.24

Six expressions [2] were observed during programming. The system was tested on 12 people. The long total programming time (45 min per person) made it difficult for subjects to remember their emotions at end of the session. Another limitation was that the system sometimes showed incorrect recognition of gestures.

SCFG [4] is a database containing eight combinations of hand gesture and body postures for each emotion. The gesture emotion recognition system consisted of an Ada-Boost filter to find whether the frame contain body or hand features. Additionally, frames were passed through a Haar cascade classifier, which classified as finger, hand, and body features. For recognition of exact emotions, classifier output was applied to the global body motion neural network, local finger motion neural network, and the global hand motion neural network. The accuracy increased as the number of combinations in SCFG increased from 65.1% to 75%. The recognition rate for the number of combinations in SCFG/emotion for 2, 4, 6, and 8 were 65.1%, 67.5%, 67.0%, and 75%, respectively for controlled scenarios. Table 1 shows the recognition rate for six standard emotions: 86.4%. The overall recognition rate by Bartlett et al. [58] and Prabhu and Jayagopi [10] were 53.44% and 83.83%, respectively. The overall recognition rate obtained from Ju and Kang [11] was 87.98%.

The difference in feature peak locations for facial expression and hand gesture [5] in IPL and FO showed no visible difference in STS peak. However, differential responses outside the MNS were observed in Figure. 4. Responses outside of MNS are shown in Figure. 4. There was a large response for facial expression compared to hand expression during the activation of perceptual areas. For extra-striate body area, there was a larger response for viewing hand gestures compared to facial expressions (illustrated in red). There was large activity in the facial expression of the bilateral precentral gyrus for facial expressions and a more dorsal (backside) hand region of the left precentral gyrus for hand gestures

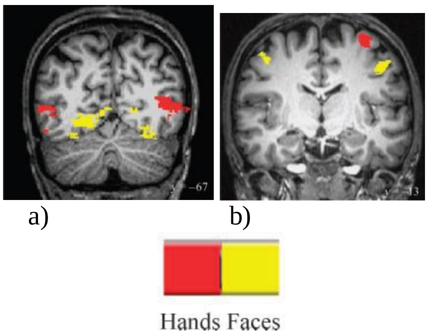

Figure 4: Responses outside MNSa) Differential Activation in perceptual areas B) Differential activation in the precentralgyrus

The distances [9] between center-of-head, left hand, and right hand features are classified as hand clapping (high-frequency and low frequency), hand-lifting (low speed and high speed), hands over the head, and Italianate postureusing HMMs. Fifteen video sequences were thus acquired:three were used for the initialization of the HMM parameters,seven were used for training and parameter re-estimation, andfive were used for testing. The overall recognition rate for the seven gestures was 94.3%.

VI. CONCLUSION

A human's emotional state can be defined by their facial expression and hand gestures. In this paper, we summarized the existing hand gestures and facial expression techniques, limitations, and results. Even if hand gestures do not comprise a universal language, most of the gestures are the same, universally, and are related to facial expressions and emotions.

REFERENCES

[1] Qing Chen; Marius D. Cordea; Emil M. Petriu; Thomas E. Whalen; Imre J. Rudas; Annamaria, Varkonvi-Koczy, "Hand-Gesture and Facial-Expression Human-Computer Interfaces for Intelligent Space Applications",2008 IEEE International Workshop on Medical Measurements and Applications, 2008 PP: 1 - 6.

[2] Thomas James Tiam-Lee; Kaoru Sumi, "Analyzing Facial Expressions and Hand Gestures in Filipino Students' Programming Sessions", 2017 IEEE Conference on Culture and Computing (Culture and Computing) , 2017, PP: 75 - 81

[3] Zhipeng Zhao and Ahmed Elgammal, "Spatiotemporal pyramid representation for recognition of facial expressions and hand gestures", 8th IEEE International Conference on Automatic Face & Gesture Recognition, 2008, PP: 1 - 6

[4] Mihai Gavrilescu, "Recognizing emotions from videos by studying facial expressions, body postures and hand gestures", 23rd Telecommunications Forum Telfor (TELFOR), 2015, PP: 720 – 723

[5] Kimberly J. Montgomery and James V. Haxby, "Mirror Neuron System Differentially Activated by Facial Expressions and Social Hand Gestures: A Functional Magnetic Resonance Imaging Study", Journal of Cognitive Neuroscience, Vol.: 20 (10), PP: 1866 - 1877, 2008

[6] Swapna Agarwal and Saiyed Umer, "MP-FEG: Media Player controlled by Facial Expressions and Gestures", Fifth National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG), 2015, PP: 1- 4

[7] Naresh Kumar, "Motion trajectory based human face and hands tracking for sign language recognition", 4th IEEE Uttar Pradesh Section International Conference on Electrical, Computer and Electronics (UPCON), 2017, PP: 211 – 216, 2017

[8] Hatice Gunes, Massimo Piccardi and Tony Jan, "Face and Body Gesture Recognition for a Vision-Based Multimodal Analyzer", Pan-Sydney Area Workshop on Visual Information Processing (VIP2003), Sydney, 2003

[9] Kostas Karpouzis, Amaryllis Raouzaiou, Athanasios Drosopoulos, Spiros Ioannou, Themis Balomenos, Nicolas Tsapatsoulis, and Stefanos Kolli, "Facial Expression and Gesture Analysis for emotionally-rich Man-Machine interaction", Chapter V, Emotionally-Rich Man-Machine Interaction, 2014

[10] Mahesh Krishnananda Prabhu and Dinesh Babu Jayagopi, "Real Time Multimodal Emotion Recognition System using Facial Landmarks and Hand over Face Gestures", International Journal of Machine Learning and Computing, Vol. 7, No. 2,2017

[11] Myung-Ho Ju and Hang-Bong Kang, "Emotional Interaction with a Robot Using Facial Expressions, Face Pose and Hand Gestures", Int J Adv Robotic Sy , Vol. 9, 2012

[12] Amit Pimpalkar, Chaitali Nagalkar, Sonal Waghmare, Keshari Ingole, "Gesture Recognition of Human Behavior using Multimodal Approach", IJCAT – International Journal of Computing and Technology, Vol. 1, No.1, February 2014

[13] Maria Pateraki and Haris Baltzakis and Panos Trahanias, "An integrated approach for visual tracking of hands, faces and facial features", Proc. of the Workshop on Measuring and Understanding Human Movements and Emotions, held within the IEEE International Conference on Robotics and Automation (ICRA), Shanghai, China, 2011

[14] A. Kapoor and R. W. Picard. "Real-time, fully automatic upper facial feature tracking" Proceedings of 5th International Conference on Automatic Face and Gesture Recognition, May 2002

[15] A. Bobick, "Movement, activity, and action: the role of knowledge in the perception of motion," Royal Society Workshop on Knowledge based Vision in Man and Machine London, England, February, 1997.

[16] P. Doll´ar, V. Rabaud, G. Cottrell, and S. Belongie, "Behavior recognition via sparse spatio-temporal features", VS-PETS, October 2005.

[17] Wong, T. Kim, and R. Cipolla "Learning motion categories using both semantic and structural information", CVPR07, pages 1–6, 2007.

[18] Kanade, T., Cohn, J. F., Tian, Y., "Comprehensive database for facial expression analysis", 4th IEEE International Conference onAutomatic Face and Gesture Recognition, pp. 46-53, Grenoble, 2000.

[19] Pantic, M., Valstar, M. F., Rademaker, R., Maat, L., "Web-based database for facial expression analysis", IEEE InternationalConference on Multimedia and Expo, pp. 317 - 321, July 2005.

[20] P. Ekman and W. V. Friesen. "The Facial Action Coding System: A Technique for Measurement of Facial Movement" Consulting Psychologists Press, San Francisco, CA, 1978.

[21] P. Ekman. "Emotions in the Human Faces", Studies in Emotion and Social Interaction. Cambridge University Press, second edition, 1982.

[22] J. N. Bassili "Facial motion in the perception of faces and of emotional expression", Journal of Experimental Psychology, PP:373–379,1978.

[23] A. Kendon, "How gestures can become like words", Cross-cultural perspectives in nonverbal communication, New York, 1988.

[24] Affectiva Developer Portal. Available: https://developer.affectiva.com/

[25] Wilson, P. I. ,Fernandez, J., "Facial feature detection using Haar classifiers", Journal of Computing Sciences in Colleges, 21(4), pp. 127-133, April 2006.

[26] Valstar M. and Pantic M, "Fully Automatic Facial Action Unit Detection and Temporal Phases of Factial Actions", IEEETransactions on Systems, Man, and Cybernetics, 42(1), pp. 28-43, February 2012.

[27] James Jenn-Jier Lien, Kanade Takeo, Cohn Jeffrey and Ching-chang Li, "Detection, Tracking, and Classification of Action Units in Facial Expression", Journal of Robotics and Autonomous Systems, 31(3), pp. 131-146, 1999

[28] Pantic M., Tomc M., Rothkrantz L and J. M., "A hybrid approach to mouth features detection", Proceedings of IEEE InternationalConference on Systems, Man, Cybernetics, pp. 1188 - 1193, Tucson, Arizona, October 2001.

[29] Wilson P. I. and Fernandez J., "Facial feature detection using Haar classifiers", Journal of Computing Sciences in Colleges, 21(4), pp.127-133, April 2006

[30] Feijun Jiang, Ekenel H. K. and Shi B. E., "Efficient and robust integration of face detection and head pose estimation", 21st International Conference on Pattern Recognition (ICPR), pp.1578- 1581, Tsukuba, 11-15 Nov. 2012.

[31] Viola P. and Jones M., "Robust real-time face detection", IEEEInternational Conference on Computer Vision, vol. 2, pp. 747, Vancouver, 2001.

[32] Wren C., Azarbayejani A., Darrel T. and Pentland A,." P-finder: Realtime tracking of the human body" IEEE Trans. Pattern Anal. Machine Intell., 9(7), pp. 780-785, 1997

[33] Kjeldsen R. and Kender J., "Finding skin in color images", Proc. 2nd Int.Conf. Automatic Face and Gesture Recognition, pp. 312-317,1996

[34] Darrell T., Essa I. and Pentland A., "Task-Specific Gesture Analysis in Real-Time Using Interpolated Views", IEEE Trans. Pattern Analysis andMachine Intelligence, 18(12), 1,236-1,242,1996

[35] Ekman P. and Friesen W., "Unmasking the Face", New York: Prentice-Hall, 1975

[36] I. Essa and A. Pentland, "Coding, analysis, interpretation and recognition of facial expressions", Pattern Analysis and Machine Intelligence, 7:757–763, July 1997.

[37] K. Mase, "Recognition of facial expressions for optical flow",

IEICE Transactions, Special Issue on Computer Vision and its Applications, E 74(10), 1991.
[38] Y. Yacoob and L. Davis, "Computing spatio-temporal representations of human faces", Proceedings of the Computer Vision and Pattern Recognition Conference IEEE Computer Society,, pp 70–75, 1994.
[39] Y. Tian, T. Kanade, and J. F. Cohn, "Recognizing action units for facial expression analysis", Pattern Analysis and Machine Intelligence, 23(2), February 2001.
[40] A. Lanitis, C. J.Taylor, T. F. Cootes, "Automatic Interpretation and Coding of Face Images Using Flexible Models"Transactions of pattern analysis and machine learning, Vol. 19, No. 7, pp. 743-756, 1997
[41] M. A. Bartlett, J. C. Hager, P. Ekman, and T. Sejnowski, "Measuring facial expressions by computer image analysis Psychophysiology" 36(2):253–263, March 1999.
[42] M.S. Bartlett and T.J. Sejnowski, "Viewpoint Invariant Face Recognition Using Independent Component Analysis and Attractor zNetworks," Advances in Neural Information Processing Systems, 1997
[43] Rajeshree Rokade, Dharmpal Doye and Manesh Kokare, "Hand Gesture Recognition By Thinning Method", Proceedings of IEEE International conference on Digital Image Processing, March 2009, pp.284-287.
[44] Ketki Kshirsagar, Dharmpal Doye, "Comparing Key Frame Selection for One-Two Hand Gesture Using Different Methods", Int. J. Signal and Imaging Systems Engineering(2015), Vol. 8, No. 5, pp.273–285.
[45] M.K. Bhuyan, D. Ghosh and P.K. Bora, "Finite state representation of hand gesture using Key video object plane", Proc. IEEE Region 10 Asia-Pacific Conference, pp. 579-582, 2004.
[46] P. Hong, M. Turk, T.S. Huang, "Gesture Modeling and Recognition using Finite State Machines", Proceeding of the IEEE Conference on Face and Gesture Recognition, (2000) 410-415.
[47] T. Meier, K.N. Nagan, "Automatic Segmentation of Moving Objects for Video Object Plane Generation", IEEE Transaction on Circuits and Systems for Video Technology, 8 (5) (1998) 525-538.
[48] Rajeshree Rokade and Dharmpal Doye, "Spelled Sentence Recognition Using Radon Transform", IEEEScience and Information Conference (SAI), 2014 , pp. 351-354
[49] James MacLeant, Rainer Herperst, "Fast hand gesture recognition for real-time teleconferencing applications", Proceedings of IEEEICCV workshop on Recognition, Analysis and Tracking of facesand gestures in real time systems, July 2001, pp.133 140.
[50] M.K. Bhuyan, D. Ghosh and P.K. Bora "Designing of Human Computer Interactive Platform for Robotic Applications", TENCON 2005 IEEE Region 10, Nov. 2005.
[51] Ming-Hsuan Yang, Member, IEEE, Narendra Ahuja, Fellow, IEEE, and Mark Tabb, Member, IEEE, "Extraction of 2D Motion Trajectories and Its Application to Hand Gesture Recognition", IEEEtransactions on pattern analysis and machine intelligence, vol. 24, no. 8, august 2002.
[52] Ho-Joon Kim, Joseph S. Lee, and Jin-Hui Park, "Dynamic hand gesture recognition using a CNN model with 3d receptive fields", IEEE Int. Conference Neural Networks and Signal Processing Zhenjiang, China, June 8~10, 2008.
[53] Rajeshree Rokade and Dharmpal Doye, "SpelledT signT wordT recognitionT usingT keyT frame", International Journal of IETT ImageT ProcessingT , Volume:9, Issue: 5 pp. 381 - 388 , 2015
[54] P. Ekman and W. V. Friesen, "Unmasking the face: a guide to recognizing emotions from facial clues",Cambridge (Mas.): Malor Books, 2003.
[55] Koelstra, S., Pantic, M., Patras, I., "A dynamic texture based approach to recognition of facial actions and their temporal models", IEEE Transactions on Pattern Analysis and MachineIntelligence, 32(11), pp. 1940-1954, November 2010.
[56] Valstar, M., Pantic, M., "Fully Automatic Facial Action Unit Detection and Temporal Phases of Factial Actions", IEEETransactions on Systems, Man, and Cybernetics, 42(1), pp. 28-43, February 2012.
[57] Gavrilescu, M., "Proposed architecture of a fully Integrated Modular Neural Network-based Facial Emotion Recognition system based on Facial Action Coding system (FACS)", 10th InternationalConference on Communication, pp. 1-6, Bucharest, May 2014.
[58] M. S. Bartlett, G. Littlewort, M. Frank, C. Lainscsek, I. Fasel, and J. Movellan, "Fully automatic facial action recognition in spontaneous behavior," in Proc. 7th International Conference on Automatic Face and Gesture Recognition, 2006, pp. 223–230.
[59] Kanade, T., Cohn, J. F., Tian, Y., "Comprehensive database for facial expression analysis", 4th IEEE International Conference onAutomatic Face and Gesture Recognition, pp. 46-53, Grenoble, 2000.
[60] N. Aifanti, C. Papachristou, and A. Delopoulos, "The mug facial expression database," in Proc. 2010 11th International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS) , pp. 1–4, 2010,

AUTHORS PROFILE

Dr. Rajeshree Rokadehas completed her graduation (B.E.)(Electronic and Telecommunication Engineering) from Walchand Institute of Technology, solapur, shivaji university, Kolahapur and post graduation(M. Tech)(Electronics)from Shri Guru Gobind Singhji Institute of Engineering and Technology, Nanded and Ph.D (Digital Image Processing) from Shri Guru Gobind Singhji Institute of Engineering and Technology, Nanded , Swami Ramanand Teerth Marathwada University. She has more than 10 years teaching experience. Presently she is working as Assistant Professor in Electronic and Telecommunication Department, Lokmanya Tilak College of Engineering, NaviMumbai. She presented and published paper in varies reputed National, International conferences and Journals. She published a book on Digital Image rocessing. She has paper review committee member in International Journal and conferences. Her keen interest is in Image processing, Computer Vision, Communication Engineering, Digital Electronics and C, C++, MATLAB. She is Member if IETE and ISTE.

Dr. Ketki Kshirsagar, has completed her graduation (B.E.)(Electronic and Telecommunication Engineering)from Walchand Institute of Technology, solapur, shivaji university, Kolahapur and post graduation(M. Tech)(Electronics)from Shri Guru Gobind Singhji Institute of Engineering and Technology, Nanded and Ph.D (Digital Image Processing) from Shri Guru Gobind Singhji Institute of Engineering and Technology, Nanded , Swami Ramanand Teerth Marathwada University. She has more than 10 years teaching experience. Presently she is working as Associate Professor in Electronic and Telecommunication Department, Vishwakarma Institute of Information Technology, Pune. She presented and published paper in varies reputed National, International conferences and Journals. She published a book on Digital Image rocessing. She has been paper review committee member in International Journal and conferences. Her keen interest is in Image Processing, Computer Vision, Communication Engineering, Digital Electronics and C, C++, MATLAB.She is Member if IETE and ISTE.

Ms. Jayashree Sonawane has completed her post graduation(M. Tech) from Somiyya college of Engineering Mumbai. She has more than 5 years teaching experience. Presently she is working as Assistant Professor in Electronic and Telecommunication Department, Lokmanya Tilak College of Engineering, NaviMumbai. She presented and published paper in varies reputed National, International conferences and Journals. Her keen interest is in Image processing, Computer Vision, Communication Engineering.She is Member if IETE and ISTE.

Ms. Sunita Mundehas completed her post graduation(M.Tech). She has more than 15 years teaching experience. Presently she is working as Assistant Professor in Electronic and Telecommunication Department, Lokmanya Tilak College of Engineering, NaviMumbai. She presented and published paper in varies reputed National, International conferences and Journals. Her keen interest is in Image processing, Computer Vision, Communication Engineering.She is Member if ISTE.

Retrieval Number: I32870789S319/2019©BEIESP
DOI: 10.35940/ijitee.I3287.0789S319

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication