International Journal of Electronics Engineering (ISSN: 0973-7383)
Volume 10 • Issue 2 pp. 45-57 June 2018-Dec 2018 www.csjournals.com

Multiple Feature Fusion for Facial Expression Recognition in video

Varun Kumar Singhal
Research Scholar & Assistant professor in Patel College of Engineering, Bhopal

Abstract: Video based facial expression recognition has been a long standing issue and pulled in developing consideration as of late. Thekey to a fruitful facial expression recognition framework is to abuse the possibilities of varying media modalities and plan vigorous featuresto successfully portray the facial appearance and setup changes caused by facial movements. We propose an effectiveframework to address this issue in this paper. In our investigation, both visual modalities (confront pictures) and sound modalities (discourse) areutilized. Another feature descriptor called Histogram of Oriented Gradients from Three Orthogonal Planes (HOG-TOP) is proposed toextract dynamic surfaces from video successions to portray facial appearance changes. Furthermore, another viable geometric featurederived from the warp change of facial points of interest is proposed to catch facial design changes. In addition, the part ofaudio modalities on recognition is likewise investigated in our examination. We connected the multiple feature fusion to handle the video-based facialexpression recognition issue under lab-controlled condition and in the wild, individually. Examinations directed on the extendedKohn-Kanada (CK+) database and the Acted Facial Expression in Wild (AFEW) 4.0 database demonstrate that our approach is hearty indealing with video-based facial expression recognition issue under lab-controlled condition and in the wild contrasted and theother best in class techniques.
Keywords: Facial expression recognition; Multiple feature fusion; HOG-TOP; Geometric warp feature; Acoustic feature.

I. INTRODUCTION

Facial expression, as a capable nonverbal channel,plays a critical part for individuals to pass on feelings and transmit messages. Automatic facial expression recognition (AFEC) can be broadly connected in numerous fields, for example, restorative evaluation, lie location and human PC interaction [1]. AFEC has pulled in awesome enthusiasm for as far back as two decades. Be that as it may, facial expression examination is an extremely difficult errand since facial expressions caused by facial muscle developments are unpretentious and transient [2]. To catch and speak to these developments is a key issue to be tended to in facial expression analysis.Two standards of facial expressions examination are generally received in the ebb and flow innovative work.

One stream is to identify facial actions. The investigation announced in [3], [4] demonstrated that every facial expression contains a one of a kind gathering of facial action units. The Facial Action Coding System (FACS), which was first proposed by Ekman andFriesen in 1978 [5] and after that upgraded in 2002 [6], is the best known framework produced for individuals to portray facial actions. Another flood of facial expression investigation is to complete facial effect (feeling) recognition straightforwardly. Most specialists manage the recognition errand of six all inclusive feelings: cheerful, tragic, fear, sicken, irate and shock [7].Many endeavors have been made for facial expression recognition.

The procedures utilized are usually arranged into appearance based techniques and geometry based methods[8]. An appearance based technique applies feature descriptors to demonstrate facial surface changes. A geometry based strategy catches facial designs in which an arrangement of facial fiducial focuses is utilized to describe the face shape.Previous works primarily centered on static and single face picture based facial expression recognition. Recently,facial expression recognition in video has pulled in incredible intrigue. Contrasted and a static picture, a video succession can give spatial appearance as well as incorporate facial movements and went with discourse. The way to take care of the issue of video based facial expression recognition is to misuse the portrayal capacity of multi modalities(e.g. visual and sound data) and plan powerful features to adequately portray the facial appearance and setup changes caused by facial strong activities.To accomplish this objective, we propose a viable system in light of multiple feature fusion for facial expression recognition in video. We investigate the possibilities of visual modalities (confront pictures) and sound modalities (discourse) in our examination. In tending to visual modalities, we broaden the Histograms of Oriented Gradients (HOG) [9] to worldly Three Orthogonal Planes (TOP), propelled by a fleeting augmentation of Local Binary Patterns, LBP-TOP [10]. The proposed HOG-TOP is utilized to portray facial appearance changes.

We demonstrate that HOG-TOP executes and also LBP TOP for facial expression recognition. What's more, contrasted and LBP-TOP, HOG-TOP is more minimized and viable to portray facial appearance changes. Also, a viable geometric warp feature got from the warp change of facial milestones is proposed to catch facial design changes. We demonstrate that the proposed geometric warp feature is more viable contrasted and other proposed geometric features [11], [12].

International Journal of Electronics Engineering (ISSN: 0973-7383)
Volume 10 • Issue 2 pp. 45-57 June 2018-Dec 2018 www.csjournals.com

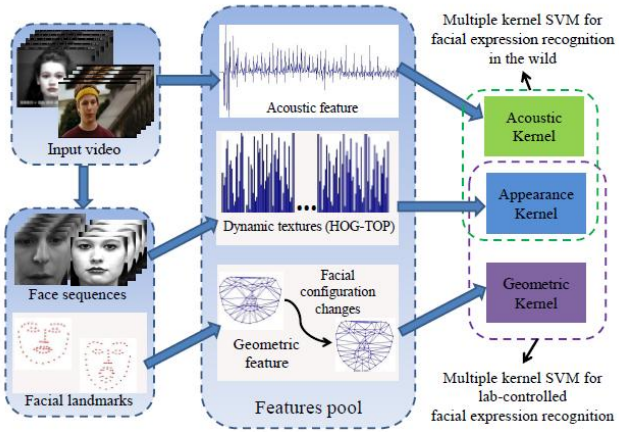

Fig. 1. Block diagram of our proposed framework. Geometric features coupled with dynamic textures (HOT-TOP) are usedto deal with lab-controlled facial expression recognition while acoustic features and dynamic textures (HOG-TOP) are fusedto tackle facial expression recognition in the wild.

We additionally investigate the part of sound modalities on influence recognition. We find that sound modalities can give some reciprocal data, particularly for facial expression recognition in nature. We additionally apply a multiple feature fusion strategy to manage facial expression recognition under lab-controlled condition and in the wild, individually. As appeared in Fig. 1, geometric features combined with dynamic surfaces (HOT-TOP) are utilized to manage lab-controlled facial expression recognition. Acoustic features and dynamic surfaces (HOG-TOP) are intertwined to handle facial expression recognition in the wild.Our commitments are abridged as takes after:

• We build up a system which can viably tacklefacial expression recognition in video. A multiple featurefusion strategy is utilized to manage facial expressionrecognition under lab-controlled condition and in thewild, separately.

• We propose another feature descriptor HOG-TOP tocharacterize facial appearance changes and another effectivegeometric feature to catch facial arrangement changes.

• We demonstrate that multiple features can improve differentcontributions and can accomplish execution thanthe singular features connected alone. We likewise demonstrate thatmultiple feature fusion can improve the discriminativepower of multiple features.

The rest of the paper is sorted out as takes after. In Section 2, we survey some related work. Segment 3 shows our proposed approach. Test results and dialogs are displayed in Section 4. The paper is finished up in Section 5.

II. RELATED WORK AND MOTIVATION

A. Static Image Based Methods

Numerous scientists apply static picture based models to handle outward appearance issue. One or a few pinnacle outlines are normally chosen for removing appearance or geometric highlights. For example, the strategies revealed in [11], [12], [13] connected the facial landmarks to portray the entire face shape. And the technique [14] estimated the relocations of a few chose candidate fiducial focuses. Bag of Words (BoW) in light of the multi-scale thick SIFT highlights were connected to speak to facial appearance surfaces in [15]. Local Fisher discriminant analysis (LFDA) was utilized for highlight extraction in [16]. The technique [17] connected Gabor channels to separate facial development highlights. A novel system for demeanor acknowledgment by utilizing appearance highlights of chosen facial patches was proposed in [18]. In any case, naturally selecting key casings from a video arrangement is typically troublesome. The techniques [19], [20] endeavored to group each casing first and receive a voting procedure to mark the video succession. It is important to remove highlights from each edge. LBP was connected in [19]. Pyramid of Histograms of Oriented Gradients (PHOG) and Local Phase Quantization (LPQ) highlights were utilized as a part of [20].

B. Dynamic Texture Based Methods

There exists a downside for a static picture based technique: separating highlights from an individual casing neglects to use dynamic data which is vital to depict facial movements. Dynamic surface based strategies can adequately manage this issue. Dynamic surface based techniques endeavor to at the same time demonstrate the spatial appearance and dynamic movements in a video arrangement. Zhao et al. [10] proposed LBP-TOP, a transient expansion of local binary patterns, for outward appearance analysis in video. A facial component LBP-TOP was proposed in [21]. A Local Gabor Binary Patterns from Three Orthogonal Planes (LGBP-TOP) was proposed in [22]. A Spatio Temporal Local Monogenic Binary Pattern (STLMBP) include descriptor was proposed in [23], [24]. What's more, Long et al. [25] utilized Independent Component Analysis (ICA) to take in spatio transient channels from recordings, and then separated dynamic surfaces utilizing the educated channels. Bite et al. [26] utilized sparsetemporal portrayal to display the fleeting elements offacial articulations in video. Li et al. [27] built up a dynamic Bayesian system to all the while and rationally speak to the facial evolvement at various levels.

C. Audiovisual Based Methods

Static picture based techniques or dynamic surface construct strategies just depend with respect to visual modalities. Notwithstanding, sound or discourse is additionally vital for individuals to pass on feelings and aims. Sound modalities can give some corresponding data notwithstanding visual modalities. As of late, varying media based techniques for influence acknowledgment have pulled in developing consideration from the full of feeling processing group. Various methodologies have been proposed to join sound and visual modalities for influence

International Journal of Electronics Engineering (ISSN: 0973-7383)
Volume 10 • Issue 2 pp. 45-57 June 2018-Dec 2018 www.csjournals.com

acknowledgment (e.g. [28], [29], [30], [31]). A far reaching study can be found in [32]. Acoustic highlights extricated from voice or discourse and visual highlights separated from confront pictures are consolidated to handle this issue. For instance, voice and lip action were utilized as a part of [33]. Face pictures and discourse were utilized in [34]. The techniques announced in [35], [36] connected a few component descriptors, for example, SIFT,HOG, PHOG and so forth to encode confront pictures and joined them with acoustic highlights to perceive outward appearance in nature.

D. Motivation

We can see that element extraction assumes a middle part on influence acknowledgment in video. Planning a compelling component is essential and significant. LBP-TOP is generally utilized for displaying dynamic surfaces. Be that as it may, there are two impediments of LBP-TOP. One is the high dimensionality. The span of LBP-TOP coded utilizing a uniform pattern is 59_3 [10]. Additionally, in spite of the fact that LBP-TOP is hearty to manage light changes, it is harsh to facial muscle misshapenings. In this work, we propose another element called HOG-TOP,which is smaller and successful to describe facial appearance changes. More subtle elements on HOG-TOP can be found in Section 3.1.In expansion, design and shape portrayals assume a critical part in human vision for the impression of outward appearances [37]. We trust that past works have not yet completely misused the possibilities of design portrayals. Describing face shape [11], [12] or estimating relocations of fiducial focuses [14], [38] just are not adequate to catch facial arrangement changes,especially the inconspicuous non-inflexible changes. In this work, we present a morehearty geometric component to catch facial setup changes. More exchange on our proposed geometric element is given in Section 3.2.

Fig. 2 The textures in XY, XT and YT planes.

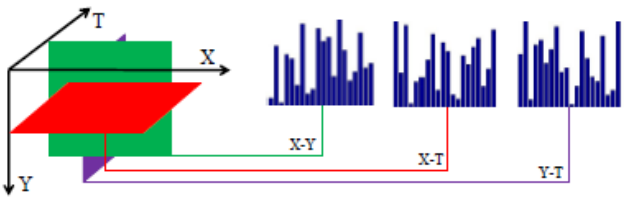

Fig. 3.The HOG from Three Orthogonal Planes (TOPs).

III. METHODOLOGY

This section presents the details of our proposed approach.We introduce the three types of features and multiple featurefusion employed in our study.

A. Histogram of Oriented Gradients from Three OrthogonalPlanes

Histograms of oriented gradients (HOG) [9] were first proposed for human discovery. The fundamental thought of HOG is thatlocal protest appearance and shape can often be described somewhat well by the appropriation of local force gradients or edge headings. HOG is touchy to question misshapenings. Outward appearances are caused by facial muscle developments. For instance, mouth opening and cocked eyebrows will produce an unexpected outward appearance. These developments could be viewed as kinds of misshapenings. HOG can successfully catch and speak to these disfigurements [39]. Be that as it may, the first HOG is restricted to manage a static image.In request to display dynamic surfaces from a video succession with HOG, we stretch out HOG to 3-D to process the oriented gradients on three orthogonal planes XY, XT, and YT (TOP),i.e. HOG-TOP. The proposed HOG-TOP is utilized to portray facial appearance changes. A video grouping incorporates three orthogonal headings, i.e. X, Y, and T (time) bearings. The XY plane gives spatial appearance, and XT and YT planes record worldly or movement data. Fig. 2 outlines the surfaces removed from the three orthogonal planes. In our investigation, we figure the disseminations of oriented gradients of each plane and acquire HOG highlights, to be specific HOG-XY, HOG-XT and HOGYT, as appeared in Fig. 3. Each point in a video succession incorporates three orthogonal neighborhoods lying on XY, XTand YT planes, separately. We initially figure the gradients along X, Y and T headings with a 3 Sobel veil. The gradient orientations are defined as $\theta_{XY} = tan^{-1}(G_Y/G_X)$, $\theta_{XT} = tan^{-1}(G_T/G_X)$, $\theta_{YT} = tan^{-1}(G_T/G_Y)$, where GX, GY, and GT are the gradients along the X, Y and Tdirections, respectively. These angles are quantized into K(K is 9 in our work) orientation bins with a range of $0°-360°$ or $0° - 180°$.We enumerate the appearance of these gradient orientationsand obtain a histogram in each plane. The threehistograms are concatenated to form a global descriptionwith the spatial and temporal features. Fig. 3 shows that thethree histograms from the three planes are combined into asingle one. The HOG-TOP computation algorithm is shownin Algorithm 1.

International Journal of Electronics Engineering (ISSN: 0973-7383)
Volume 10 • Issue 2 pp. 45-57 June 2018-Dec 2018 www.csjournals.com

Algorithm 1 Compute the HOG-TOP.

Input: Video sequence V, which contains N frames with the same width and height.
Output: The histograms of oriented gradients from three orthogonal plans (HOG-TOP).
Algorithm:
Get the number of frames N, frame width W and height H.
for $t = 2 : N - 1$ **do**
  for $x = 2 : W - 1$ **do**
    for $y = 2 : H - 1$ **do**
      get the local patch in XY, XT, and YT planes.
      $P_{xy} = V(x-1:x+1, y-1:y+1, t)$;
      $P_{xt} = V(x-1:x+1, y, t-1:t+1)$;
      $P_{yt} = V(x, y-1:y+1, t-1:t+1)$;
      Compute the gradients G_X, G_Y, and G_T; and gradient orientations θ_{XY}, θ_{XT}, θ_{YT}. Quantize the θ_{XY}, θ_{XT}, θ_{YT} into one of 9 bins. Get a histogram in each plan, i.e. HOG-XY, HOG-XT and HOG-YT.
    end for
  end for
end for
Normalize the HOG-XY, HOG-XT and HOG-YT respectively. Concatenate the three histograms into a long histogram.

LBP-TOP processes the distinction of a pixel as for its neighborhood, influencing LBP-To top hearty in managing enlightenment changes. HOG-TOP figures the oriented gradients of a pixel, which is more delicate to protested developments [9]. Outward appearances are caused by facial muscle developments, which can be viewed as kinds of muscle misshapenings. HOG-TOP is subsequently more successful to portray facial appearance changes than LBP-TOP. Another favorable position of HOG-TOP is the component dimensionality. Contrasted and LBP-TOP, the measure of HOG-TOP ismuch littler than that of LBP-TOP. The extent of LBP-TOPcoded utilizing a uniform pattern is 59 3 [10], [40]. and thesize of HOG-TOP quantized into 9 receptacles is 9 3, which ismuch more minimized than that of LBP-TOP.In request to use local spatial data, a square basedmethod is presented in our examination, as appeared in Fig. 4. Wecan isolate the picture succession into numerous pieces and extractthe HOG-TOP highlights from each square. The HOG-TOPfeatures of the considerable number of squares can be linked to representthe entire succession. In our analyses, the face is firstcropped from the first picture and resized to 128. We parcel the face picture into 8 hinders with each blockhaving a size of 16.

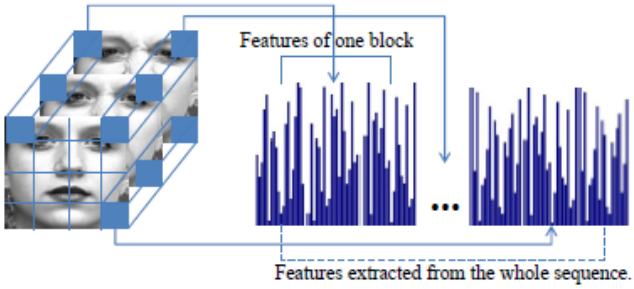

Fig.4.The HOG-TOP features extracted from each block areconcatenated together to represent the whole sequence.

The quantity of containers is set to 9 withan angle range of $0° - 180°$.

B. Geometric Warp Feature

In this segment, we present a more hearty geometric feature namely geometric twist include, which is gotten from the warp change of the facial landmarks. Facial expressions are caused by facial muscle developments. These movements result in the relocations of the facial landmarks. Here we assume that each face picture comprises of numerous sub-regions.These sub-districts can be framed with triangles with theirvertexes situated at facial landmarks, as appeared in Fig. 5. Thedisplacements of facial landmarks cause the deformationsof the triangles. We propose to use the distortions torepresent facial arrangement changes.Facial demeanor can be considered as a dynamic processincluding beginning, pinnacle and offset. We consider the displacementof the comparing facial landmarks between onset(neutral face) and pinnacle (expressive face). Given a set offaciallandmarks $s = (x_1, y_1, x_2, y_2, ...x_n, y_n)$, where (x_i, y_i)denote the coordinates of the i-th facial landmark. Thesefacial landmarks make up the mesh of a face, as shown inFig. 5.

As should be obvious, there are numerous little triangles in the face, and every triangle is controlled by three facial points of interest. Facial muscle developments cause the disfigurements of the triangles when an impartial face changes to an expressive face. We think about a pixel (x; y) which lies in a triangle $\Delta A'B'C'$ having a place with the impartial face and the relating pixel (u; v) lies in a triangle having a place with the expressive face, as appeared in Fig. 6. From [41], we realize that the pixel (x; y) can be communicated with a direct combinationof the three vertexes.

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} + \lambda_1 \begin{bmatrix} x_2 - x_1 \\ y_2 - y_1 \end{bmatrix} + \lambda_2 \begin{bmatrix} x_3 - x_1 \\ y_3 - y_1 \end{bmatrix} \quad \textbf{(1)}$$

And the coefficients _1; _2 can be obtained as

$$\lambda_1 = \frac{(x - x_1)(y_3 - y_1) - (y - y_1)(x_3 - x_1)}{(x_2 - x_1)(y_3 - y_1) - (y_2 - y_1)(x_3 - x_1)} \quad \textbf{(2)}$$

$$\lambda_2 = \frac{(x_2 - x_1)(y - y_1) - (y_2 - y_1)(x - x_1)}{(x_2 - x_1)(y_3 - y_1) - (y_2 - y_1)(x_3 - x_1)} \quad \textbf{(3)}$$

The point (u; v) in the triangle $\Delta A'B'C'$ of the expressiveface can be defined with the three vertexes and $\lambda_1, \lambda_2,$

International Journal of Electronics Engineering (ISSN: 0973-7383)
Volume 10 • Issue 2 pp. 45-57 June 2018-Dec 2018 www.csjournals.com

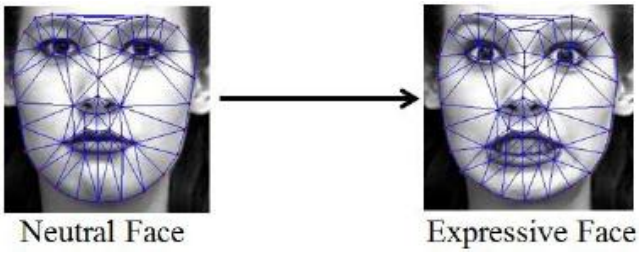

Fig. 5.Facial landmarks describe the shape of a face.

Fig. 6. A pixel (x; y) in a triangle ΔABC of the neutral faceis transformed to another pixel (u; v) in a triangle $\Delta A'B'C'$ of the expressive face.

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} u_1 \\ v_1 \end{bmatrix} + \lambda_1 \begin{bmatrix} u_2 - u_1 \\ v_2 - v_1 \end{bmatrix} + \lambda_2 \begin{bmatrix} u_3 - u_1 \\ v_3 - v_1 \end{bmatrix} \quad \textbf{(4)}$$

Combining Eq. (2) with Eq. (3), Eq. (4) can be rewrittenas:

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} a_1 + a_2 x + a_3 y \\ a_4 + a_5 x + a_6 y \end{bmatrix} \quad \textbf{(5)}$$

Each combine of triangles between the unbiased face and theexpressive face can characterize a remarkable change and eachaffine change is dictated by 6 parameters a1, a2,… a6.We process the 6 parameters for each twist change and link every one of the parameters as a long worldwide component vector, which is utilized to portray facial design changes. We will appear by tests that the proposed geometric twist highlight is more successful than the other geometric highlights [11], [14], [38].

C. Acoustic Feature

Visual modalities (confront pictures) and sound modalities(speech) can both pass on the feelings and aims ofhuman creatures. Sound modalities additionally give some helpful intimations to influence acknowledgment in video. For example, with voicesignal, the strategy [42] proposed an improved autocorrelation (EAC) highlight for feeling acknowledgment in video.One fruitful acoustic component extraction is to obtainthe time arrangement of various paralinguistic descriptors andthen utilizing pooling activities on each time arrangement to extractfeature vectors. Schuller et al. [43] indicated how tocompute the acoustic highlights by taking 21 functionals of 38low level descriptors and their first relapse coefficients.The 38 low-level descriptors appeared in Table 1 are firstextracted and smoothed by straightforward moving normal lowpassfiltering. From that point forward, 21 functionals are utilized and 16zero-data highlights are dispensed with. At long last, two singlefeatures: the quantity of onsets (F0) and turn span areadded. A sum of 1,582 acoustic highlights are extricated from

TABLE I: Acoustic Features: 38 Low Level Descriptors Along With Their First Regression Coefficients And 21 Functionals [43].

Descriptors	Functionals
PCM loudness	Position max./min.
MFCC (0-14)	Arithmetic Mean
log Mel Freq. Band (0-7)	skewness, kurtosis
LSP Frequency (0-7)	lin. regression coeff.
F0	lin. regression error
F0	Envelope quartile
Voicing Prob.	quartile range
Jitter local	percentile
Jitter consec. frame pairs	percentile range
Shimmer local	up-level time

every video. These acoustic highlights incorporate vitality/phantom Low Level Descriptors (LLD) (top 6 things in Table 1) and voice related LLD (base 4 things in Table 1).We investigate the portrayal capacity of acoustic featuresfor influence acknowledgment in our examination. Tests demonstrate thataudio modalities (discourse) can give valuable complementaryinformation notwithstanding visual modalities. The visualfeatures combined with acoustic highlights can accomplish betterperformance for outward appearance acknowledgment in nature.

D. Multiple Feature Fusion

Highlights from various modalities can make distinctive contributions.Traditional SVM connects diverse featuresinto a solitary component vector and constructed a solitary portion for allthese distinctive highlights. In any case, developing a portion foreach kind of highlights and incorporating these bits optimallycan upgrade the discriminative energy of these features.The consider in [44] demonstrated that utilizing numerous pieces withdifferent sorts of highlights can enhance the execution ofSVM. A numerous portion SVM is intended to learn both thedecision limits between information from various classes andthe bit mix weights through a solitary optimizationproblem [45].

Given a training set with labeled samples $D = \{(\mathbf{x}_i, y_i) | \mathbf{x}_i \in \mathbf{R}^n, y_i \in \{-1, 1\}\}_{i=1}^N$ A decision line isobtained by solving the following primal optimization problem,

$$\min_{\mathbf{w},b} \frac{1}{2} \|\mathbf{w}\|^2$$
$$s.t. \quad y_i(\mathbf{w}\mathbf{x}_i + b) \geq 1, i = 1, 2, ...N \quad \textbf{(6)}$$

In general, we solve the dual form of the primal optimizationproblem. The dual formulation of the

International Journal of Electronics Engineering (ISSN: 0973-7383)
Volume 10 • Issue 2 pp. 45-57 June 2018-Dec 2018 www.csjournals.com

traditionalsingle kernel SVM optimization problem is given by

$$\max_{\alpha} \left[\sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j K_{ij} \right]$$
$$s.t. \quad \sum_{i=1}^{N} \alpha_i y_i = 0, 0 \le \alpha_i \le C \qquad (7)$$

whereKij is the kernel matrix, and $K_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$,here $k(\cdot,\cdot)$ is the kernel function and $\mathbf{x}_i, \mathbf{x}_j$ are the feature vectors.

Multiple kernel fusion applies a linear combination ofmultiple kernels to substitute for the single kernel. In ourstudy, we adopt the formulation proposed in [46] in whichthe kernel is actually a convex combination of basis kernels:

$$K_{ij} = \sum_{i=1}^{M} \beta_m k_m(\mathbf{x}_i, \mathbf{x}_j)$$
$$s.t. \quad \beta_m \ge 0, \quad \sum_{i=1}^{M} \beta_m = 1 \qquad (8)$$

We apply a various part combination structure to manage outward appearance acknowledgment under lab-controlled condition and in the wild, individually, as appeared in Fig. 1.HOG-TOP and acoustic component are ideally intertwined to deal with the issue of outward appearance acknowledgment in the wild, while HOG-TOP and geometric twist include are joined to handle the issue of outward appearance acknowledgment under lab-controlled condition. In the followings, we detail how to locate an ideal mix of HOG-TOP and acoustic component for outward appearance acknowledgment in nature. It can be effectively reached out to the issue of outward appearance acknowledgment under lab controlled condition. We mean the dynamic surface HOG-TOP as x and acoustic element as z, at that point we have

$$K_{ij} = \beta k_1(\mathbf{x}_i, \mathbf{x}_j) + (1-\beta) k_2(\mathbf{z}_i, \mathbf{z}_j) \qquad (9)$$

with $0 \le \beta \le 1$, whereK is the kernel matrix, $k_1(\cdot,\cdot), k_2(\cdot,\cdot)$ are the basis kernels. The basis kernels could be linearkernel, radial basis function (RBF) kernel and polynomialkernel, etc. We need to learn the kernel weight α, andcoefficients α. In our study, we construct a linear kernel foreach type of feature and build a two-step method to searchfor the optimal values of β and α.We set two nested iterativeloops to optimize both the classifier and kernel combinationweights. In the outer loop, we adopt the grid search to findthe kernel weight β. In the inner iteration, a solver of SVM(LIBSVM [47] is used in our work) is implemented by fixingthe kernel weight β to find the coefficients α. Then given anew sample which contains visual feature HOG-TOP x andacoustic feature z, the predict label y can be obtained by

$$y = sgn(\sum_{i=1}^{N} y_i \alpha_i (\beta k_1(\mathbf{x}_i, \mathbf{x}) + (1-\beta) k_2(\mathbf{z}_i, \mathbf{z})) + b) \qquad (10)$$

In our work, the one-versus one technique is utilized to dealwith the multiclass-SVM issue and we receive the maxwinvoting methodology to do the characterization. At last, the esteem and qualities with the most elevated general classificationaccuracy in the approval informational index are gotten as theoptimal portion weight and coefficients.

IV. EXPERIMENTAL RESULTS AND DISCUSSIONS

A. Data sets

Keeping in mind the end goal to assess our strategies, we direct the experiments on three open informational indexes: the Extended Cohn-Kanade(CK+) informational collection [11], GEMEP-FERA 2011 informational collection [19] and the Acted Facial Expression in Wild (AFEW) 4.0 informational index [48].We first give a short depiction of the three informational collections. The Extended Cohn-Kanade (CK+) informational collection contains 593image groupings from 123 subjects. The face pictures in thedatabase are lab-controlled. The picture successions change induration from 10 to 60 outlines. Altogether, 327 of 593 imagesequences have feeling marks and each is sorted intoone of the accompanying seven feeling classes: outrage (A), contempt(Co), sicken (Di), fear (Fe), satisfaction (Ha), sadness(Sa) and astonishment (Su). Each picture arrangement changes fromthe beginning (the nonpartisan edge) to the pinnacle (the expressiveframe). Furthermore, the X-Y directions of 68 facial landmarkpoints were given for each picture in the database. Thelandmark purposes of key edges inside every video sequencewere physically marked, while the rest of the edges wereautomatically adjusted utilizing the AAM fitting calculation [41]. The GEMEP-FERA 2011 informational index contains 289 sequencesof 10 performing artists, who are prepared by an expert executive. Itis separated into a preparation set of 155 successions and a testset of 134 groupings. Each succession is classified into thefollowing five feelings: outrage (A), fear (Fe), happiness(Ha), help (Re) and trouble (Sa). Just the preparation setprovides feeling marks. This database is more challengingthan the CK+ database, since there are head developments andgesture varieties in picture groupings.

The Acted Facial Expression in Wild (AFEW) 4.0 dataset incorporates video cuts gathered from various motion pictures

Which are accepted to be near genuine conditions. The database parts into a preparation set, an approval set and a test set. There are 578 video cuts in the preparation set. The approval and test sets have 383 video cuts and407 video cuts, separately. Every video cut has a place with one of the seven classes: outrage (An), appall (Di), fear(Fe), joy (Ha), unbiased (Ne), pity (Sa), and surprise(Su). This database gives unique video cuts and adjusted face successions. They connected the model proposed in [49] to extricate the

International Journal of Electronics Engineering (ISSN: 0973-7383)
Volume 10 • Issue 2 pp. 45-57 June 2018-Dec 2018 www.csjournals.com

countenances from video cuts and adjusted the appearances. Unique in relation to the CK+ and GEMEP-FERA 2011data sets, outward appearances in AFEW 4.0 are more normal and unconstrained. The varieties in brightening, posture and foundation in picture groupings increment the unpredictability of outward appearance investigation. Fig. 7 demonstrates the chose picture arrangements from the three databases. The primary line is the face pictures from the CK+ database, which are frontal-view and lab-controlled appearances. The center line demonstrates the pictures from the GEMP-FERA2011 database; there exist head developments and signal varieties. The base line is a picture succession from AFEW4.0 database. We can see that the foundation is unpredictable and there exist light changes and stance varieties.

B. Feature Extraction

In our investigations, three sorts of highlights are employed,namely HOG-TOP, geometric twist highlight and acousticfeature.In separating HOG-TOP from picture arrangements, each faceimage is first edited and resized to 128. The resizedface picture is then apportioned into 8 obstructs with a sizeof 16. The container number is set to 9 with an edge extend

Fig. 7. The selected image sequences from the three databases. From top to bottom: CK+, GEMEP-FERA2011 and AFEW4.0.

of $0° - 180°$. In each block, we can obtain a HOG-TOP with a dimension of $3 \times 9 = 27$. We then concatenate the HOGTOPof the 8×8 blocks into a long feature vector with adimension of $3 \times 9 \times 8 \times 8 = 1728$.

Facial points of interest are utilized to process the geometricwarp highlights. We register the twist change of faciallandmarks between the nonpartisan face and an expressive face.Each confront contains 68 facial milestones. These facial landmarksdivide the face into numerous non-cover sub regionsby Delaunay triangulation. In our work, we take 109 pairof triangles (the most modest number of triangles accessible inface pictures). Each combine of triangles between the neutralface and an expressive face can characterize a one of a kind transformand every relative change is controlled by six parameters(see Section 3.2). The twist change coefficients are finallyconcatenated as a component vector of 6 109 = 654 elementsto speak to the geometric twist highlight. The acoustic highlights with a length of 1582 utilized as a part of ourwork are given by the database [34], [48]. The acousticfeatures are separated by applying the open-source EmotionAffect Recognition (openEAR) toolbox [50] backend OpenSMILE[51].

C. Experimental Results

A Comparison of HOG-TOP and LBP-TOP: We first look at the execution of HOG-TOP proposedin our work with LBP-TOP proposed in [10]. At the point when wecompute the LBP-TOP highlights, we take the general settingsadopted in most detailed works. The resized confront picture ispartitioned into 4 pieces. The LBP-TOP is coded with auniform design. The LBP-TOP histogram of each square is afeature vector of 3 59 = 177 components. The length of thefeature vector comprises of 4 pieces is 3 59 4 = 2832.There are 327 picture groupings with feeling names belongingto 118 subjects in the CK+ database. We take after theprotocol proposed in [11] and withdraw one-subject-outcross approval system. Each time the examples from onesubject are utilized for testing and the rest of the examples fromall different subjects are utilized for preparing. With the end goal for eachsubject to be assessed once, we complete 118 validations.The order precision procured on the CK+ database byusing two kinds of highlights is indicated Table 2.We additionally look at the execution of the two highlights inthe GEMEP-FERA 2011 database. Since just the feeling.

TABLE II: The classification accuracy of LBP-TOP and HOGTOPon the CK+ database (%).

	LBP-TOP	HOG-TOP
Anger	75.6	88.9
Contempt	88.9	66.7
Disgust	93.2	94.9
Fear	80.0	76.0
Happiness	98.5	95.6
Sadness	78.6	67.9
Surprise	92.8	97.6
Overall	89.3	89.6

TABLE III: The classification accuracy of LBP-TOP and HOGTOPon the GEMEP-FERA 2011 database (%)

	LBP-TOP	HOG-TOP
Anger	56.2	43.7
Fear	26.7	36.7
Joy	58.1	61.3
Relief	51.6	54.8
Sad	74.2	74.2
Overall	53.6	54.2

marks of preparing set are openly accessible, we do theevaluation on the preparation set. There are seven subjects inthe preparing set. We receive the forget one-subject strategyand complete seven cross approvals. Table 3 indicates theperformance acquired by applying the two features.As for the AFEW 4.0 database, we use the trainingset to prepare a SVM classifier and test the classifier on thevalidation set. The database gave a pattern

International Journal of Electronics Engineering (ISSN: 0973-7383)
Volume 10 • Issue 2 pp. 45-57 June 2018-Dec 2018 www.csjournals.com

strategy [34]which utilized LBP-TOP to speak to the dynamic texturesof the video succession and prepared a SVM with nonlinearRBF portion for feeling grouping. The accuracyacquired on the AFEW 4.0 database by applying two typesof highlights is appeared in Table 4.We utilize the general precision to assess the execution.

TABLE IV: The Classification Accuracy of LBP-TOP and HOGTOPon Validation set of the AFEW 4.0 Database (%)

	LBP-TOP	HOG-TOP
Neutral	19.0	58.7
Anger	50.0	73.4
Disgust	25.0	22.5
Fear	15.2	4.3
Happiness	57.1	60.3
Sadness	16.4	4.9
Surprise	21.7	2.2
Overall	30.6	35.8

The overall accuracy is defined as

$$O_{acc} = \frac{\sum_{n=1}^{N}\sum_{k=1}^{K} m_{nk}}{\sum_{n=1}^{N}\sum_{k=1}^{K} M_{nk}} \quad (11)$$

where K is the quantity of classes, N is the quantity of crossvalidation folds, mnk is the quantity of effectively predictedsamples of the k-th class in the n-th overlay, and Mnk denotesthe add up to tests of the k-th class in the n-th crease. Theclassification rate of every individual outward appearance (k-thclass) is $\frac{\sum_{n=1}^{N} m_{nk}}{\sum_{n=1}^{N} M_{nk}}$.

From the exploratory outcomes, we can see that the general arrangement precision got by utilizing HOG-TOP onthe CK+ database and GEMEP-FERA 2011 database is 89.6%and 54.2%, separately. It is focused with the outcome of89.3% and 53.6% got by applying LBP-TOP on the two databases. While the general grouping rate of HOG-TOPon the AFEW 4.0 database is 35.8%, which is better than30.6% got by utilizing LBP-TOP, implying that HOG-TOPis more strong in catching the unobtrusive facial appearance changes in nature. What's more, HOG-TOP with a length of1728 is more minimal than LBP-TOP with a length of 2832.We further look at the certainty interims (the variancesacross cross-approval folds) of two capabilities. Since eachfold in the CK+ database contains a few examples only(118 folds (subjects) all together), the fluctuation crosswise over crossvalidation folds is vast and along these lines it isn't verymeaningful to report the differences for this informational index. On the other hand, the preparation set and approval set are settled inthe AFEW 4.0 informational index and hence no fluctuation cross the folds for this informational collection can be accounted for. We just compare the differences of the two capabilities on the GEMEP-FERA2011 database. The differences of HOG-TOP and LBP-TOP are12.7% and 12.6%, individually, which are practically identical with each other. We likewise think about the computational rates ofthe two highlights under the 64-bit Win 7 working system with a Core i7 CPU. We processed the two highlights withMatlab 8.2. The calculation time relies upon the piece sizeand succession span. With a similar piece measure (16 16)and succession span (11 outlines), the calculation time ofHOG-TOP and LBP-TOP is 0.042s, respectively,showing the computational productivity of HOG-TOP.

TABLE V: The Comparison Results Of Different Geometric Features On The CK+ Database (%). (GWF Is Our Proposed Geometric Warp Feature)

	GWF	[11]	[14]	[38]
Anger	86.7	35.0	75.6	62.2
Contempt	94.4	25.0	–	72.2
Disgust	96.6	68.4	88.2	86.4
Fear	36.0	21.7	76.0	56.0
Happiness	98.5	98.4	97.1	91.3
Sadness	75.0	4.0	89.3	39.3
Surprise	96.4	100.0	98.7	95.2
Overall	89.0	66.7	87.5	79.2

TABLE VI: The Classification Accuracy Obtained By Using Four Different Feature Sets On The Ck+ Database (%)

	HOG-TOP	Geometric Feature	Hybrid Feature I	Hybrid Feature II
Anger	88.9	86.7	95.6	100.0
Contempt	66.7	94.4	94.4	94.4
Disgust	94.9	96.6	94.9	96.6
Fear	76.0	36.0	52.0	84.0
Happiness	95.6	98.5	98.5	100.0
Sadness	67.9	75.0	78.6	78.6
Surprise	97.6	96.4	96.4	98.8
Overall	89.6	89.0	91.4	95.7

Facial Expression Recognition Under Lab-controlledEnvironment: We build up a model which joins HOG-TOP and geometric twist to deal with the issue of outward appearance acknowledgment under lab-controlled condition. We assess the accompanying distinctive capabilities: geometric twist include, dynamic appearance highlight (HOG-TOP), half breed highlight I and cross breed highlight II. Mixture highlight I indicates the component vector of linking HOG-TOP and geometric twist include straightforwardly and half breed include II is the ideal mix of the HOG-TOP and geometric twist highlight. We first contrast our proposed geometric twist

International Journal of Electronics Engineering (ISSN: 0973-7383)
Volume 10 • Issue 2 pp. 45-57 June 2018-Dec 2018 www.csjournals.com

include and the other geometric highlights on the CK+ informational collection. Every one of the strategies disappear one-subject-out cross approval. The correlation comes about are appeared in Table 5. The technique [11] connected an arrangement of facial points of interest to portray the face shape. The relative separation of eight chose fiducial focuses is estimated to speak to the geometric component in [14]. The movements of the facial points of interest between the nonpartisan face and the expressive face are processed to speak to the geometric element in [38]. We can see that our proposed geometric twist include accomplishes an unrivaled execution contrasted and the other geometric highlights, implying that the geometric twist include is more powerful to catch facial arrangement changes.

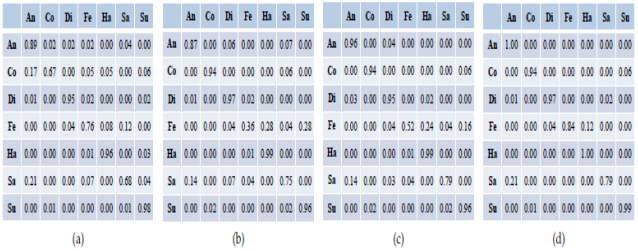

Fig. 8. The confusion matrices obtained by using the four feature sets on the CK+ database: (a) HOG-TOP, (b) geometricfeature, (c) hybrid feature I and (d) hybrid feature II. (An: Anger, Co: Contempt, Di: Disgust, Fe: Fear, Ha: Happiness, Sa:Sadness and Su: Surprise).

We additionally assess half breed include I and cross breed highlight

II with the forget one-subject cross approval on the CK+database and contrast the execution and that obtainedby applying geometric element and HOG-TOP alone. Table 6shows the arrangement exactness acquired by the four differentfeature sets. Fig. 8 demonstrates the disarray lattices of usingthe four distinctive capabilities. We can see that the emotions"disgust", "satisfaction" and "amazement" have higher classification rates than alternate feelings, showing that these three feelings are simpler to recognize than the others. Weal so take note of that cross breed include I (91.4%) and half breed highlight II(95.7%) beat the geometric twist include (89.0%) andHOG-TOP (89.6%) connected independently. We can conclude that distinctive highlights (cross breed include I) can give complementary information and numerous element combination (hybrid feature II) can additionally improve the discriminative capacity ofthe consolidated highlights .

We additionally contrast our strategy and the other methods. All the strategies we thought about take after the baseline method [11] and disappear one-subject-out cross validation. The techniques [11], [12] joined geometric feature and appearance include and prepared a SVM to perform the classification. In [21], a weighted segment based feature descriptor to extricate dynamic appearance include was utilized and numerous piece learning was connected to train the SVM for acknowledgment. A meager worldly representation classifier was proposed for outward appearance recognition in [26].

The strategy in [24] connected spatiotemporal localmonogenic twofold example (STLMBP) highlight to deal with the problem of outward appearance recognition.As can be found in Table 7, the HOG-TOP (89.6%) and geometricfeature (89.3%) proposed in our technique can achievea focused execution contrasted and SPTS+CAPP [11](88.38%), CLM [12] (82.4%) and STLMBP [24] (88.4%). It demonstrates the adequacy of our proposed highlights.

The half breed include II as the ideal blend of HOGTOPand geometric component accomplishes an unrivaled performancecompared with alternate techniques tried, demonstrating the effectivenessof the numerous element combination.

Facial Expression Recognition in the Wild:HOG-TOP and acoustic feature are fused to tackle theproblem of facial expression recognition in the wild.We first

TABLE VII: Performance Comparison With Other Methods On CK+ Database

Method	Accuracy (%)
HOG-TOP	89.6
Geometric Feature	89.3
Hybrid Feature II	**95.7**
SPTS+CAPP [11]	88.4
CLM [12]	82.4
STLMBP [24]	88.4
STR [26]	94.9
CFD [21]	93.2

TABLE VIII: The Classification Accuracy Obtained By Using Four Feature Sets On Validation Set Of The AFEW 4.0 Database (%)

	HOG-TOP	Acoustic Feature	Hybrid Feature I	Hybrid Feature II
Neutral	58.7	57.1	65.1	69.8
Anger	73.4	64.1	75.0	76.6
Disgust	22.5	15.0	12.5	17.5
Fear	4.3	26.1	8.7	15.2
Happiness	60.3	34.9	57.1	63.5
Sadness	4.9	14.7	13.1	9.8
Surprise	2.1	0.0	4.4	2.1
Overall	35.8	32.9	37.6	40.2

assess our strategy on the approval set. Four capabilities assess our strategy on the approval set. Four capabilities are

International Journal of Electronics Engineering (ISSN: 0973-7383)
Volume 10 • Issue 2 pp. 45-57 June 2018-Dec 2018 www.csjournals.com

investigated: HOG-TOP just, acoustic element just, cross breed highlight I and cross breed include II. Half and half component I concatenates the HOG-TOP and acoustic element straightforwardly. Half and half featureII is the ideal mix of the HOG-TOP and acousticfeature.Table 8 demonstrates the arrangement exactness got byapplying four distinctive capabilities. The relating confusionmatrices are appeared in Fig. 9. We can see that the

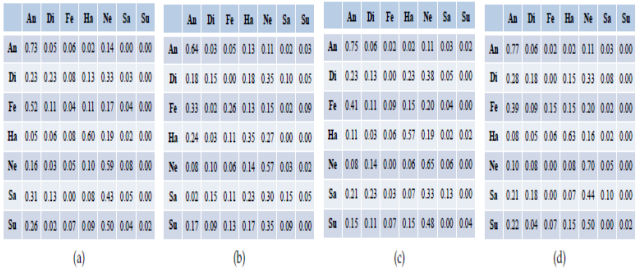

Fig. 9. The confusion matrices obtained by using the four feature sets on the validation set of AFEW 4.0 database: (a) HOGTOP,(b) acoustic feature, (c) hybrid feature I and (d) hybrid feature II. (An: Anger, Di: Disgust, Fe: Fear, Ha: Happiness,Ne: Neutral, Sa: Sadness and Su: Surprise).

TABLE IX: Performance Comparison With Other Methods On The Test Set Of The AFEW 4.0 Database

Method	Accuracy (%)
Hybrid Feature II (our method)	**45.2**
LBP-TOP + Voice [34]	33.7
Lip activity + Voice [33]	35.3
STLMBP [23]	41.5
EAC [42]	40.1
ELM [52]	44.2

arrangement rates are much lower than the outcomes shownin Table 6. Unique in relation to outward appearances under labcontrolledenvironment in which the on-screen characters or subjects canpose recognized outward appearances, the facial expressionsin the wild might be more inconspicuous. The components including headmovements, posture varieties and so forth additionally increment classificationdifficulties. What's more, in some cases, a few outward appearances in thewild may seem together, which makes a facial expressionto be mistaken for different articulations.

We watch thatthe order rate of feeling "astonish" is the lowest.From the disarray networks appeared in Fig. 9, we findthe feeling "amaze" is generally misclassified as emotions"anger", "satisfaction" and "nonpartisan". The feelings "anger"and "impartial" have higher acknowledgment correctnesses than theother feelings. Cross breed include I and mixture highlight IIoutperform the HOT-TOP and acoustic component utilized individually,indicating that two capabilities are complementarywith each other. Half breed highlight II accomplishes a predominant performancecompared with mixture include I, exhibiting thatthe adequacy of the various component combination in dealingwith the outward appearance acknowledgment issue in the wild.We additionally apply cross breed highlight II which accomplishes thebest execution on the approval set (the piece weightsof HOG-TOP and acoustic element are 0.73 and 0.27) toevaluate the test set. The general acknowledgment exactness onthe test set is 45.2%. Table 9 demonstrates the outcomes comparedwith alternate strategies. The gauge technique [34] combinedLBP-TOP and the acoustic component. Lip movement was incorporatedwith voice in [33] to handle the feeling acknowledgment

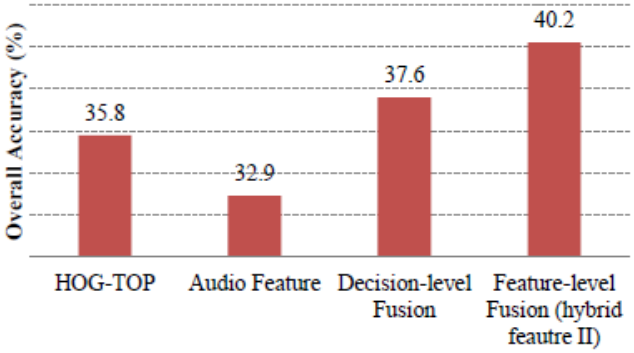

Fig. 10. The comparison results of different methods on thevalidation set of AFEW 4.0 database.

issue. The strategy [23] utilized dynamic surfaces just and

the strategy [42] connected the voice as it were. The technique [35]employed varying media include for feeling acknowledgment. Wecan see that our strategy (45.2%) enhances altogether comparedwith the pattern technique [34] and the strategy [33],with a change of around 11% and 10%, respectively.Our technique is likewise superior to [23] (41.5%) and EAC [42](40.1%). Contrasted and the strategy [52] (44.2%), our performanceis still aggressive. In addition, we connected ourmethod to partake in the second feeling recognitionin the wild test (EmotiW 2014) [34] and accomplished thesecond sprinter up grant.

Decision-Level Fusion VsFeature-Level Fusion: The different element combination connected in our work is a kindof highlight level combination technique. Another procedure, namelydecision-level combination, is additionally generally utilized as a part of PC visioncommunity to manage numerous arrangements of highlights. In apreliminary think about, we investigate the viability of the twotechniques for outward appearance acknowledgment in video. A preliminaryexperiment is directed on AFEW 4.0 database.As we have said above, we utilize the one-versus onetechnique to handle the multiclass-SVM issue, and usemax-win voting technique to lead the order. Fordecision-level combination, we initially apply the HOG-TOP and acousticfeature independently and after that spared the anticipate comes about,

International Journal of Electronics Engineering (ISSN: 0973-7383)
Volume 10 • Issue 2 pp. 45-57 June 2018-Dec 2018 www.csjournals.com

TABLE X: The Parameters Used For Extracting HOG-TOP

Image Size	Block Size	Overlap	Blocks
96 × 96	12 × 12	No	8 × 8
96 × 96	24 × 24	Half	7 × 7
128 × 128	16 × 16	No	8 × 8
128 × 128	32 × 32	Half	7 × 7

i.e. the quantity of votes in favor of each class of the two features,respectively. From that point forward, we include the votes got by eachindividual highlight together and in view of the combinedvotes, we complete the maximum win voting system againto settle on an official choice. The general arrangement rateis processed as the execution of choice level combination

method.Fig. 10 demonstrates the consequences of the distinctive strategies tested.The general precision of HOG-TOP, acoustic element andfeature-level combination (crossover include II) appeared in Fig. 10 is thesame as appeared in Table 8. From the test comes about, wefind that element level combination outflanks the choice levelfusion strategy, in spite of the fact that they both use the same multiplesets of highlights. We likewise find that the change acquiredby choice level combination over individual highlights sets is notas huge as that accomplished by include level combination.

The Effect of Block Size on HOG-TOP: We additionally investigate the portrayal capacity of HOG-TOPwith diverse piece sizes, from 12 to 32. Table10 demonstrates the parameters utilized for separating HOG-TOP. Theblocks with a little size (12 and 16) are not covered andlarge squares (24 and 32) are half covered. Weemploy HOG-TOP on the CK+ database and the AFEW 4.0database. Test comes about are appeared in Tables 11 and 12.We can see that the HOG-TOP with different piece sizes achieve the comparable general exactness. We can accordingly conclude that HOG-TOP is hearty to scales. We can further examine the test comes about. For facial expressions under lab-controlled condition (Table 11), HOG-TOP with a little size (12) is more powerful to perceive the facial articulations "dread" and "hatred" which have subtle facial muscle exercises. A little piece estimate is more strong to capture nearby unobtrusive appearance changes than a substantial block size. For outward appearances in the wild (Table 12), HOG-TOPwith an extensive size (24 and 32) accomplishes a prevalent performancefor outward appearance "shock", showing that HOG-TOPwith a huge piece measure is more vigorous to recognize thisexpression from others in nature. Table 12 likewise demonstrates thatHOG-TOP with different piece sizes outflanks the LBPTOP(30.6%) for outward appearance acknowledgment in nature.

D. Discussion

From the trial comes about announced above, we can seethat our proposed structure can effectively deal with theproblem of outward appearance acknowledgment in video. Outward appearances under lab-controlled condition are differentfrom those in the wild which are more common and spontaneous.We propose two ways to deal with handle the twodifferent outward appearance acknowledgment issues. The twoapproaches both apply HOG-TOP, demonstrating that facial appearanceplays an essential part for both facial expressionrecognition issues.

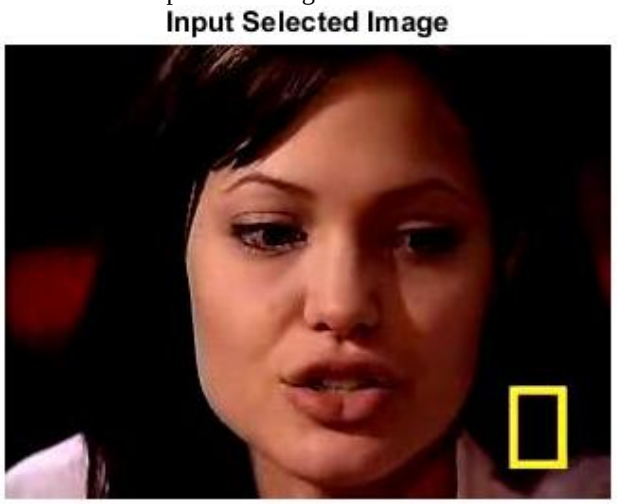

Figure 11. Input Image

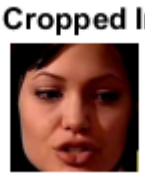

Figure 12. Face Cropped Image

Figure 13. Gaussian Filtered Image.

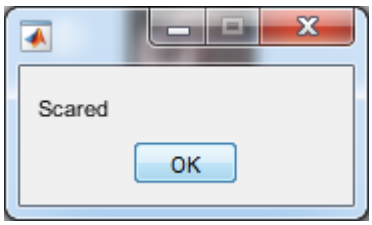

Figure 14. Resultant Expression.

International Journal of Electronics Engineering (ISSN: 0973-7383)
Volume 10 • Issue 2 pp. 45-57 June 2018-Dec 2018 www.csjournals.com

TABLE XI: The Performance of HOG-TOP with VariousBlock Sizes on the CK+ Database (%)

	12 × 12	24 × 24	16 × 16	32 × 32
Anger	84.4	80.0	88.9	84.4
Contempt	72.2	66.7	66.7	66.7
Disgust	93.2	93.2	94.9	96.6
Fear	88.0	80.0	76.0	72.0
Happiness	95.7	95.7	95.7	97.1
Sadness	64.3	64.3	67.9	60.7
Surprise	96.4	96.4	97.6	97.6
Overall	89.3	87.8	89.6	88.7

TABLE XII: The Performance of HOG-TOP with VariousBlock Sizes On The Validation Set of AFEW 4.0 Database (%)

	12 × 12	24 × 24	16 × 16	32 × 32
Neutral	54.0	38.1	58.7	46.0
Anger	71.9	71.9	73.4	73.4
Disgust	20.0	15.0	22.5	20.0
Fear	6.50	15.2	4.30	13.0
Happiness	57.1	63.5	60.3	60.3
Sadness	4.90	9.80	4.90	9.80
Surprise	2.20	13.0	2.20	8.70
Overall	34.2	35.2	35.8	36.0

Contrasted and LBP-TOP, HOG-TOPis more minimized and viable to describe facial appearancechanges. Facial setup changes likewise give usefulclues to outward appearance examination. The facial landmarkscan be found precisely on a face picture under lab-controlledenvironment, speaking to the facial setup changescaused by facial muscle developments. We propose a neweffective geometric component in light of twist change offacial points of interest and the proposed geometric twist featureis powerful to catch facial arrangement changes. On theother hand, it is exceptionally testing to find facial landmarkson confront pictures in nature. Be that as it may, the discourse additionally playsan vital part on influence acknowledgment. Rather than usinggeometric highlight, acoustic element is utilized for facialexpression acknowledgment in nature.

Figure 15. Input Image – Case2

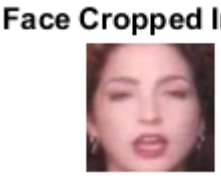

Figure 16. Face Cropped Image

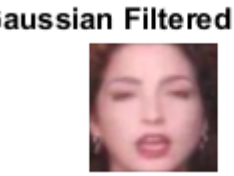

Figure 17. Gaussian Filtered Image.

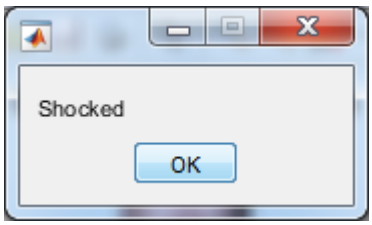

Figure 18. Resultant Expression

Trial resultsshow that diverse highlights can make distinctive contributions to outward appearance acknowledgment and the different feature fusion can improve the discriminative capacity of the multiple features. We additionally take note of that for outward appearance recognitionin the wild, despite the fact that our technique outflanks the baseline method, the execution is by and large not on a par with thatin outward appearance acknowledgment under lab-controlled environment.Facial articulation acknowledgment in the wild is muchmore testing and it will be one of our future researchfocuses.

V. CONCLUSION

Video based outward appearance acknowledgment is a challengingand long standing issue. In this paper, we misuse thepotentials of varying media modalities and propose an effective framework with different element combination to deal with thisproblem. Both the visual modalities (confront pictures) and audiomodalities (discourse) are used in our examination. Another featuredescriptor called Histogram of Oriented Gradients fromThree Orthogonal Planes (HOG-TOP) is proposed to extract dynamic surfaces from video groupings to portray facial appearance changes. Investigations led on threepublic databases (CK+, GEMEP-FERA 2011, AFEW4.0) have shown that HOG-TOP executes and in addition a broadly used feature LBP-TOP in speaking to dynamic surfaces from video arrangements. In addition, HOG-TOP is more effective to catch unobtrusive facial appearance changes and strong in dealing with outward appearance acknowledgment in nature. In addition, HOG-TOP is more smaller. With a specific end goal to capture facial arrange transforms, we present a compelling geometric feature

getting from the twist change of the facial landmarks. Understanding that voice is another intense wayfor people to transmit message, we additionally explorethe part of discourse and utilize the acoustic component foraffect acknowledgment in video. We connected the various featurefusion to manage outward appearance acknowledgment under labcontrolledenvironment and in nature. Tests conductedon two outward appearance datasets, CK+ and AFEW4.0, show that our approach can accomplish a promisingperformance in outward appearance acknowledgment in video.

VII. REFERENCES

[1] R. A. Calvo and S. D'Mello, "Affect Detection An Interdisciplinary Review of Models, Methods, and Their Applications,"IEEE Transactions on Affective Computing, vol. 1, pp. 18-37, 2010.
[2] Y. l. Tian, T. Kanade, and J. F. Cohn, "Recognizing action units forfacial expression analysis," IEEE Transactions on Pattern Analysisand Machine Intelligence, vol. 23, pp 97-115, 2001.
[3] K. Scherer and P. Ekman,"Handbook of Methods in NonverbalBehavior Research," UK: Cambridge Univ. Press, 1982.
[4] J. F. Cohn and P. Ekman,"Measuring facial action," 2005.
[5] P. Ekman and W. V. Friesen,"Facial Action Coding System: ATechnique for the Measurement of Facial Movement," ConsultingPsychologists Press, 1978.
[6] P. Ekman, W. V. Friesen, and J. C. Hager,"Facial Action CodingSystem: The Manual on CD ROM. A Human Face," 2002.
[7] P. Ekman,"An argument for basic emotions," Cognition & Emotion,vol. 6, pp. 169-200, 1992.
[8] S. Z. Li and A. K. Jain,"Handbook of face recognition," springer, 2011.
[9] N. Dalal and B. Triggs,"Histograms of Oriented Gradients forHuman Detection," IEEE Conference on Computer Vision and PatternRecognition, 2005, pp. 886-893.
[10] G. Zhao and M. Pietikainen,"Dynamic texture recognition usinglocal binary patterns with an application to facial expressions,"IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.29, pp. 915-928, 2007.
[11] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, andI. Matthews,"The Extended Cohn-Kanade Dataset (CK+) Acomplete dataset for action unit and emotion-specified expression,"IEEE Conference on Computer Vision and Pattern RecognitionWorkshops (CVPRW), 2010, pp. 94-101.
[12] S. W. Chew, P. Lucey, S. Lucey, J. Saragih, J. F. Cohn, andS. Sridharan,"Person-independent facial expression detectionusing constrained local models," IEEE International Conference onAutomatic Face & Gesture Recognition and Workshops, 2011, pp. 915-920.
[13] S. Taheri, P. Turaga, and R. Chellappa, "Towards view-invariantexpression analysis using analytic shape manifolds," IEEE InternationalConference on Automatic Face & Gesture Recognition andWorkshops, 2011, pp. 306-313.
[14] A. Saeed, A. Al Hamadi, R. Niese, and M. Elzobi, "Effectivegeometric features for human emotion recognition," IEEE 11thInternational Conference on Signal Processing (ICSP), 2012, pp. 623-627.
[15] K. Sikka, T. Wu, J. Susskind, and M. Bartlett, "Exploring bag ofwords architectures in the facial expression domain," in ComputerVision-ECCV Workshops and Demonstrations, 2012, pp. 250-259.
[16] Y. Rahulamathavan, R. C. W. Phan, J. A. Chambers, andD. J. Parish, "Facial Expression Recognition in the EncryptedDomain Based on Local Fisher Discriminant Analysis," IEEETransactions on Affective Computing, vol. 4, pp. 83-92, 2013.
[17] L. Zhang and D. Tjondronegoro, "Facial expression recognition using facial movement features," IEEE Transactions on Affective Computing, vol. 2, pp. 219-229, 2011.
[18] S. Happy and A. Routray, "Automatic facial expression recognition using features of salient facial patches," IEEE Transactions on Affective Computing, vol. 6, pp. 1-12, 2015.
[19] M. F. Valstar, B. Jiang, M. Mehu, M. Pantic, and K. Scherer, "Thefirst facial expression recognition and analysis challenge," IEEEInternational Conference on Automatic Face & Gesture Recognitionand Workshops, 2011, pp. 921-926.
[20] A. Dhall, A. Asthana, R. Goecke, and T. Gedeon, "Emotionrecognition using PHOG and LPQ features," IEEE InternationalConference on Automatic Face & Gesture Recognition and Workshops,2011, pp. 878-883.

AUTHOR's

Varun Kumar Singhal,has completed B.E (ECE) fromRajiv Gandhi Technical University,Bhopal, M.Tech (VLSI)from Rajiv Gandhi TechnicalUniversity, Bhopal, Currentlyhe is working as an AssistantProfessor of ECE Departmentin Patel College of Engineering,Bhopal, India.