This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2017.2704087, IEEE Access

Facial Expression Recognition from Depth Video with Patterns of Oriented Motion Flow

Md. Hasanul Kabir, *Member, IEEE,* Md Sirajus Salekin, Md. Zia Uddin, *Member, IEEE,* and M. Abdullah-Al-Wadud, *Member, IEEE*

***Abstract*—In this paper, we propose a novel feature representation method by a new feature descriptor, named Patterns of Oriented Motion Flow (POMF) from the optical flow information, to recognize the proper facial expression from the facial video. The POMF computes different directional motion information and encodes those directional flow information with enhanced local texture micro patterns. As it captures the spatial temporal changes of facial movements through optical flow and enables to observe both local and global structures, it shows its robustness in recognizing facial information. Finally, the POMF histogram is used to train the expression model through Hidden Markov Model (HMM). To train through the HMM, the objective sequences are produced by the generation of codebook using K-means clustering technique. The performance of the proposed method has been evaluated over the RGB and Depth camera based video. Experimental results demonstrate that the proposed POMF descriptor is more robust in extracting facial information and provides higher classification rate compared to other existing promising methods.**

***Index Terms*—Depth Image, Facial Expression Recognition, Local Binary Pattern, Optical Flow, Patterns of Oriented Motion Flow.**

I. Introduction

FACIAL expressions provide non-verbal cues which are the representation of a person's emotions or intentions. We can easily capture anyone's behavior or reaction based on these natural indications. Facial expression recognition systems have attracted the researchers a lot in the last few decades because of the increasing demand in the field of automatic human-computer interaction system. Basically, the natural identity of facial expression makes it more applicable over other biometrics. An automatic facial expression recognition system refers to a computer system which tries to analyze and recognize the facial features from the visual perspective. During the last two decades, many methods have been proposed for different face-related problems, where different facial feature extraction techniques have been introduced. Based on the types of features used, facial feature extraction approaches can be approximately classified into two different categories: geometric feature-based methods and appearance-based methods.

In geometric feature-based methods, the feature vector is formed based on the geometric relationships, such as positions, angles or distances between different facial components (eyes, ears, nose etc.). Earlier methods for facial recognition were mostly based on these geometric feature representations. For facial expression recognition, facial action coding system (FACS) [1], [2] is a popular geometric feature-based method which represents facial expression by means of a set of action units (AU). Each action unit represents the physical behavior of a specific facial muscle. Later, Zhang [3] proposed a feature extraction method based on the geometric positions of 34 manually selected fiducial points. A similar type of representation was employed by Guo and Dyer [4], where they utilized linear programming in order to perform simultaneous feature selection and classifier training. Valstar et al. [5], [6] have studied facial expression analysis based on tracked fiducial point data and reported that geometric features provide similar or better performance than appearance-based methods in action unit recognition. However, the effectiveness of geometric methods is heavily dependent on the accurate detection of facial components, which is a difficult task in changing and unconstrained environment, thus making geometric methods difficult to accommodate in many scenarios [7].

On the other hand, appearance-based methods extract the facial appearance by applying image filter or filter bank on the whole face image or some specific facial regions. Basically, two types of approaches can be observed in appearance-based methods. One type of approach tries to apply some feature reduction or class separation methods directly on the intensity values to minimize the feature size. Another type of approach uses any descriptor on the image intensity values and generate some key features from the image. In case of feature minimization or class separation approaches Principal Component Analysis (PCA) [8], [9], Linear Discriminant Analysis (LDA) [10], [11], Independent Component Analysis (ICA) [12], [13], Gabor wavelets [14] are the commonly-used appearance-based methods for facial expression recognition. On the contrary, key feature generation type approaches apply any descriptor on the image intensity values. These type of approaches try to generate some fruitful information from the neighborhood regions of an image and generate the key features. LBP [15], LDP [16], POEM [17] are some popular descriptors for the feature extraction in case of facial expression recognition system.

This work was supported by the Research Center of the College of Computer and Information Sciences, King Saud University, Riyadh, Saudi Arabia.

Md. Hasanul Kabir and Md Sirajus Salekin are with the Department of Computer Science and Engineering, Islamic University of Technology, Gazipur, Dhaka, Bangladesh (email: {hasanul, salekin}@iut-dhaka.edu). Md. Hasanul Kabir is the corresponding author.

Md. Zia Uddin is with the Department of Informatics, University of Oslo, Norway (email: mdzu@ifi.uio.no).

M. Abdullah-Al-Wadud is with the Department of Software Engineering, College of Computer and Information Sciences, King Saud University, Riyadh 11543, Kingdom of Saudi Arabia (e-mail: mwadud@ksu.edu.sa)

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2017.2704087, IEEE Access

Besides, most of the time RGB camera is used to capture facial video. But nowadays, many of the researchers are attracted to the depth camera [18], [19]. As depth camera provides the depth information of any image which highly exhibits important features of the facial image, so facial expression recognition can be more reliable and efficient on depth based facial video. Apart from this, the privacy of the individuals is also highly reserved in depth video which makes it more viable in the real life.

In this paper, a novel feature descriptor named Patterns of Oriented Motion Flow (POMF) is proposed to identify the facial expression from the depth video. The POMF computes different directional motion information and encodes those directional flow information with enhanced local texture descriptor. Both RGB and Depth camera-based experiments are performed with different conventional facial expression approaches and superior results are achieved using the POMF over the depth video images.

The rest of the paper is organized as follows: In section II, overall idea of our proposed POMF descriptor is discussed. Then, in section III, how to extract the POMF feature and how to model and recognize the system are explained. Later, in section IV, experimental setup, experimental results and performance analysis of our proposed descriptor with various promising methods are explicated. Finally, in section V, our research contributions are concluded pointing out its future potential developments.

II. Proposed Motion Feature Extraction by POMF

Our proposed POMF descriptor works based on the motion changes of the images which are captured by the optical flow information. Later on, from the directional motion information, a robust pattern will be generated using local texture pattern.

A. Optical Flow Estimation

Optical flow features have been used increasingly since the past decade in the field of any motion detection or object tracking. As it defines the changes in image from frame to frame, nowadays, it is being used for facial expression recognition from video [20]–[23] and already it has expressed its robustness. From the video image of expression, our first step is to calculate the motion change from frame to frame. And it is done by estimating optical flow.

The optical flow methods try to calculate the motion between two image frames which are taken at times t and $t+\Delta t$ at every voxel position. For a $2D+t$ dimensional case ($3D$ or $n-D$ cases are similar) a voxel at location (x, y, t) with intensity $I(x, y, t)$ will have moved by Δx, Δy and Δt between the two image frames, and the following image constraint equation can be given.

$$I(x, y, t) = I(x + \Delta x, y + \Delta y, t + \Delta t) \tag{1}$$

If we consider that the movement is very small, the image constraint at I(x,y,t) with Taylor series can be developed as follows:

$$\begin{aligned} I(x + \Delta x, y + \Delta y, t + \Delta t) &= I(x, y, t) \\ &+ \frac{\delta I}{\delta x}\Delta x + \frac{\delta I}{\delta y}\Delta y + \frac{\delta I}{\delta t}\Delta t + H.O.T \end{aligned} \tag{2}$$

Here, the main challenge of optical flow estimation is which property to track and how to track it. More precisely, it needs to track a property which includes motion information more robustly. Several image properties have been used for this purpose throughout different optical flow estimation methods [24], [25].

From any optical flow estimation, two kinds of flow information are found which are known as horizontal flow (u) and vertical flow (v). Each of the u and v reveals two directional flow information from two consecutive images. The positive u and the negative u represent the flow information from left to right and right to left respectively. On the other hand, the positive v and the negative v represent the flow information from top to bottom and bottom to top respectively. In our method, we have used Lucas-Kanade [24] method to estimate the optical flow information.

B. Local Binary Pattern (LBP)

Local Binary Pattern (LBP), a gray-scale invariant texture pattern has gained much popularity among the researchers for encoding the spatial information of image texture. The basic LBP [15] was developed based on the presumption that image texture will be represented by two aspects, a pattern, and its strength. It encodes the gray-scale structure of an image using a binary code. It generates a label for each pixel of an image by thresholding its neighbor values with the center value. The resulting pattern represents a binary number which is converted to a decimal number before assigning to each pixel according to the following equations.

$$LBP_{P,R}(x_c, y_c) = \sum_{p=0}^{P-1} s(g_p - g_c)2^p \tag{3}$$

$$s(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases} \tag{4}$$

Here, g_c denotes the gray value of the center pixel (x_c, y_c) and g_p corresponds to the gray values of equally spaced pixels P on the circumference of a circle with radius R. This encoded pattern ensures the rotation invariance of the gray-scale structure of the image. For further improvement of rotation invariance and finer quantization of the angular space, a variation of LBP was also proposed which is known as Uniform LBP. As a certain local binary pattern appears frequently in the significant image area, they contain very few transitions from 0 to 1 and 1 to 0 in the circular bit sequence. Ojala et al. [15] observed that LBP patterns with $U \leq 2$ are the fundamental properties of texture, which provide a vast majority of all the 8-bit binary patterns present in any texture image. Therefore, uniform patterns are able to describe significant local texture information, such as bright spot, flat area or dark spot, and edges of varying positive and negative curvature. All the other patterns ($U > 2$) are grouped under a miscellaneous label.

C. Patterns of Oriented Motion Flow (POMF) Descriptor

In this paper, we propose a directional optical flow based descriptor which is called Patterns of Oriented Motion Flow

(POMF). The rudimentary idea of the POMF is to discretize the motion change information and capture the encoded micro pattern from those motion changes. At first, discretized motion change will be enhanced by the local motion changes and then further incorporated by the self-similarity measurements of LBP micro pattern. Taking the directional changes of image information and encoding those directional image velocity by LBP in POMF descriptor, a robust pattern will be generated.

1) POMF Code: **Optical flow computation and orientation quantization:** The first step in extracting the POMF feature is the computation of optical flow between consecutive two image frames from the video. The optical flow of the image produces the two flows information: horizontal (u) and vertical (v) for each pixel. From these motion flow information, four directional flows information are produced discretely over the u and v. Positive u_p represents the horizontal flow from left to right whereas the negative u_n represents the reverse direction. Similarly, positive v_p represents the vertical flow from top to bottom and negative v_n represents the reverse direction. As a facial expression starts from a neutral state and then exposes the expression, and again ends with a neutral expression, certainly a prominent motion change is elicited through the different facial part like mouth, eye, eyebrow, nose, chin, forehead etc. Fig. 1 shows prominent motion changes of two consecutive depth image frames. Here, the our main intention is to make some directional motion vectors from these horizontal and vertical motion information. A total number of four directional patterns are created which represent flow energy information in four discrete directions U_pV_p, U_pV_n, U_nV_n, U_nV_p. As a result, a robust pattern from the responses of the directional approach can be achieved (Fig. 2).

Local flow energy accumulation: The second step is to introduce the motion flow information from the neighboring regions. A local histogram of the motion orientation over the cell pixels is estimated. Here, flow energy $(u^2 + v^2)$ can be used as a vote weight or some function of the flow energy. In our original POMF descriptor, flow energy is used. As a result, each significant motion pixel is identified as any of the four directional motion change contributing pixels where it contains the flow energy.

Global self-similarity estimation: The last step of the POMF descriptor is to encode the accumulated directional flow information by a micropattern. The self-similarity of the cells in a global region is estimated by the LBP descriptor. LBP is a robust texture descriptor which is robust for encoding any pattern information providing the rotation invariance of the structure. In the original LBP descriptor [15], a uniform LBP was suggested which was more robust with finer angular space. Since the significant facial expression appears frequently the uniform LBP is used in our experiment and it performs better than the LBP. We continue LBP encoding process on the accumulated flow energy across four different flow directions to build the final POMF descriptor. A POMF feature is calculated at every pixel position p over each discretized flow direction (Fig. 3) according to the following equation.

$$POMF^{\theta_i}_{B,C,N}(p) = \sum_{j=1}^{N} f(S(E^{\theta_i}_p, E^{\theta_i}_j))2^j \tag{5}$$

Here, θ_i represents the directional flow at i, where $(i = U_pV_p, U_pV_n, U_nV_p, U_nV_n)$ direction; $E^{\theta_i}_p, E^{\theta_i}_j$ are the directional optical flow values of central pixel p, and surrounding pixels respectively; S is the similarity function; B, C refer to the size of blocks and cells respectively; N is the number of pixels surrounding the considered central pixel p; and f is defined as:

$$f(x) = \begin{cases} 1 & x \geq \alpha \\ 0 & x < \alpha \end{cases} \tag{6}$$

Here, the value α is slightly larger than zero which ensures some stability in uniform regions. Finally, from two consecutive image frames, we get the descriptor which will be the concatenation of the unidirectional POMF of four directional motion flows as follows:

$$POMF_{B,C,N}(p) = \{POMF^{U_pV_p}, POMF^{U_pV_n}, POMF^{U_nV_p}, POMF^{U_nV_n}\} \tag{7}$$

2) Robustness of POMF descriptor: POMF descriptor contains not only the local significant motion changes from two consecutive frames but also represents the directional motion information in neighboring regions. It determines the following properties:

- POMF is a directional approach. So it clarifies its robustness against any directional motion changes with different levels of accuracy.
- POMF descriptor reveals multi-resolution feature because of different scales of cells and blocks. At the same time it contains both local and global information.
- Introducing flow energy information at each pixel, it represents horizontal and vertical motion effects at the same time which makes it more robust for identifying proper motion changes.
- As the motion flow information from the image frames is considered, the facial problems like age, beard, pose, gender etc. can easily be overcome.

Therefore, the POMF descriptor contains richer information from the facial expression video. It considers the relationship between frames by the directional optical flow. As a result, the expression changes from neutral status to final status is robustly represented through the POMF. Moreover, the rest of the part except the expression is almost same throughout the frames of any particular video, and so, only the expression exposed information will be captured which makes POMF robust against the variations like illumination, background pattern, beard, facial hair, gender, pose etc.

III. Facial Expression Recognition Based on POMF

Our proposed method starts with the optical flow information from image frames. For facial video, depth camera-based video is preferred. Then from the optical flow information, the proposed POMF feature will be generated and those will be trained by HMM. Finally, the desired expression will be recognized from the maximum likelihood response of HMM. Fig. 4 shows the general steps of the proposed Facial Expression Recognition (FER) system.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2017.2704087, IEEE Access

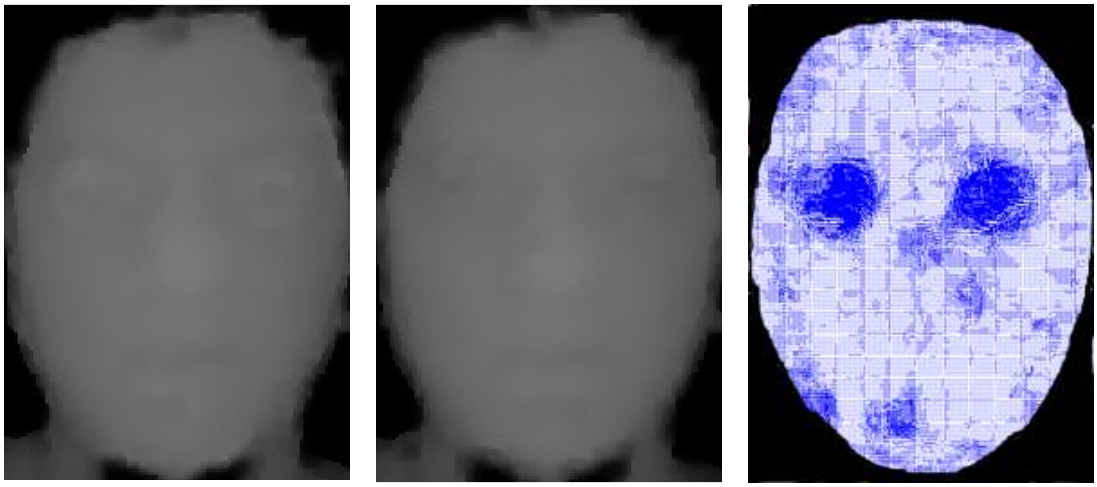

Fig. 1. Consecutive two frames of an anger expressed depth video sample and corresponding optical flow response respectively (from left to right).

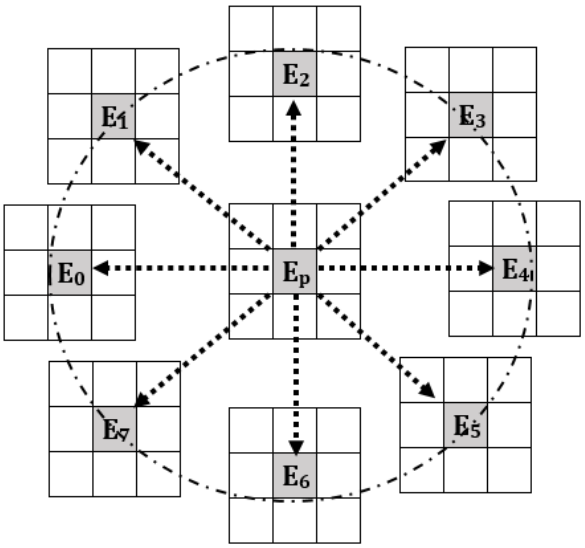

Fig. 3. Estimation of self-similarity over a region.

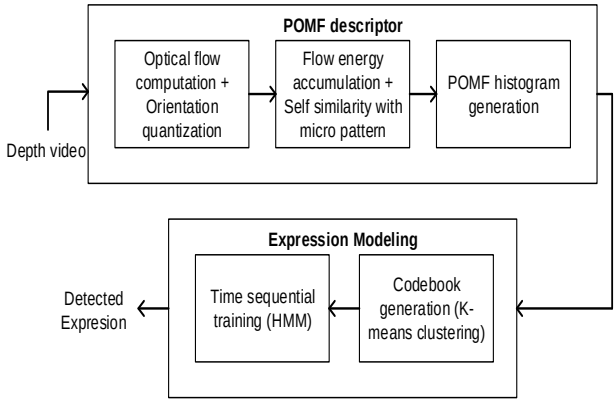

Fig. 4. Steps of the proposed facial expression recognition system.

A. POMF Histogram Sequences for Facial Expression

For facial expression recognition from a video sequence, at first consecutive image frames are extracted from the video where an expression will be represented by some sequential static images. Here, image frame should be extracted in such a way that a significant motion change is present. As optical flow information will be used for POMF, the presence of significant motion change will provide better results. Then for every two consecutive image frames, an optical flow information is estimated. In our experiment, Lucas-Kanade [24] method is used for estimating the optical flow information. From the optical flow information, two kinds of motion flow information: horizontal (u) and vertical (v) are generated. On this u and v the proposed POMF code is applied and it produced the four directional encoded motion flow information. Now from each consecutive two image frames, a POMF feature vector is generated. Basically, from the optical flow information of each frame, four new directional frames will appear. Each of them is divided into multiple non-overlapping regions and the histogram is extracted from each region which will create the POMF-HS. So if the video is divided into an n number of image frames, then $(n-1)$ number of POMF Histogram (POMF-HS) will be generated. All the accumulated feature vectors are used as the feature representation of that particular facial expression from the video. Fig. 5 shows the steps of POMF feature extraction from two consecutive image frames.

B. Modeling and Recognizing the Expression

After the POMF feature extraction, a left to right feed forward discrete HMMs are applied for expression training. Hidden Markov Model (HMM) [26], [27] is doubly stochastic process based on discrete Markov process. It is not observable but can only be observed through another set of stochastic processes that produce the sequences of observed symbols. The observations are probabilistic functions of each state while state sequences are not known. The state sequence with the highest probability is obtained from another process. It is known for its wide application in temporal pattern recognition

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2017.2704087, IEEE Access

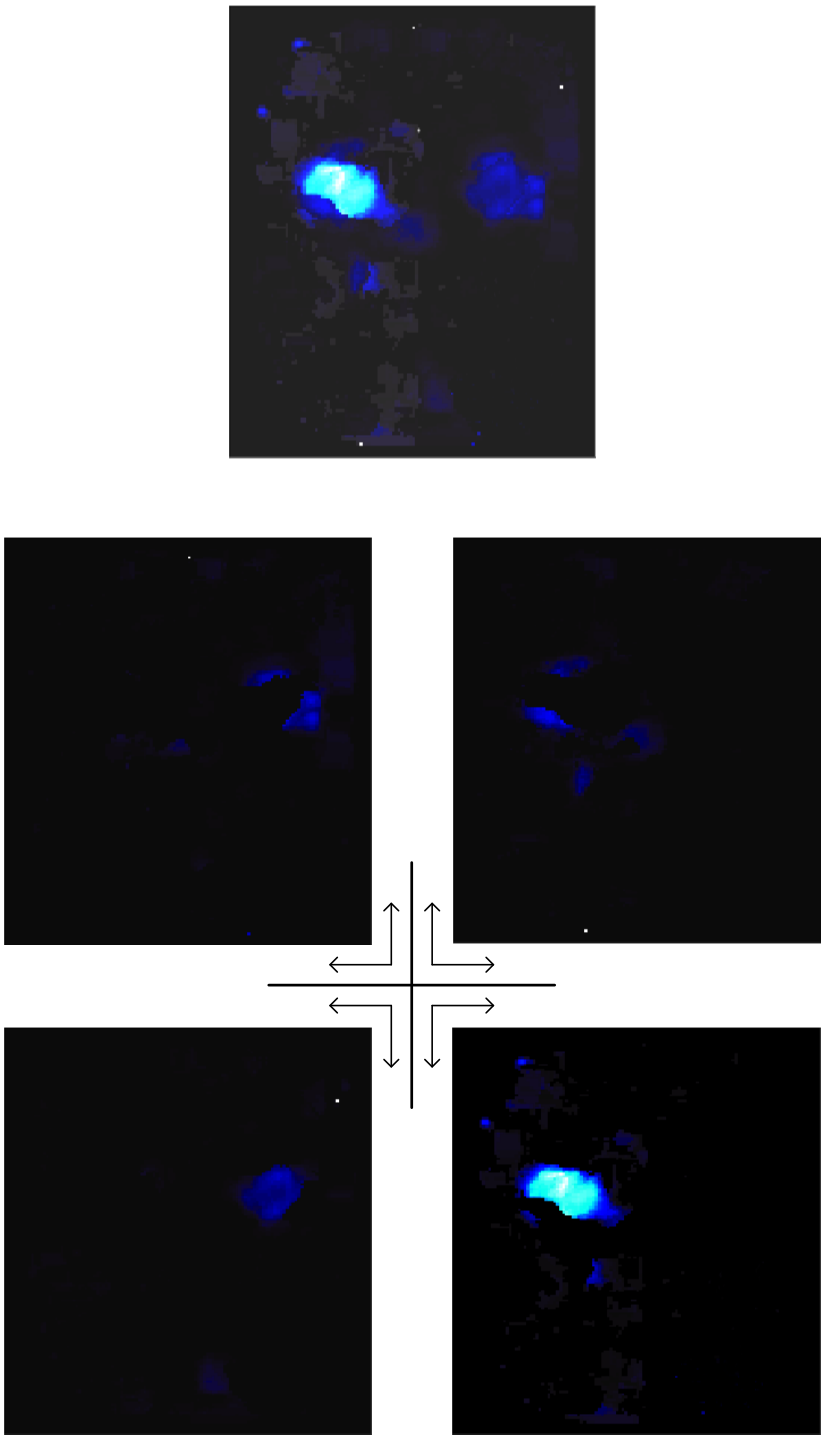

Fig. 2. Responses of the two consecutive frames after discrete flow orientations. Flow energy response of two consecutive anger expressed depth video sample (upper row), directional flow U_nV_p (middle row, left), directional flow U_pV_p (middle row, right), directional flow U_nV_n (lower row, left), and directional flow U_pV_n (lower row, right).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2017.2704087, IEEE Access

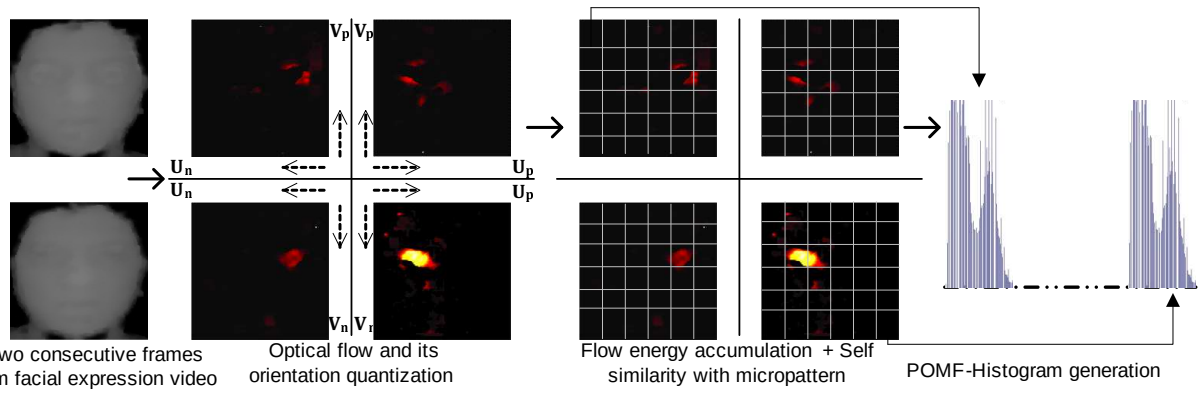

Fig. 5. Steps of POMF feature extraction from optical flow of two consecutive frames.

especially in the field of machine learning and data mining. As HMM takes the observation sequences value, K-means [28] clustering technique is used to generate the desired observation sequences. K-means clustering technique takes all the samples and clusters the samples based on their features. For all type of expressions, all sequences are clustered by the codebook and observation symbol sequences are produced which is trained by HMMs.

A basic HMM is represented by a set of parameters $\lambda = \{\pi, A, B\}$ where, π = prior probabilities of the states, A = transition probabilities of one state to another, B = observation symbol probability matrix. The main purpose of HMM is to find out the model type (λ_k) with the highest probability of the likelihood $P(O|\lambda_k)$ for the observation sequence O. If we denote the states in the model by $S = \{s_1, s_2, ..., s_N\}$ and different states at t-th time by $Q = \{q_1, q_2, ..., q_t\}$, then the HMM parameters can be presented as follows.

$$A = \{a_{ij}\}, a_{ij} = P(q_{t+1} = S_j | q_t = S_i), where\ 1 \leq i, j \leq N \quad (8)$$

$$B = \{b_j(O_t)\}, b_j = P(O_t | q_t = S_j), where\ 1 \leq j \leq N \quad (9)$$

$$\pi = \{\pi_j)\}, \pi_j = P(q_1 = S_j) \quad (10)$$

To test a sample video of facial expressions, at first proper feature vectors are extracted by POMF feature descriptor using the same procedure. Then each of the trained expression models (λ_k) is used to generate the likelihood response for the particular sample observation sequence (O). Finally, to determine the test observation sequence, highest likelihood response from all N-trained expression HMMs evokes the corresponding desired class of the facial expression as follows:

$$\text{Detected Expression} = \arg\max_{k=1}^{N}(P(O|\lambda_k)) \quad (11)$$

In our experiments, the Baum-Welch algorithm [29] was applied to estimate the parameters of the expression HMMs.

IV. Facial Expression Recognition Experimental Results

Human is able to percept RGB images but we are dealing with Machine. Machine's perception is different than human. So we can provide some more information rich image to the Machine. Hence, the concept comes about depth image. In the depth image, high pixel value represents a near distance and low pixel value represents a far distance. Depth information greatly contributes to the facial expression. Besides, it also ensures the privacy of the individuals. The Depth database [18] of facial expression which was used in our experiment was built using Zcam [30] depth camera. Head motion was assumed to be small and neglected. Some threshold values were used empirically to extract the face from the video based on the depth information. The Depth database [18] was developed based on both RGB and Depth camera-based image sequences. Examples of different expressions of the Depth database are shown in Fig. 6. In our experiment, there were six kinds of expressions to be recognized by system, which were anger, disgust, fear, happiness, sadness, and surprise. For each of the cases, the expression video clips were started and ended with a neutral expression. In our experiments, a total of 120 video clips of variable length from all expressions were used in the experiment. To train and test each facial expression model, 20 and 40 image sequences were applied respectively. To train HMM, the features were symbolized by the K-means clustering [28] technique using a cluster size of 40 and there were 5 intermediate hidden states throughout all the experiments which were selected empirically.

For facial expression recognition, the performance of the proposed POMF was evaluated with some promising methods, namely Optical flow-PCA [31], POEM [17], LDP-PCA [18]. All these approaches were performed based on the same experimental setup. The recognition rates of these approaches on RGB faces are shown in Tables I, II, III and IV while the performance on Depth faces are shown in Tables V, VI, VII and VIII. Besides, the average recognition rates are also represented in Table IX to draw the analogy among the different approaches. It is evident from the experimental results that all the approaches perform well on Depth video rather than RGB video and POMF shows its superiority over other approaches.

In the case of RGB video images, POEM and LDP-PCA show almost similar performance. POEM achieved 86.67% while LDP-PCA achieved 87.08%. Both of the approaches use a feature representation instead of intensity values. LDP uses the neighborhood edge responses values whereas, POEM considers the gradient value. On the other hand, proposed

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2017.2704087, IEEE Access

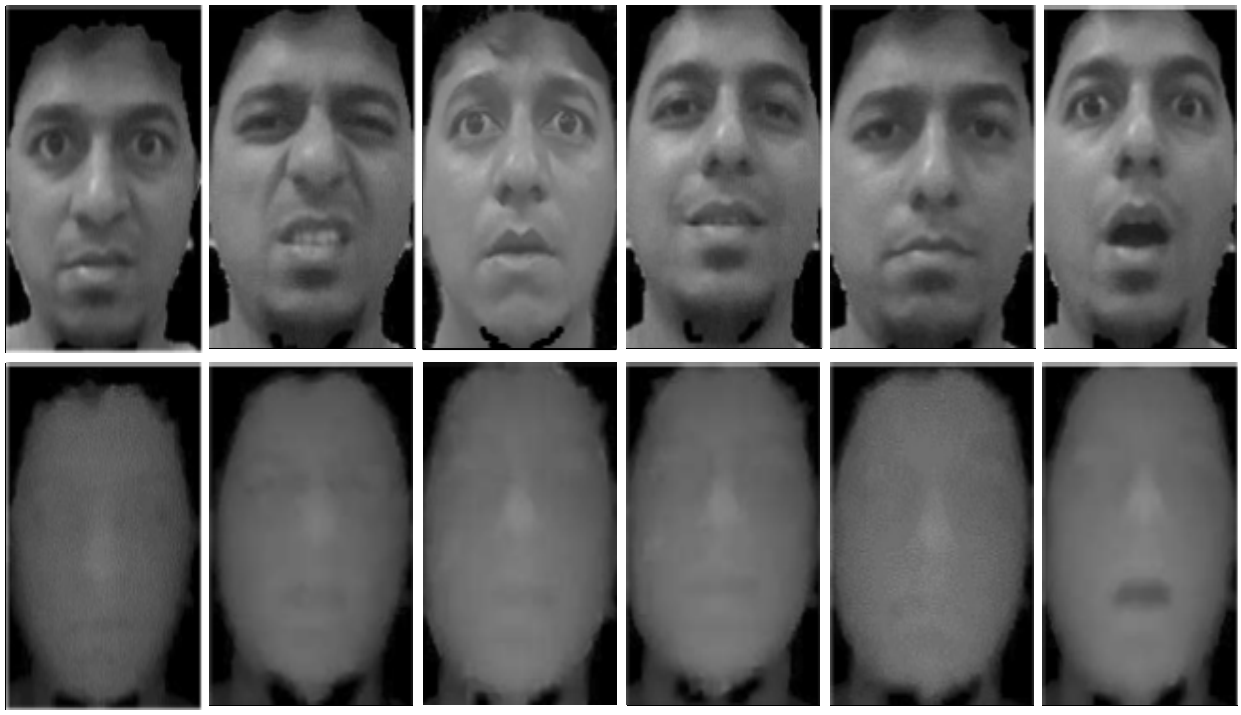

Fig. 6. Examples of different expressions of the Depth database. Anger, Disgust, Fear, Happiness, Sadness, Surprise expression respectively (from left to right); normal gray faces (upper row) and the corresponding depth faces (lower row).

TABLE I
CONFUSION MATRIX USING RGB FACES WITH OF-PCA.

Expression	Anger	Disgust	Fear	Happiness	Sadness	Surprise
Anger	72.50	0	10	0	17.50	0
Disgust	10	75	0	0	15	0
Fear	0	0	75	5	20	0
Happiness	10	0	0	80	0	10
Sadness	0	0	15	0	77.50	7.50
Surprise	10	10	0	0	0	80

TABLE II
CONFUSION MATRIX USING RGB FACES WITH POEM.

Expression	Anger	Disgust	Fear	Happiness	Sadness	Surprise
Anger	90	0	0	0	10	0
Disgust	10	85	0	0	5	0
Fear	0	0	85	0	15	0
Happiness	0	0	0	85	0	15
Sadness	10	0	0	0	90	0
Surprise	12.50	2.50	0	0	0	85

TABLE III
CONFUSION MATRIX USING RGB FACES WITH LDP-PCA.

Expression	Anger	Disgust	Fear	Happiness	Sadness	Surprise
Anger	85	10	0	0	5	0
Disgust	0	87.50	0	2.50	10.50	0
Fear	0	0	85	5	10	0
Happiness	2.50	0	5	87.50	0	5
Sadness	0	0	0	2.50	87.50	10.5
Surprise	0	7.50	2.50	0	0	90

TABLE IV
CONFUSION MATRIX USING RGB FACES WITH POMF.

Expression	Anger	Disgust	Fear	Happiness	Sadness	Surprise
Anger	90	0	0	0	10	0
Disgust	5	85	0	0	10	0
Fear	0	0	85	0	15	0
Happiness	0	0	0	87.50	0	12.50
Sadness	10	0	0	0	90	0
Surprise	0	0	0	10	0	90

TABLE V
CONFUSION MATRIX USING DEPTH FACES WITH OF-PCA.

Expression	Anger	Disgust	Fear	Happiness	Sadness	Surprise
Anger	77.50	0	5	0	17.50	0
Disgust	5	85	0	0	10	0
Fear	0	0	77.50	5	17.50	0
Happiness	10	0	0	85	0	5
Sadness	0	0	10	0	80	10
Surprise	10	5	0	0	0	85

POMF outperforms others achieving the highest accuracy 87.91% while the direct Optical flow-PCA approach was only 76.67%.

On the contrary, in the case of Depth video images, all the approaches manifest a better recognition rate than the RGB video images. POEM and LDP-PCA both approaches reached a better recognition rate to 92.50% and 92.92% respectively. An improved recognition rate of 81.67% is also found for Optical flow-PCA approach than its prior RGB video. But POMF gave the best recognition rate of 94.17%. However, PCA-based method takes a higher computation time for large feature vector as it needs to calculate the covariance matrix for the computation of eigen vector and eigen value.

HMM is popular to decode time-sequential events and considered to be better than others [18], [32], [33]. Our depth image-based experiments are further enhanced in this work to show HMM's superiority over other traditional classifiers such as multiclass Support Vector Machine (SVM) [34] using polynomial kernel, K-Nearest Neighbours [35] and Nave-Bayes [36] classifier. Table X shows the comparative performance of different classifiers using POMF on Depth database [18].

TABLE VI
CONFUSION MATRIX USING DEPTH FACES WITH POEM.

Expression	Anger	Disgust	Fear	Happiness	Sadness	Surprise
Anger	92.50	0	0	0	7.50	0
Disgust	7.50	92.5	0	0	0	0
Fear	0	0	90	0	10	0
Happiness	0	0	0	95	0	5
Sadness	5	0	0	0	95	0
Surprise	5	5	0	0	0	90

TABLE VII
CONFUSION MATRIX USING DEPTH FACES WITH LDP-PCA.

Expression	Anger	Disgust	Fear	Happiness	Sadness	Surprise
Anger	92.50	0	0	0	7.50	0
Disgust	0	92.50	0	0	7.50	0
Fear	0	0	92.50	0	7.50	0
Happiness	7.50	0	0	92.50	0	0
Sadness	0	0	0	0	92.50	7.50
Surprise	0	5	0	0	0	95

TABLE VIII
CONFUSION MATRIX USING DEPTH FACES WITH POMF.

Expression	Anger	Disgust	Fear	Happiness	Sadness	Surprise
Anger	95	0	0	0	5	0
Disgust	7.50	92.50	0	0	0	0
Fear	0	0	92.50	0	7.50	0
Happiness	0	0	0	95	0	5
Sadness	5	0	0	0	95	0
Surprise	0	0	0	5	0	95

TABLE IX
AVERAGE EXPRESSION RECOGNITION RATES FOR DIFFERENT APPROACHES ON RGB AND DEPTH VIDEO.

Approach	Recognition Rate (%) on RGB Video	Recognition Rate (%) on Depth Video
Optical Flow-PCA	76.67	81.67
POEM	86.67	92.50
LDP-PCA	87.08	92.92
POMF	87.91	94.17

TABLE X
EXPRESSION RECOGNITION PERFORMANCE (%) USING DIFFERENT CLASSIFIERS ON DEPTH VIDEO.

Activity	SVM	KNN	Naïve-Bayes	HMM
Anger	85	80	82.50	95
Disgust	77.50	82.50	87.50	92.50
Fear	77.50	87.50	92.50	92.50
Happiness	82.50	82.50	92.50	95
Sadness	77.50	87.50	82.50	95
Surprise	85	82.50	82.50	95
Average	80.83	83.75	86.67	94.17

V. CONCLUSION

In this paper, an optical flow based facial expression recognition system is proposed where the directional pattern encoded information is used from the optical flow of consecutive depth images. We proposed a novel and robust facial descriptor called Patterns of Oriented Motion Flow (POMF). Using this descriptor POMF histogram is generated from the sample frames to produce the expression feature vectors. Finally, the objective sequences of the feature vectors are trained through the Hidden Markov Model (HMM) to produce the expression model. As we work with the optical flow information and it represents only the changing information from a video, the significant changes can be easily captured which occur because of the facial expression. Besides, different challenges of expression recognition such as age, gender, beard, and glasses can easily be suppressed. Moreover, the directional optical flow information ensures more robust feature description by generating an oriented pattern. An experimental analysis on both RGB and Depth camera based video images is performed including some salient approaches to evaluate the strength of our proposed method. From the empirical results, it is obvious that our proposed POMF descriptor represents better recognition rate for depth based facial expression recognition system. Besides, it also turns out that Depth image shows superior performance over RGB image. In our future work, we are planning to enhance the performance of POMF by introducing the solution for nonlinearity since human face images with large pose variation demonstrate significant nonlinearity.

REFERENCES

[1] P. Ekman and W. V. Friesen, "Facial action coding system," 1977.

[2] J. Hager, P. Ekman, and W. Friesen, "Facial action coding system. salt lake city, ut: A human face," ISBN 0-931835-01-1, Tech. Rep., 2002.

[3] Z. Zhang, "Feature-based facial expression recognition: Sensitivity analysis and experiments with a multilayer perceptron," *International journal of pattern recognition and Artificial Intelligence*, vol. 13, no. 06, pp. 893–911, 1999.

[4] G. Guo and C. R. Dyer, "Simultaneous feature selection and classifier training via linear programming: A case study for face expression recognition," in *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, vol. 1. IEEE, 2003, pp. I–346.

[5] M. F. Valstar, I. Patras, and M. Pantic, "Facial action unit detection using probabilistic actively learned support vector machines on tracked facial point data," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)-Workshops*. IEEE, 2005, pp. 76–76.

[6] M. Valstar and M. Pantic, "Fully automatic facial action unit detection and temporal analysis," in *2006 Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'06)*. IEEE, 2006, pp. 149–149.

[7] C. Shan, S. Gong, and P. W. McOwan, "Facial expression recognition based on local binary patterns: A comprehensive study," *Image and Vision Computing*, vol. 27, no. 6, pp. 803–816, 2009.

[8] C. Padgett and G. W. Cottrell, "Representing face images for emotion classification," *Advances in neural information processing systems*, pp. 894–900, 1997.

[9] G. Donato, M. S. Bartlett, J. C. Hager, P. Ekman, and T. J. Sejnowski, "Classifying facial actions," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 21, no. 10, pp. 974–989, 1999.

[10] A. J. Calder, A. M. Burton, P. Miller, A. W. Young, and S. Akamatsu, "A principal component analysis of facial expressions," *Vision research*, vol. 41, no. 9, pp. 1179–1208, 2001.

[11] S. Dubuisson, F. Davoine, and M. Masson, "A solution for facial expression representation and recognition," *Signal Processing: Image Communication*, vol. 17, no. 9, pp. 657–673, 2002.

[12] I. Buciu, I. Pitas *et al.*, "Ica and gabor representation for facial expression recognition," in *Image Processing, 2003. ICIP 2003. Proceedings. 2003 International Conference on*, vol. 2. IEEE, 2003, pp. II–855.

[13] C. Fan and K. Kotani, "Facial expression recognition by supervised independent component analysis using map estimation," *IEICE transactions on information and systems*, vol. 91, no. 2, pp. 341–350, 2008.

[14] M. J. Lyons, J. Budynek, and S. Akamatsu, "Automatic classification of single facial images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 12, pp. 1357–1362, 1999.

[15] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 24, no. 7, pp. 971–987, 2002.

[16] T. Jabid, M. H. Kabir, and O. Chae, "Robust facial expression recognition based on local directional pattern," *ETRI journal*, vol. 32, no. 5, pp. 784–794, 2010.

[17] E. Silva, C. Esparza, and Y. Mejía, "Poem-based facial expression recognition, a new approach," in *2012 XVII Symposium of Image, Signal Processing, and Artificial Vision (STSIVA)*. IEEE, 2012, pp. 162–167.

[18] M. Z. Uddin, "An efficient local feature-based facial expression recognition system," *Arabian Journal for Science and Engineering*, vol. 39, no. 11, pp. 7885–7893, 2014.
[19] J. Yang, H. Wang, Z. Ding, Z. Lv, W. Wei, and H. Song, "Local stereo matching based on support weight with motion flow for dynamic scene," *IEEE Access*, vol. 4, pp. 4840–4847, 2016.
[20] M. Kenji, "Recognition of facial expression from optical flow," *IEICE TRANSACTIONS on Information and Systems*, vol. 74, no. 10, pp. 3474–3483, 1991.
[21] M. J. Black and Y. Yacoob, "Tracking and recognizing rigid and non-rigid facial motions using local parametric models of image motion," in *Computer Vision, 1995. Proceedings., Fifth International Conference on*. IEEE, 1995, pp. 374–381.
[22] Y. Yacoob and L. S. Davis, "Recognizing human facial expressions from long image sequences using optical flow," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 18, no. 6, pp. 636–642, 1996.
[23] I. A. Essa and A. P. Pentland, "Coding, analysis, interpretation, and recognition of facial expressions," *IEEE transactions on pattern analysis and machine intelligence*, vol. 19, no. 7, pp. 757–763, 1997.
[24] B. D. Lucas, T. Kanade *et al.*, "An iterative image registration technique with an application to stereo vision." in *IJCAI*, vol. 81, no. 1, 1981, pp. 674–679.
[25] H. B. S. B. G. Determining, "Optical flow," *Artificial Intelligence*, vol. 17, no. 1-3, pp. 185–203, 1981.
[26] L. E. Baum and T. Petrie, "Statistical inference for probabilistic functions of finite state markov chains," *The annals of mathematical statistics*, vol. 37, no. 6, pp. 1554–1563, 1966.
[27] L. E. Baum, T. Petrie, G. Soules, and N. Weiss, "A maximization technique occurring in the statistical analysis of probabilistic functions of markov chains," *The annals of mathematical statistics*, vol. 41, no. 1, pp. 164–171, 1970.
[28] J. MacQueen *et al.*, "Some methods for classification and analysis of multivariate observations," in *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, vol. 1, no. 14. Oakland, CA, USA., 1967, pp. 281–297.
[29] L. R. Rabiner, "A tutorial on hidden markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, 1989.
[30] G. J. Iddan and G. Yahav, "Three-dimensional imaging in the studio and elsewhere," in *Photonics West 2001-Electronic Imaging*. International Society for Optics and Photonics, 2001, pp. 48–55.
[31] M. Z. Uddin, T.-S. Kim, and B. C. Song, "An optical flow feature-based robust facial expression recognition with hmm from video," *Int. J. Innovative Comput. Inf. Control*, vol. 9, no. 4, pp. 1409–1421, 2013.
[32] I. Cohen, N. Sebe, A. Garg, L. S. Chen, and T. S. Huang, "Facial expression recognition from video sequences: temporal and static modeling," *Computer Vision and image understanding*, vol. 91, no. 1, pp. 160–187, 2003.
[33] Y. Zhu, L. C. De Silva, and C. C. Ko, "Using moment invariants and hmm in facial expression recognition," *Pattern Recognition Letters*, vol. 23, no. 1, pp. 83–91, 2002.
[34] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.
[35] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE transactions on information theory*, vol. 13, no. 1, pp. 21–27, 1967.
[36] A. McCallum, K. Nigam *et al.*, "A comparison of event models for naive bayes text classification," in *AAAI-98 workshop on learning for text categorization*, vol. 752. Citeseer, 1998, pp. 41–48.