

An Efficient Algorithm for Motion Detection Based Facial Expression Recognition using Optical Flow

Ahmad R. Naghsh-Nilchi, and Mohammad Roshanzamir

Abstract—One of the popular methods for recognition of facial expressions such as happiness, sadness and surprise is based on deformation of facial features. Motion vectors which show these deformations can be specified by the optical flow. In this method, for detecting emotions, the resulted set of motion vectors are compared with standard deformation template that caused by facial expressions. In this paper, a new method is introduced to compute the quantity of likeness in order to make decision based on the importance of obtained vectors from an optical flow approach. For finding the vectors, one of the efficient optical flow method developed by Gautama and VanHulle[17] is used. The suggested method has been examined over Cohn-Kanade AU-Coded Facial Expression Database, one of the most comprehensive collections of test images available. The experimental results show that our method could correctly recognize the facial expressions in 94% of case studies. The results also show that only a few number of image frames (three frames) are sufficient to detect facial expressions with rate of success of about 83.3%. This is a significant improvement over the available methods.

Keywords—Facial expression, Facial features, Optical flow, Motion vectors.

I. INTRODUCTION

FACIAL expression delivers rich information about human emotion and plays an important role in human communications. The study on this subject was pioneered by Darwin's works [1] and has been extensively studied in psychology during the 1980's [2]. Research in psychology has indicated that at least six emotions are universally associated with distinct facial expressions. Note that, several other emotions and many of their combinations have also been studied but remain unconfirmed as universally distinguishable. These six principle emotions are: happiness, sadness, surprise, fear, anger and disgust. In Fig. 1, six human face images, selected from Cohn-Kanade AU-Coded Facial Expression Database [3], representing each of these emotions, are shown.

For intelligent and natural human-computer interaction, it is essential to recognize facial expression automatically. Various techniques have been developed for automatic facial expression recognition, which differ in the kind of data used

(still images vs. video sequences), feature extraction methods used, and classifiers used.



Fig. 1 Six images with six universal facial expressions selected from Cohn-Kanade AU-Coded Facial Expression Database with permission [3]

Edwards, Hong and et al. [4, 5] introduced still image to recognize facial expression while Black, Cohen and et al. [6, 7] made the use of a sequence of images. Facial expression recognition from still image has less precision with respect to video sequence because a single image offers much less information than a sequence of images for expression recognition processing. The advantage of using a sequence of images over a single image is shown by Bassili shows this fact and conclude that the facial expression were more accurately recognized from dynamic images than from a single static image [8]. Table I summarized Bassili's results.

TABLE I
 THE COMPARISON OF THE ACCURACY OF FACIAL RECOGNITION BETWEEN MOVING AND STATIC DISPLAYS [8]

	Happiness	Sadness	Fear	Surprise	Anger	Disgust
Moving face	90.0	44.2	43.3	88.3	39.2	53.3
Static face	30.8	25.8	12.5	47.5	16.7	21.7

Another approaches used by many scientists is based on optical flow. Yacoob and Davis [9] used optical flow to track the motion of brows, eyes, nose and mouth. In their approach, they used a lookup table to classify six standard facial expressions, we introduced earlier. Barlett et al. [10] combined optical flow and principal component analysis

Manuscript received June 29, 2006.

A. R. Naghsh-Nilchi is an assistant professor of computer engineering at the University of Isfahan, Isfahan, 81744, Iran (phone: +98-311-7932662; fax: +98-311-7932670; e-mail: nilchi@yahoo.com).

M. Roshanzamir is a graduate student at the University of Isfahan, computer engineering department (e-mail: mroshanzamir@eng.ui.ac.ir).

(PCA) for facial expression recognition. Otsuka and Ohya [11] computed the 2D Fourier transform coefficients of optical flow and hidden Markov model was used to classify facial expressions. The performance of all of these methods depends on the reliability of optical flow estimation.

Most of worked done by many scientists used Ekman and Friesen [13] description to show facial expressions apex and classify them from static images. In dynamic images Bassili [8] describes motion patterns of the face for facial expressions. Fig. 2 shows the Bassili's observation on the motion-based cues for facial expressions.

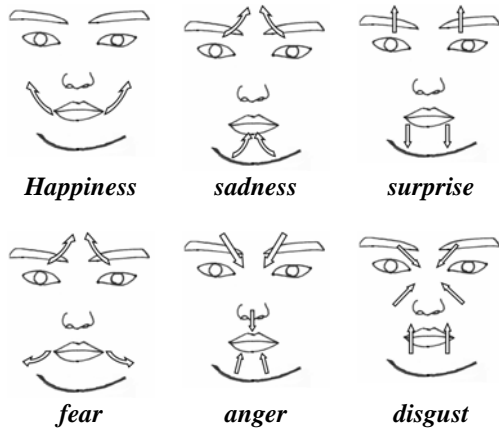


Fig. 2 The cues for facial expression as suggested by Bassili [8]

In this paper we use image sequences of expressive human faces to accurately estimate optical flows in order to recognize facial expressions. In this method, after a rough and approximate detection of the locations of facial feature, the face is segmented to six areas, according to its facial features and deformation. Then, the image motions of facial features are detected using optical flow algorithm introduced by Gautama and Vanhulle, and consequently, by analysis of the vector set, the motion vectors are extracted. This analysis is done by specifying the deformation caused by the display of emotions in the face.

Section 5 discussed that how facial expression is recognized according to face deformation and comparing these deformations with source vectors. The results of extensive testing of the recognition method are also presented.

II. FACIAL FEATURE TRACKING AND FACE SEGMENTATION

In our approach it is not necessary to track facial features exactly and approximate tracking is enough to have a sufficient result. To localize facial features approximately we use an active infra-red (IR) illumination as shown in Fig. 3(a), so that it can detect pupils under large variation in lighting and head orientations [14]. To assist pupil detection, the person's face is illuminated with an IR illuminator, which produces the dark/bright pupil effect [15, 16], as shown in Fig. 3(b). The pupil positions are then used to constrain the possible positions of other facial features. Details on pupil detection and tracking may be found in [16].

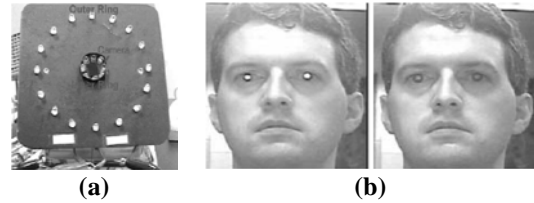


Fig. 3 (a) IR camera, (b) Light/ dark pupil effect

After tracking pupil positions by IR camera, face could be divided into six parts, as shown in Fig. 4. In this segmentation approach, first, conjunction line between two pupils is determined and drawn (Line 1). Then another line parallel to Line 1 is drawn such that it divides the area between the line and the bottom edge of the image frame in half (Line 2). A perpendicular line dividing the face in half is also drawn (Line 3). Using this proportion, line 3 often lies on nose. Since in the proposed method it does not matter for the facial features to be in exact position, a little deviation of line would cause no problem.

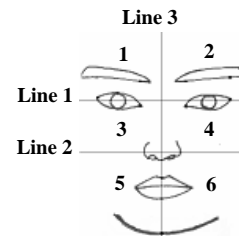


Fig. 4 Face segmentation in to six parts

In this approach, face is segmented into six parts, and each of motion vectors will be a member of one of these parts. The segmentation is done based on deformations shown in Fig. 2. Most of motion vectors that located in the same region have a relatively unique direction.

III. MOTION VECTORS COMPUTATIONS USING OPTICAL FLOW TECHNIQUES

Optical flow reflects the image changes due to motion during a time intervals. Optical flow calculation is a prerequisite for higher level processing which can solve motion problems. There are different ways to calculating optical flow. Barron et al. [18] have evaluated nine different techniques, representative for various approaches, namely the differential, matching, energy-based and phase-based ones. They have tested these algorithms on several standard image sequences.

Results show that the phase-based technique of Fleet and Jepson [19] and the differential technique of Lucas and Kanade [20] produced the more overall accurate results. Fleet and Jepson [19] have shown that phase-base approaches are more robust with respect to smooth shading and lighting variations, and more stable with respect to small deviations from image translations.

Gautama and VanHulle [17] introduced a new phase-based approach to the estimation of the optical flow field, which is based on spatially filtering the images using a bank of

quadrature pair filters. They estimate the optical flow field of an image sequence in three stages. First the image sequence is spatially filtered using a bank of quadrature pairs of Gabor filters, and the temporal phase gradient is computed, yielding estimates of the velocity component in directions orthogonal to the filter pairs' orientations. Second, the reliability of these component velocities is examined and the unreliable ones are rejected. Third, the remaining component velocities at a single spatial location are combined and a recurrent neural network is used to derive the full velocity. We use this approach to estimate the optical flow needed in our algorithm.

IV. FACIAL EXPRESSION RECOGNITION METHODOLOGY

A. Source Vector Sets

Source vectors are the collection of vectors which shows motion and deformation expressed in face due to emotion representation. To obtain source vectors, a series of facial images is chosen which its motion vectors show facial expression correctly similar to Bassili description [8]. Then incorrect vectors are eliminated and for others a pair of vector angle and its area is saved. We use face segmentation introduced in section 3 to eliminate incorrect vectors. For example when we want to save source vectors of happiness, none of the vectors should be at the upper areas.

Fig. 5 shows an example of source vectors which used for detection of happiness in a happy face. In this case, six frames of a happy faces' image are programmed as the input image sequences. The vectors are then produced and it is shown in the last image.

This procedure is run for all of basic emotions. Note that emotions have different forms in different. Also, for some of the emotions, it is possible to have more than one form in one person. An example is the two form of happiness in one person's face shown in Fig. 6.

In these cases, we should save separate source vector for each form of facial expressions. Because the number of these states is limited, it is relatively easy to create suitable source vectors for each of the forms.

B. Emotion Recognition Methodology

Emotions are specified according to the estimated similarities between the source vectors and the extracted motion vectors. In order to determine similarity between extracted vectors and source vectors, each of vectors in an area is compared with all source vectors corresponding to that area in the database.

If the difference between the angle of each of extracted vectors and any of a number of vectors is less than a predetermined value (θ), it's considered as a suitable vector. After comparing all vectors of different areas, the number of suitable vectors represents the similarity between two sets of vectors. The mathematical expression for the comparison is given in Equation (1) below:

$$SV = \sum_{k=1}^6 \sum_{i=1}^n Cnt(\sum_{j=1}^m DegDiff(v_{ik}, v_{jk})), \quad (1)$$

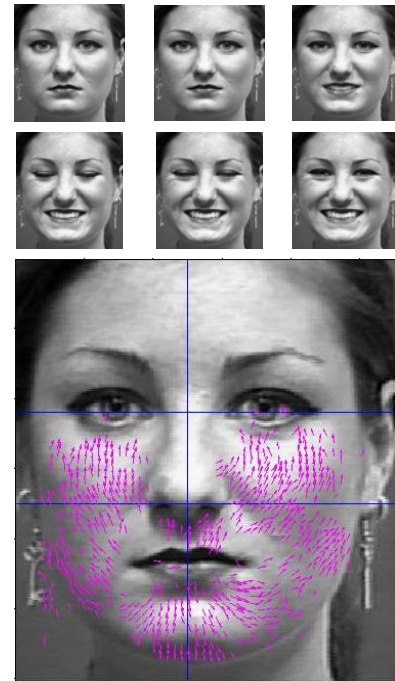


Fig. 5 An instance of source vectors

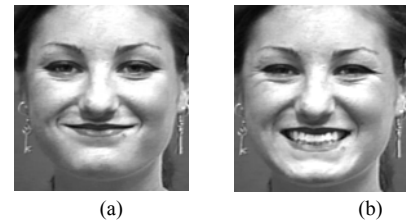


Fig. 6 Two different cases of happiness used for happiness

where SV is the number of suitable vectors, k show areas related to face segmentations, m is a number of extracted vector belong to k area, n is source vector belong to k area, v_{ik} is extracted vector of k area, v_{jk} is source vector belong to k area and $DegDiff()$ function is defined by:

$$DegDiff(v_1, v_2) = \begin{cases} 1 & |v_1 - v_2| \geq \theta \\ 0 & otherwise \end{cases} \quad (2)$$

and $Cnt()$ in Eqn (1) is defined by:

$$Cnt(x) = \begin{cases} 1 & x \geq a \\ 0 & otherwise \end{cases} \quad (3)$$

The advantage of this approach is that all vectors participate in the decision making process according to their impact on the display of emotion. A correct vector has the same direction as that of the majority of the vectors, and so it has the highest probability value. Therefore, it is considered as the suitable vector. Note that those vectors that are created as the result of

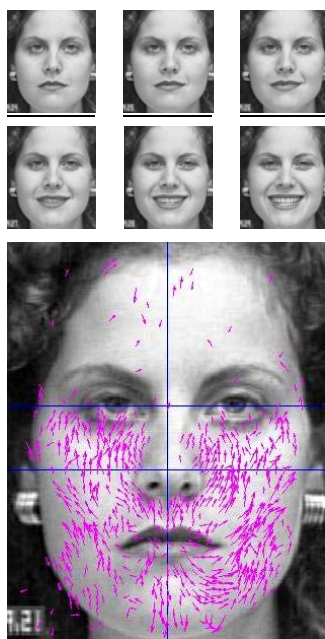


Fig. 7 Facial expression detection using source vectors matching

noise, would not be considered suitable vector because noisy vectors often have a random directions, and so would not appear as the majority.

To select suitable values for variables a and θ in the algorithm, a pre-process of running the algorithm with different values of these variables is needed. Our experimental results show that instead of selecting some single values for these parameters, it is better to take an interval for each of variables. This procedure has increased the accuracy of results.

According to experimental results obtained from different case studies, we choose the sets $A=\{15,16,\dots,20\}$ for values of a and $\Theta=\{10,11,\dots,15\}$ for θ , the number of suitable vectors are calculated for each of $(a,\theta)\in A\times\Theta$.

Then, the extracted vectors are compared with all source vector sets corresponding to the six basic emotions in the database. A higher number represents the higher level of similarity between the computed vectors and the specific set of source vectors representing a unique emotion. Based on the computed SV values, the degree of similarity between the image and all the basic emotions are determined and the one with highest degree of similarity could be identified as the detected emotion

Fig. 7 shows a case of this comparison. The number of suitable vectors achieved for happiness, surprise, anger, fear, disgust and sadness were 197, 37, 0, 44, 46 and 32, respectively. It's clear that the number of suitable vectors for happiness is more than that of other emotions.

V. SOME EXPERIMENTAL RESULTS

Our method is examined on the Cohn-Kanade Facial Expression Database [3]. This database includes approximately 1000 image sequences from over 100 different

subjects ranged in age from 18 to 30 years. Sixty-five percent were female, 15 percent were African-American, and three percent were Asian or Latino. Images are taken from frontal view and subjects began each display from a neutral face and end to one of basic emotions.

Table II depicts the result of applying our algorithm on the image sequences shown in Fig. 8. Left column shows the facial expression and other cells show the numbers of suitable vectors earned for each of emotions. Table III shows the results of applying the algorithm on all cases exist in the Cohn-Kanade Facial Expression Database.

One of the property of this method is that it is not determined an emotion absolutely but shows the amount of similarity it has with every basic emotion. Also, our algorithm can determine facial expression with only a few number of frames in image sequence (three or four frames). It means that even if most of frames are lost, this method can still work. Table IV shows the results of applying our method only on first, middle and last frames of different facial expression image sequence.

The results on only three frames (first, middle and last) of the sequence shows that our algorithm is capable of recognizing facial expression with average rate of success of about 83.3%. This shows that our algorithm has a very good performance even for limited number of available frames in an image's sequence.

Slowness of optical flow algorithm, difficulty of the emotion detection in present of significant head motion and the problems related to the optical flow techniques, when the light is unfixed, are the most disadvantages of our algorithm.

VI. CONCLUSION

In this paper, we introduced an efficient and new algorithm for facial motion detection based on Gautama and VanHalle optical flow technique to extract the necessary motion vectors, and consequently, determining the relative facial expression.

In our algorithm, the image frames are segmented into six parts and then, the degree of similarities' between the calculated vectors and the source vectors, which is created and made available in a database, are computed. The strongest degree of similarity determines the image's facial emotion.

The algorithm is examined using the images' sequences available in the Cohn-Kande AU-Coded facial expression database. The experimental results for full image sequences and also for the limited number of frames in the sequences (first, middle, last) are reported.

In comparison with other method, our procedure has a very good rate of success. One of the most important advantages of our method is that it is not necessary to determine exact facial feature locations and only the approximate values are sufficient. In addition, the experimental results show that our method could correctly recognize the facial expressions in 94% of case studies. The results also show that only a few number of image frames (three frames) are sufficient to detect facial expressions with rate of success of about 83.3%. These advantages should make a high performance facial recognition based on optical flow techniques, feasible.



Fig. 8 Some instances of emotion representation for happiness, sadness, fear, surprise, anger and disgust respectively

TABLE II
 RESULTS OF PROPOSED SYSTEM ON IMAGE SEQUENCES SHOWN IN FIG. 8

<i>represented emotions</i>	<i>Average of suitable vectors obtained for each emotion</i>					
	<i>happiness</i>	<i>sadness</i>	<i>fear</i>	<i>surprise</i>	<i>anger</i>	<i>disgust</i>
<i>happiness</i>	352	88	17	144	16	160
<i>sadness</i>	89	140	50	104	2	10
<i>fear</i>	93	144	255	115	17	29
<i>surprise</i>	152	255	83	329	22	141
<i>anger</i>	10	26	84	6	231	51
<i>disgust</i>	287	249	2	214	124	479

TABLE III
 CORRECTION RATE OF FACIAL EXPRESSION RECOGNITION ON COHN-KANADE DATABASE

<i>happiness</i>	<i>sadness</i>	<i>fear</i>	<i>surprise</i>	<i>anger</i>	<i>disgust</i>
100	89.5	86.03	96.8	98.48	93.9

TABLE IV
 CORRECTION RATE OF FACIAL EXPRESSION RECOGNITION ON COHN-KANADE DATABASE (FIRST, MIDDLE, AND THE LAST FRAME ARE USED)

<i>happiness</i>	<i>sadness</i>	<i>fear</i>	<i>surprise</i>	<i>anger</i>	<i>disgust</i>
90.2	85.1	73.7	80.1	82.1	88.4

ACKNOWLEDGMENT

The authors would like to thank Jeffrey Cohn of Robotics Institute at the Carnegie Mellon University for his kind permission to use Cohn-Kanade AU-Coded Facial Expression Database.

REFERENCES

- [1] D. Darwin, "The Expression of Emotion in Man and Animals," *John Murray, 1872, reprinted by University of Chicago Press, 1965.*
- [2] E.W. Young and H.D. Ellis (Eds.), "Handbook of Research on Processing," *Elsevier Science Publishers, 1989.*
- [3] Cohn-Kanade AU-Coded Facial Expression Database. Available: http://vasc.ri.cmu.edu/idb/html/face/facial_expression/
- [4] G.J. Edwards, T.F. Cootes, and C.J. Taylor, "Face Recognition Using Active Appearance Models," *Proc. European Conf. Computer Vision*, vol. 2, pp. 581-695, 1998.
- [5] H. Hong, H. Neven, and C. von der Malsburg, "Online Facial Expression Recognition Based on Personalized Galleries," *Proc. Int'l Conf. Automatic Face and Gesture Recognition*, pp. 354-359, 1998.
- [6] M.J. Black and Y. Yacoob, "Recognizing Facial Expressions in Image Sequences Using Local Parameterized Models of Image Motion," *Int'l J. Computer Vision*, vol. 25, no. 1, pp. 23-48, 1997.
- [7] J.F. Cohn, A.J. Zlochower, J.J. Lien, and T. Kanade, "Feature-Point Tracking by Optical Flow Discriminates Subtle Differences in Facial Expression," *Proc. Int'l Conf. Automatic Face and Gesture Recognition*, pp. 396-401, 1998.
- [8] J.N. Bassili, "Emotion recognition: The role of facial movement and the relative importance of upper and lower areas of face," *Journal of Personality and Social Psychology*, Vol. 37, 2049-2059, 1979.
- [9] Y. Yacoob, L. Davis, "Recognizing human facial expressions from long image sequences using optical flow," *IEEE Trans. Pattern Anal. Machine Intell.* 16 (6), 636-642. 1994.
- [10] M. Barlett, P. Viola, T. Sejnowski, L. Larsen, J. Hager, P. Ekman, "Classifying facial action," In: *Touretzky, D., Mozer, M., Hasselmo, M. (Eds.), Advances in Neural Information Processing Systems*. MIT Press, Cambridge, MA. 1996.
- [11] T. Otsuka, J. Ohya, "Recognizing multiple persons facial expressions using HMM based on automatic extraction of significant frames from image sequence," In: *Proc. Int. Conf. On Image Processing*, pp. 546-549. 1997.
- [12] X. Chen, T. Huang, "Facial expression recognition: A clustering-based approach," In: *Pattern Recognition Letters, Elsevier Science*, pp 1295-1302, 2003.
- [13] P. Ekman, W. Friesen, "Unmasking the face," *Prentice-Hall, 1975.*
- [14] Y. Zhang, Q. Ji, "Facial Expression Understanding in Image Sequences Using Dynamic and Active Visual Information Fusion," *Proceedings of the Ninth IEEE International Conference on Computer Vision (ICCV 2003)*, 2003.
- [15] C. Morimoto, D. Koons, A. Amir, and M. Flicker, "Framerate pupil detector and gaze tracker," In *Proc. of the IEEE ICCV99 Frame-Rate Workshop*, (Kerkyra, Greece), Sep. 1999.
- [16] J.Z. Zhu, Q. Ji, K. Fujimura, and K. Lee, "Combining Kalman filtering and mean shift for real time eye tracking under active IR illumination," In *Proc. Int'l Conf. Pattern Recognition*, Aug. 2002.
- [17] T. Gautama, M. M. VanHulle, "A phase-based approach to the estimation of the optical flow field using spatial filtering," *IEEE Trans. Neural Networks*, VOL. 13(5), September 2002.
- [18] J.L. Barron, D.J. Fleet, and S. Beauchemin, "Performance of optical flow techniques," *International Journal of Computer Vision*, vol. 12(1), pp. 43-77, 1994.
- [19] D.J. Fleet, and A.D. Jepson, "Computation of Component Image Velocity from Local Phase Information," *Int. J. Comput. Vision*, vol. 5(1), pp. 77-104, 1990.
- [20] B.D. Lucas, and T. Kanade, "An Iterative Image Registration Technique with an Application to Stereo Vision," *7th International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 674-679, 1981.