

Geometric-based Feature Extraction and Classification for Emotion Expressions of 3D Video Film

Salwa A. Al-agma

Department of Electrical Engineering, University of Technology, Baghdad, Iraq

Email: salwa_alagma2009@yahoo.com

Hilal H. Saleh and Rana F. Ghani

Department of Computer Sciences, University of Technology, Baghdad, Iraq

Email: {hhsrq888, ranafghany}@yahoo.com

Abstract—Feature extraction is the most significant step in the operation of emotion expressions recognition. Discrimination operation of emotion expressions has gained the attention of many researchers in the field of pattern recognition because of its significant impact on the various aspects of applications, especially in the application of human-computer interaction, both for the image or to the video. Based on pattern recognition theory, the process of facial expression recognition can be divided into features extraction operation and classification operation. In this paper, the geometric-based features extraction operation is used for extracting the local characteristics (landmarks) of a set of emotion expressions (anger, happiness, sadness, surprise) for images of BOSPHERUS database as training stage, then the classification operation is done by using of the threshold method (Euclidean distance) between the distances of neutral image and the expression image. The trained system is used for feature extraction and classification for 3D video film (stereoscopic) as testing stage. This method is implemented on 40 3D video films that were recorded, 10 video films for each expression of the four basic emotion; the ratio of discrimination is 85%.

Index Terms—local feature extraction, 3D video film, basic emotion expression, emotion classification, 3D video classification

I. INTRODUCTION

Feature extraction process is essentially the facial expression representation, which transforms the genuine information from a low-level 2D pixel or 3D vertex based representation, into a higher-plane representation of the face regarding its landmarks, spatial setup, shape, appearance and/or movement. The dimensionality of the original input facial data is usually reduced by features extraction [1].

Based on pattern recognition theory, the process of facial expression recognition can be divided into features extraction operation and classification operation. Many methods have been suggested, involving principal component analysis, linear discriminant analysis, nonparametric discriminant analysis, optical flow, Fisher weight maps, and local binary pattern [2].

Many researchers have worked on geometric-based as well as holistic-based feature extraction approaches.

This paper presents the proposed technique of geometric-based features extraction and classification for emotion recognition operation with four basic emotion expressions (anger, happiness, sadness, surprise), which depends on finding the local features (landmarks) of the face for the neutral image and the expression image for each frame of 3D video film (stereoscopic).

II. GEOMETRIC-BASED FEATURE EXTRACTION

In geometric-based approach, the local features (local statistics and locations) include mouth, eyes, eyebrows, and nose are at first extracted from face images, as shown in Fig. 1, which is used for classification [3]. The most successful of geometric-based methods is Active Appearance graph Models (AAM) [4].

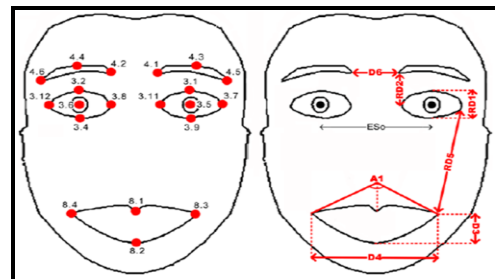


Figure 1. Landmarks features [5]

Most ways to deal with facial expression examination endeavor to understand a little arrangement of prototypic passionate facial expressions, i.e., dread, misery, repugnance, outrage, shock, and satisfaction [6] and [7]. This practice might be contrived from the exploration of Darwin [8], and all the more as of late of Ekman [9], who recommended that fundamental feelings have relating prototypic expressions.

Facial Action Coding System (FACS) is the best known among a few different techniques for acknowledgment of facial signals and most generally utilized as a part of mental exploration [10, 11]. FACS portrays that the adjustments in the facial expression are as far as 44 Different Action Units (AUs), each of which is anatomically identified with the choking of either a particular facial muscle or an arrangement of facial muscles. FACS gives the guidelines to AU location in a face picture, alongside the meaning of different AUs [12].

III. THE PROPOSED TECHNIQUE

In this work, geometric-based feature extraction and classification are implemented in two stages: geometric-based features extraction and classification for training level, and geometric-based features extraction and classification for testing level.

A. Geometric-based Features Extraction as Training Level

Local features extraction is accomplished for an arrangement of BOSPHERUS database images as training information; this is performed via preparing this framework to discover the distances between particular focuses which are allocated manually on the particular territories of face of database images. In this work, there are 14 specific points that must be taken from landmarks of face and 7 distances for each face of neutral images and expression images, as illustrated in Fig. 2:

- 1) 2 points represent the inner eyebrows points. Line between inner eyebrows (Ib), represents the distance between inner eyebrows.
- 2) 4 points: 2 points between the center of left eye and

center of left eyebrow, 2 points between the center of right eye and center of right eyebrow.

Line between the center of left eye and center of left eyebrow (Ib1) represents the distance between left eye and left eyebrow. Line between the center of right eye and center of right eyebrow (Ib2) represents the distance between right eye and right eyebrow.

3) 2 points represent the lip vertical distance. Line between lip vertical points (Lpv) represents the distance between lip vertical points.

4) 2 points represent the lip horizontal distance. Line between lip horizontal points (Lph) represents the distance between lip horizontal points.

5) 4 points: 2 points between the center of left eye and the point of left corner of mouth, 2 points between the center of right eye and the point of right corner of mouth. Line between the center of left eye and the point of left corner of mouth (Eml) represents the distance between left eye and left angle of mouth. Line between the center of right eye and point of right angle of mouth (Emr) represents the distance between the right eye and right angle of mouth.

B. Classification of Geometric-based Features of Training Level

After the operation of features extraction, analysis of all the features (in this operation, decision is needed for deciding which feature is useful in classification operation). Sequential Forward Selection (SFS) algorithm is used for selecting the suitable local features for the images of database, which serves the classification operation.

The result of SFS selection in this step are 6 features: Ib1, Lpv1, Lph1, for neutral images, Ib2, Lpv2, and Lph2 for expression images, which will be used in the classification operation for emotion expression recognition.

In the training level, threshold classification operation is used to train this system for decision making about the category of each image, this operation depends on finding the proportion of the amount of changes in distance between neutral image and expression image by using (Ib1, Lpv1, Lph1, Ib2, Lpv2, and Lph2) local features in the classification operation. Threshold classification operation involves the following:

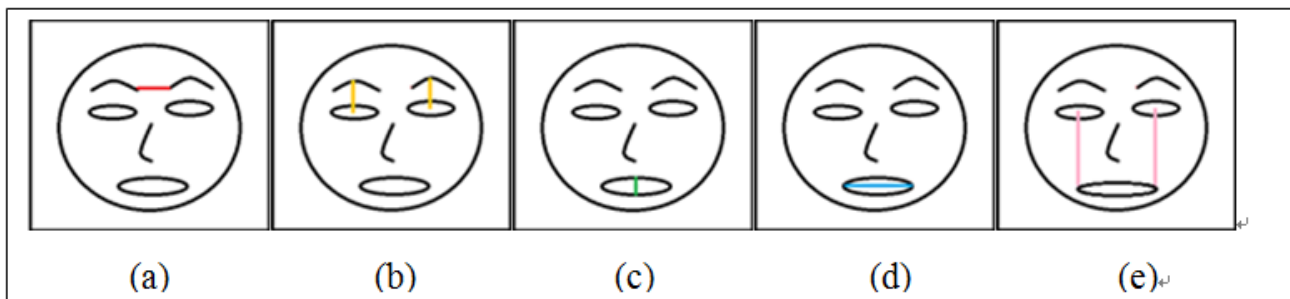


Figure 2. Local features (points and distances): (a) is Ib, (b) is Ib1 and Ib2, (c) is Lpv, (d) is Lph, and (e) is Eml and Em

```

1) for i: = 1 to No. of images of database do
2) calculate:
   Ib ← |Ib1 – Ib2|           // amount of change for
inner eyebrow
   Lpv ← |Lpv1 – Lpv2|        // amount of change for
vertical lip
   Lph ← |Lph1 – Lph2|        // amount of change for
horizontal lip
/* using of the Euclidean distance between the distances, in
order to find the amount of changes in distances between
the neutral image and the expression image */
3) Ib ← fix (Ib/10)           // normalizing value of Ib
4) if Lpv > 0 then
   Lpv ← fix (Lpv/10)         // normalizing value of
Lpv
5) end if                      // Note: choose value 10
depending on the size of
6) if Lph > 0 then             image and on the value
of distances */
   Lph ← fix (Lph/10)         // normalizing value of Lph
7) end if
8) if ((Ib<0) & (Lpv<0)) & ((Ib<Lpv) & (Ib<Lph)) then
   Decision ← ‘Anger’         // decision making for anger
expression
   else if (Lph>0) & ((Lph>Ib) & (Lph>Lpv)) then
   Decision ← ‘Happy’         // decision making for happy
expression
   else if (Lpv>0) & ((Lpv>Ib) & (Lpv>Lph)) then
   Decision ← ‘Surprise’      // decision making for surprise
expression
   else
   Decision ← ‘Sad’           // decision making for sad
expression
9) end if
10) end for i

```

C. Geometric-based Technique as Testing Level

The main objective of this work is to extract features of the sequences frames of 3D video film. Local features extraction of intensity image is done for frames of 3D video film; this is done by selecting some of these frames as neutral frames and expression frames. The trained framework finds the distances between particular focuses which are allocated manually on the particular areas of face of the 3D frames after face detection operation. There are 6 specific points that must be taken from landmarks of face:

- 2 points representing the inner eyebrow distance Ib.
 - 2 points representing the lip vertical distance Lpv.
 - 2 points representing the lip horizontal distance Lph.
- Feature extraction operation for 3D video film:

```

1) for i: = 1 to No. of 3D video frames do
2) for j: = 1 to first 10 frames of video as neutral frames do
3) if prompt == press ‘y’ then // choose suitable neutral
frame
4) for k: = 1 to No. of 6 points detected manually from each
face do

```

```

5) [x, y] ← ginput (k)       // get x and y dimensions for
each point
6) A (k) ← [x, y]           // save each dimension in array
7) end for k
8) calculate:
   Ib1= A (k1) – A (k2)      // inner eyebrow distance of
neutral image
   Lpv1= A (k3) – A (k4)     // vertical lip distance of
neutral image
   Lph1= A (k5) – A (k6)     // horizontal lip distance of
neutral image
9) else prompt == press ‘enter’ // permit for another
suitable neutral frame
10) end if
11) end for j
12) for h: = 15 to h+10 do
13) if prompt == press ‘y’ then // choose expression
frame
14) for k: = 1 to No. of 6 points detected manually from
each face do
15) [x ,y] ← ginput (k)     // get x and y dimensions for
each point
16) A(k) ← [x, y]           // save each dimension in array
17) end for k
18) calculate:
   Ib2= A(k1) – A(k2)       // inner eyebrow distance of
expression image
   Lpv2= A(k3) – A(k4)      // vertical lip distance of
expression image
   Lph2= A(k5) – A(k6)      // horizontal lip distance of
expression image
19) else prompt == press ‘enter’
20) end if
21) end for h
22) end for i

```

D. Classification for Geometric-based Features of Testing Level

After geometric-based feature extraction for some of frames of 3D video film as testing level, these local features (distances between the local points in each face) are used in the classification operation to make a decision about the class of each frame, edge characterization operation (threshold), which relies on upon finding the measure of changes between neutral frame and expression frame:

```

1) for i: = 1 to No. of 3d video frames do
2) calculate:
   Ib ← |Ib2 – Ib1| // amount of change for inner eyebrow
Lpv ← |Lpv2 – Lpv1| // amount of change for vertical lip
Lph ← |Lph2 – Lph1| // amount of change for horizontal
lip
3) Ib ← fix (Ib/5)          (normalizing value of Ib)
4) if Lpv > 0 then
   Lpv ← fix (Lpv/5)        // normalizing value of Lpv
5) end if
6) if Lph > 0 then

```

```

    Lph ← fix (Lph/10)    // normalizing value of Lph
7) end if
8) if ((Ib<0) & (Lpv<0)) & ((Ib<Lpv) & (Ib<Lph)) then
    Decision ← 'Anger' // decision making for anger
expression
    else if (Lph>0) & ((Lph>Ib) & (Lph>Lpv)) then
    Decision ← 'Happy' // decision making for happy
expression
    else if (Lpv>0) & ((Lpv>Ib) & (Lpv>Lph)) then
    Decision ← 'Surprise' // decision making for surprise
expression
    else
    Decision ← 'Sad' // decision making for sad
expression
9) end if
10) end for i

```

IV. EXPERIMENTAL RESULTS AND ANALYSIS

The proposed technique is used for local feature extraction and classification for facial expression in 3D video frames. This proposed system is implemented on well-known 3D facial expression database (BOSPHOPUS database) as training level. Results of actualizing geometric-based technique for feature extraction for some images of database, as an example, detecting the 6 points manually on specific area (2 points on inner eyebrows, 2 points on lip vertical, and 2 points on lip horizontal) of the face for each neutral image and expression image, as illustrated in Fig. 3.

Table I. introduces the distances between focuses for eyebrow distance (Ib1, Ib2), lip vertical distance (Lpv1, Lpv2), lip horizontal distance (Lph1, Lph2), for every neutral image and expression image as local features for each image of database. Table (I) presents also the category of every image of database relying upon the proportion of the measure of changes in distance between neutral image and expression image.



Figure 3. Focuses detected for neutral image and expression image: (a) anger, (b) happy, (c) sad, (d) surprise.

TABLE I. LOCAL FEATURES AS DISTANCES AND PROPORTION OF DISTANCES CHANGING OF SOME IMAGES OF DATABASE.

Images of Database	Local Features as Distances							Ratio of Changing in Distances			Classification
	No.	Images	IB1	IB2	LPV1	LPV2	LPH1	LPH2	IB	LPV	LPH
1	I1	43	21	15	8	63	60	-22	-7	-3	Anger
2	I2	46	48	14	22	64	87	2	8	23	Happy
3	I3	47	43	15	14	65	68	-4	-1	3	Sad
4	I4	48	50	14	28	63	68	2	14	5	Surprise
5	I5	32	14	15	12	72	77	-18	-3	5	Anger
6	I6	33	39	17	22	71	93	6	5	22	Happy
7	I7	30	32	16	14	71	79	2	-2	8	Sad
8	I8	34	37	16	31	69	65	3	15	-4	Surprise
9	I9	38	12	15	10	69	62	-26	-5	-7	Anger
10	I10	38	41	13	18	67	94	3	5	27	Happy
11	I11	36	33	12	14	70	69	-3	2	-1	Sad
12	I12	40	41	15	37	71	64	1	22	-7	Surprise
13	I13	44	10	13	11	65	57	-34	-2	-8	Anger
14	I14	39	40	16	19	58	82	1	3	24	Happy
15	I15	39	38	17	13	63	64	-1	-4	1	Sad
16	I16	41	40	15	32	70	64	-1	17	-6	Surprise

Results of ratio of changes in distances are shown in Table I. for some images of database. The decision making depends on values of (Ib, Lpv, Lph), when the values of (Ib, Lpv) is less than zero and the value of (Ib) is the smallest one between them, the decision is Anger, as shown in rows (1, 5, 9, 13). When the value of (Lph) is greater than zero and it is the greater one between them, the decision is Happy, as shown in rows (2, 6, 10, 14). When the value of (Lpv) is greater than zero and it is the greater one between them the decision is Surprise, as shown in rows (4, 8, 12, 16). Otherwise the decision is Sad, as shown in rows (3, 7, 11, 15).

Results of executing geometric-based technique for feature extraction for frames of some of 3D video, as an example, detecting the 6 points manually on specific area (2 points on inner eyebrows, 2 points on lip vertical, 2 points on lip horizontal) of the face for each neutral frames and expression frames (for one side of frame, left or right), as illustrated in Fig. 4.

Table II. introduces the distances between focuses for eyebrow distance (Ib1, Ib2), lip vertical distance (Lpv1, Lpv2), lip horizontal distance (Lph1, Lph2), for each neutral frame and expression frame as local features for each frame of 3D video film. Table (2) presents also the class of each frame contingent upon the proportion of the measure of changes in distance between neutral frame and expression frame.



Figure 4. Focuses detected for neutral frames and expression frames: (a) anger, (b) happy, (c) sad, (d) surprise.

TABLE II. LOCAL FEATURES AS DISTANCES AND PROPORTION OF DISTANCES CHANGING OF SOME 3D VIDEO FRAMES.

Images of Database		Local Features as Distances						Ratio of Changing in Distances			Classification
No.	Video	IB1	IB2	LPV1	LPV2	LPH1	LPH2	IB	LPV	LPH	Expression
1	V1	18	9	21	18	23	24	-9	-3	1	Anger
2	V2	12	17	15	20	23	33	5	5	10	Happy
3	V3	18	15	15	12	21	24	-3	-3	3	Sad
4	V4	15	15	14	38	21	20	0	24	-1	Surprise
5	V5	39	15	30	27	50	45	-24	-3	-5	Anger
6	V6	29	24	19	24	48	67	-5	5	19	Happy
7	V7	33	32	20	17	48	45	-1	-3	-3	Sad
8	V8	33	42	23	117	53	44	9	94	-9	Surprise
9	V9	18	9	14	9	20	20	-9	-5	0	Anger
10	V10	15	18	12	17	17	32	3	5	15	Happy
11	V11	17	16	11	10	22	23	-1	-1	1	Sad
12	V12	15	18	11	48	20	20	3	37	0	Surprise
13	V13	18	6	15	9	33	30	-12	-6	-3	Anger
14	V14	12	14	14	18	30	47	2	4	17	Happy
15	V15	14	18	18	15	32	33	4	-3	1	Sad
16	V16	17	21	17	43	33	27	4	26	-6	Surprise

Results of ratio of changes in distances are shown in Table II. for some frames of 3D video film. The decision making depends on values of (IB, LPV, LPH), when the values of (IB, LPV) are less than zero and the value of (IB) is the least one between them, the decision is Anger, as shown in rows (1, 5, 9, 13). When the value of (LPH) is greater than zero and it is the greater one between them, the decision is Happy, as shown in rows (2, 6, 10, 14). When the value of (LPV) is greater than zero and it is the greater one between them, the decision is Surprise, as shown in rows (4, 8, 12, 16). Otherwise the decision is Sad, as shown in rows (3, 7, 11, 15).

Number of testing 3D video film for 4 expression classes are 40 videos, the number of classified videos is 34 videos and the number of misclassified videos is 6, the recognition rate for all 3D video film is 85 %. The perplexity matrix is given in Table III.

TABLE III. PERPLEXITY MATRIX FOR GEOMETRIC-BASED FEATURES CLASSIFICATION METHOD OF INTENSITY FRAME

	Anger	Happy	Sad	Surprise
Anger	6	0	2	2
Happy	0	10	0	0
Sad	0	0	8	2
Surprise	0	0	0	10

V. CONCLUSIONS

In this paper, the proposed strategy is exhibited for 3D video features extraction and classification. The local features of geometric-base method (Ib, Lpv, and Lph) are straightly isolated (linearly) for images database as training images and for frames of 3D video as testing images for 4 feeling expressions. These features are figured by utilizing the basic distinction technique (Euclidean Distance) and amplify the objective function for classification.

The local features of geometric-base technique (Ib1, and Ibyr) are repeated features which give the same performance for LPV feature. (Eml, and Emr) features do not discriminate any expression of 4 emotion expressions, therefore, (Eml, Emr, Ib1, and Ibyr) has been excluded in the classification process.

The threshold classification method for geometric-based features (Ib, Lpv, and Lph) of image database during training level and successive frames (intensity frames) of 3D video films at testing level depends on finding the proportion of distance (the measure of changes between neutral image and expression image) rather than the distance itself.

Geometric-based features classification strategy for intensity frames gives the classification rate for anger as 60 %, happy as 100%, sad as 80 %, surprise as 100 %; at long last, the recognition rate for all 3D video film is 85 %.

Misclassification rate for Geometric-based features classification for intensity frames is 15%; the perplexity is

between sad feeling expression and anger with surprise feeling expressions, which prompts this proportion of misclassification.

In future work, this work is to be developed by making the process of finding the landmarks of geometric-based features classification method, automatically not manually. Visual information in 3D video films in this work is dealing with frontal view, towards the integration of the work there is need to deal with poses of the face for multiple angles.

REFERENCES

- [1] E. Cambria, I. Hupont, A. Hussain, E. Cerezo, and S. Baldassarri, "Sentic avatar: Multimodal affective conversational agent with common sense," *Toward Autonomous, Adaptive, and Context-Aware Multimodal Interfaces: Theoretical and Practical Issues*, Springer Berlin Heidelberg, pp. 81-95, 2011.
- [2] M. Albert, *Biometrics – Unique and Diverse Applications in Nature, Science, and Technology*, In – Tech Open Access Publisher, 2011.
- [3] H. F. Huang and S. C. Tai, "Facial expression recognition using new feature extraction algorithm," *Electronic Letters on Computer Vision and Image Analysis, Published by Computer Vision Center*, vol. 11, no. 1, pp. 41 – 54, 2012.
- [4] K. Roy, P. Bhattacharya, and C. Y. Suen, "Iris recognition using shape – guided approach and game theory", *Pattern Analysis and Application, Springer – Verlag*, vol. 14, no. 4, pp. 329 – 348, 2011.
- [5] H. Shin, S. D. Kim, and H. C. Choi, "Generalized elastic graph matching for face Recognition," *Pattern Recognition Letters, Elsevier*, vol. 28, no. 9, pp. 1077–1082, 2007.
- [6] M. J. Black and Y. Yacoob, "Recognizing facial expressions in image sequences using local parameterized models of image motion," *International Journal of Computer Vision, Kluwer Academic Publishers*, vol. 25, no. 1, pp. 23–48, 1997.
- [7] M. Pantic and L. J. M. Rothkrantz, "Expert system for automatic analysis of facial expression," *Image and Vision Computing, Elsevier*, vol. 18, no. 11, pp. 881–905, 2000.
- [8] C. Darwin, *The Expression of the Emotions in Man and Animals*, 1st edition, University of Chicago Press, 1872, 1965.
- [9] M. Lewis, J. M. Haviland, and L. F. Barrett, *Handbook of Emotions*, 3rd Edition, Guilford Press, New York, 2008.
- [10] P. Ekman and W. Friesen, "Facial action coding system: A technique for the measurement of facial movement," *Consulting Psychologists Press*, Palo Alto, 1978.
- [11] K. Scherer and P. Ekman, *Handbook Methods in Non – Verbal Behavior Research*, Cambridge University Press, New York, pp. 45 – 135, 1982.
- [12] S. Gangwar and K. Kumar, "3D face recognition based on extracting PCA methods," *International Journal of Engineering Research and Applications*, vol. 2, no. 2, pp. 693 – 696, 2012.



Salwa A. Alagha is currently a lecturer in University of Technology Department of Electrical Engineering and a postgraduate student for PhD degree in University of Technology Department of Computer Science. She received her Msc at University of Technology Department of Computer Science in 2003. Her research interest lie in Image and Video processing.



Hilal H. Saleh is currently a lecturer in University of Technology Department of Computer Science. He received his PhD at Higher Institute Sofia - Bulgaria in 1981. His research interest lie in Information Systems and Data Security.

Hilal H. Saleh is currently a lecturer in University of Technology Department of Computer Science. He received his PhD at Higher Institute Sofia - Bulgaria in 1981. His research interest lie in Information Systems and Data Security.

Rana F. Ghani is currently a lecturer in University of Technology Department of Computer Science. She received her PhD at University of Technology Department of Computer Science in 2006. Her research interest lies in Network Security.