# WHAT HAPPENS IN FACE DURING A FACIAL EXPRESSION?
USING DATA MINING TECHNIQUES TO ANALYZE FACIAL EXPRESSION MOTION VECTORS

**A SHORT ELICIATION OF PAPER (MY VERSION) – LINK**

**-Arshia Sali**

# INTRODUCTION

- **Facial expression** is one of the fastest and most efficient ways in communication. It prevails over words in human communication.
- Although it is easy for humans to recognize facial expressions, **Automatic Recognition** remains difficult for machines.
- **Human-computer interaction** is probably the most important application in automatic facial expression recognition.
- Other fields such as date-driven animation, psychology researches, medicine, security, education and distance learning, customer satisfaction and video conferences can use the results of these researches.
- They can even be used in the (re)production of artificial emotions.

- Machines can **analyze the changes in face** during facial expression presentation.
- The ideal automatic facial expression recognition system must be completely automatic, person independent and robust to any environmental condition.
- To do that, a three stage process must be done.
  - Face Detection
  - Facial Feature Extraction
  - Facial Expression Classification
- This research only focused on facial feature extraction and facial expression classification.
- **Assumption**: Face images are already available in suitable conditions.

# DATA MINING TECHNIQUES

- **Data mining** is the process of exploration and extraction of the knowledge from the data. It involves learning in a practical sense.

- The word **learning** means that these techniques learn from the changes appearing in the data in a way that improves their performance in the future. Thus, learning is tied to performance enhancement.

- Based on this learning process, the learning techniques can be employed to map data into decision model in order to produce the predicting output from the new data. This decision model is called **classifier**.

# I. MOTION VECTOR EXTRACTION
## (DATA COLLECTION PHASE)

- A facial expression results in some temporary shift of facial features because of facial muscle movements.

- Motion vectors that show facial deformation were extracted from image sequence of facial expression.

- In this research, **Optical flow algorithm** was used to extract motion vectors.

- Optical flow algorithm has some weaknesses. For example, luminance must not change while the image sequence is created. Otherwise, this algorithm is not able to extract motion vector correctly. However, as the changes in face because of facial expression happens in a very short time, commonly luminance change was not happen. So, this weakness of optical flow is not a critical problem in this algorithm.

## METHODOLGY

- It is based on tracking points across multiple images.
- In this research, at most, 8-image sequence was used for each test. We used image sequences of faces in a frontal view displaying various facial expressions of emotion.

**Preprocessing**

- The images were converted into gray scale images and then segmented the face in a rectangular bounding. Extra parts of images were cut. It resizes image dimensions to about 280 × 330 pixels.
- Assumption : Facial expression sequences always started with a neutral facial expression and ended with the apex of a facial expression. The faces were without hair and glasses and no rigid head movement could be acceptable for the method to work properly.

(a) shows the image sequence of disgust

(b) shows the image sequence of happiness

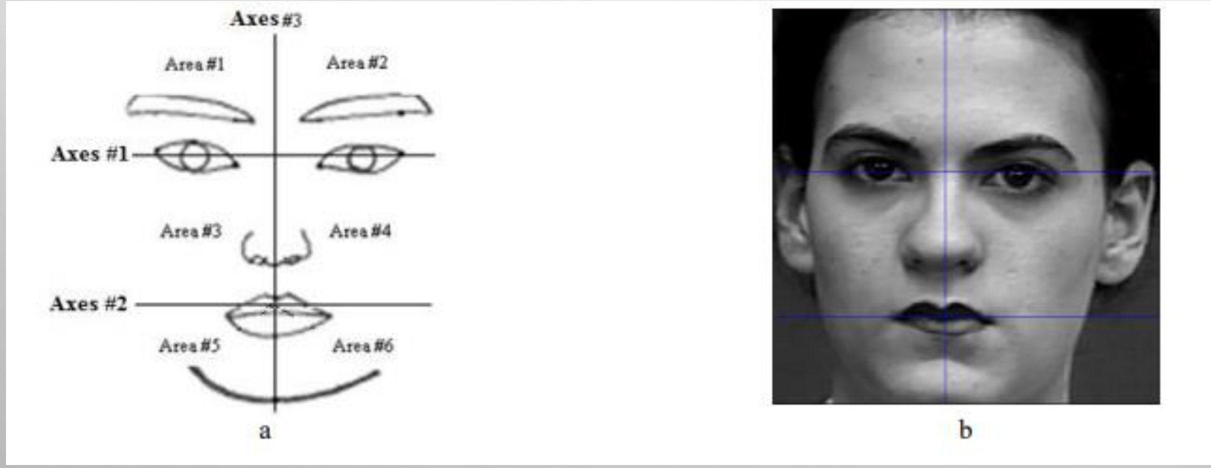
- However, there are a few cases the subject has rigid head movement while they show an expression. This disturbs the extracted motion vectors and will mislead the classification algorithms. So, we eliminated these types of extracted motion vectors.
- Different modifications have been done on optical flow algorithm. To minimize the effects of luminance variation and inaccuracies in facial point tracking, Gautama-VanHulle optical flow method has been used.
- It was claimed that this method is less sensitive to luminance variation and has very good efficiency in motion vector extraction.

Two examples of using optical flow algorithm on image sequences of facial expression have been presented.

**Figure 1**: Disgust motion vectors extracted from disgust image sequence
**Figure 2**: Happiness motion vectors extracted from happiness image sequence.

# FACE SEGMENTATION

- To classify extracted motion vectors into six basic emotions, the face was divided into six parts as shown in the figure a.
- At first, the position of pupils and mouth must be located.
- Their initial location can be detected in the first frame of the input face image sequence. As it is clear in Figure b, axis number 1 connects two pupils.
- Axis number 3 is perpendicular to axis number 1 and divides it into two equal parts.
- Axis number 2 shows the mouth position. It is not necessary to specify axes location precisely. Approximation of axis locations is enough.




Face segmentation into 6 areas in
a) A scheme
b) A real face image

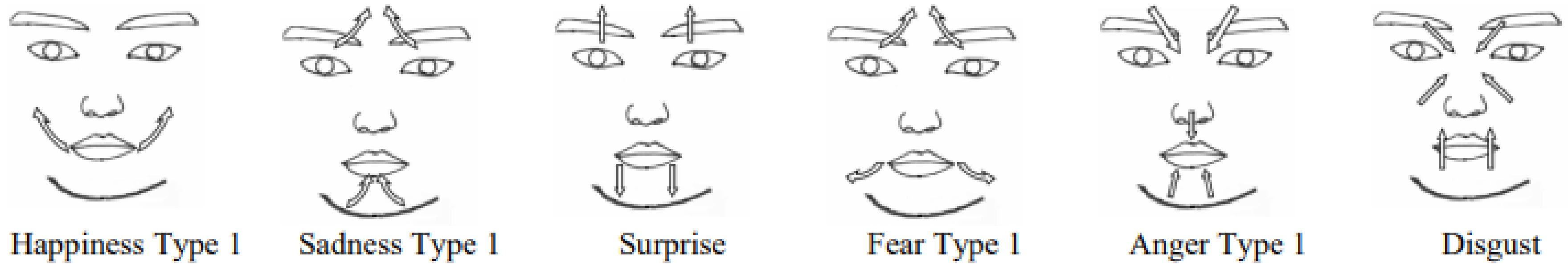



Figure 5. Bassili description of face deformation in each emotion

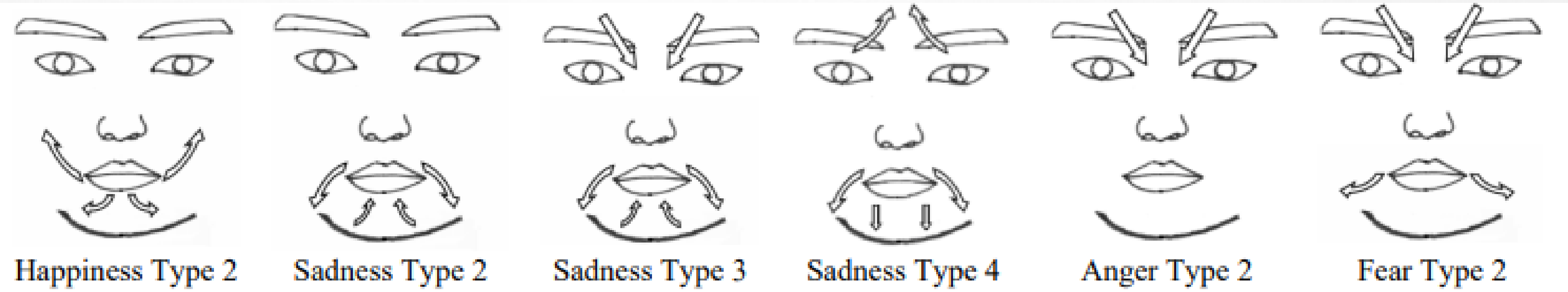



Figure 6. Other types of face deformation in different emotions

**Figure 5** shows Bassile facial expression deformation.
**Figure 6** displays other types of deformations extracted experimentally from facial expression image sequences in CK+ dataset.

- It is clear from these figures that the most important changes happen on the top of eyes, around the eyebrows and mouth.

- So these areas are divided into smaller sections as shown in Figure 7(a).
- Dividing these areas into smaller sections gives the chance to analyze the deformations in these sections more precisely.
- In Figure 7(b), nine vectors showing different directions and segmentations can be seen.
- As face is symmetric, the number and size of segments in X1 and X4, X2 and X5 and, X3 and X6 directions are the same. These segmentations can be different in width, length and number.
- In each segment, the ratio of vector numbers and their average length in x and y directions are extracted and used for mining.
- So, for each image sequence, 3n features are extracted, where n is the number of small segments. For example, in Figure 7(a), the number of small segments is 120. So 360 features are extracted and used for mining. By using this 360 feature, the system should analyze the face deformation and detect facial expressions

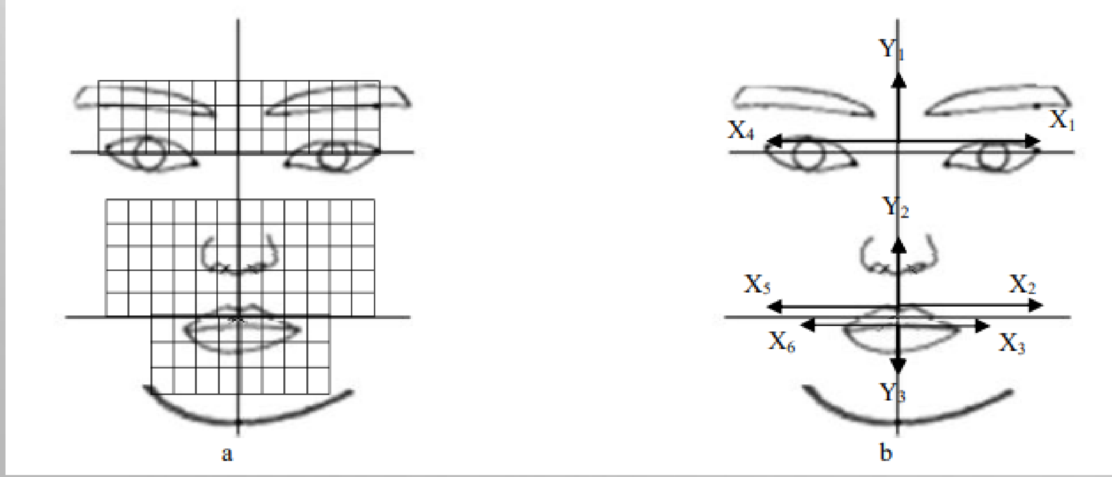



Figure 7. a) Face areas having the most important deformation during facial expression are divided into smaller sections. b) Nine different directions in which segmentations are done.

# FACIAL EXPRESSION CLASSIFICATION
## (DATA ANALYSIS STAGE)

**Data Cleaning - Handling Missing Values and Outliers**

- Outliers are replaced with the nearest value that would not be considered as an outlier.
- For e.g., if an outlier was defined to be anything above or below three standard deviations, then all of them would be replaced with the highest or lowest value within this range.
- Extreme values would be discarded.
- Missing values (sections wherein there was no motion vector) would be replaced by zero.
- Then, 10-fold cross validation was applied 50 times on these data.
- In this research, C5.0, CRT, QUEST, CHAID,  Neural Networks, Deep Learning, SVM and Discriminant algorithms were used to classify the extracted features. These classification algorithms were used to extract knowledge from a dataset consisting of motion vector features. These motion vectors formed feature vectors. These features calculated according to equations 1, 2 and 3.

These features calculated according to equations 1, 2 and 3.

$$P_{ij} = \frac{N_{ij}}{N}$$

$$LX_{ij} = \frac{1}{N_{ij}} \sum_{k=1}^{N_{ij}} lx_{ij}^{k}$$

$$LY_{ij} = \frac{1}{N_{ij}} \sum_{k=1}^{N_{ij}} ly_{ij}^{k}$$

where

- $N$ is the number of extracted motion vectors.
- $N_{ij}$ is the number of extracted motion vectors in section $ij$.
- $P_{ij}$ is the ratio of motion vector numbers in section $ij$.
- $lx_{ij}^{k}$ is the length of motion vector $k$ in direction $x$ in section $ij$.
- $ly_{ij}^{k}$ is the length of motion vector $k$ in direction $y$ in section $ij$.
- $LX_{ij}$ is the mean of motion vectors' length in direction $x$ in section $ij$.
- $LY_{ij}$ is the mean of motion vectors' length in direction $y$ in section $ij$.

As illustrated in Figure 8, each feature vector was composed of the values of features calculated by equations 1, 2 and 3. The ratio of vector number and the mean length of motion vectors in each created section were used as characteristic features.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| $P_{11}$ | $LX_{11}$ | $LY_{11}$ | $P_{12}$ | $LX_{12}$ | $LY_{12}$ | . . . | $P_{NM}$ | $LX_{NM}$ | $LY_{NM}$ |

Figure 8. An example of feature vector used for facial expression classification.

The location and size of deformation are often different from one image sequence to another. They even differ from one occurrence to another one in a person. However, these differences are not a lot in a specific facial expression.

# PSEUDO-CODE OF PROPOSED ALGORITHM

1. Prepare image sequence % Onset: natural state;
   % Apex: one of six basic emotions.
2. Apply preprocessing on image sequence
   2.1. Convert images into grayscale
   2.2. Cut images so that only the face is shown in them.
3. Extract motion vectors in image sequence using optical flow algorithm.
4. Divide images into six areas.
5. Segment each area as shown in Figure 7(a)
6. Extract data
   6.1. Calculate the percentage of the number of vectors in each segment
   6.2. Calculate the mean of vector length in x and y directions in each segment
7. Eliminate outlier and missing data from extracted motion vectors.
8. For 50 times,
   8.1. Apply 10-fold cross-validation
   8.2. Estimate the performance of classifier
9. Calculate the overall performance of classifier

# EXPERIMENTAL RESULTS

- Feature vectors were extracted from motion vectors in the image sequences.
- For each image sequence, between 18 and 631 features were used.
- For each facial expression, 50 times 10-fold cross validation were applied.
- In order to evaluate and compare the performance of different algorithms, different algorithms were tested on 25 situations (different width, height and number of segments in different directions)

Table 1. Features of different situations. Segments width and height are in pixel

| | | Segments | | Number of Segments in different directions | | | | | | Total number of features |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Width | Height | $X_1$ & $X_4$ | $Y_1$ | $X_2$ & $X_5$ | $Y_2$ | $X_3$ & $X_6$ | $Y_3$ | |
| Situation | 1 | 5 | 5 | 3 | 3 | 4 | 3 | 2 | 3 | 162 |
| | 2 | 5 | 5 | 8 | 5 | 8 | 5 | 5 | 5 | 630 |
| | 3 | 5 | 15 | 2 | 1 | 3 | 1 | 2 | 1 | 42 |
| | 4 | 5 | 15 | 8 | 2 | 8 | 2 | 5 | 2 | 252 |
| | 5 | 5 | 20 | 5 | 2 | 5 | 2 | 2 | 1 | 132 |
| | 6 | 10 | 10 | 5 | 4 | 5 | 4 | 5 | 3 | 330 |
| | 7 | 10 | 10 | 5 | 5 | 5 | 5 | 5 | 5 | 450 |
| | 8 | 10 | 10 | 5 | 6 | 5 | 6 | 5 | 4 | 480 |
| | 9 | 10 | 15 | 5 | 3 | 5 | 3 | 3 | 2 | 216 |
| | 10 | 15 | 15 | 4 | 3 | 4 | 3 | 3 | 2 | 168 |
| | 11 | 15 | 15 | 5 | 5 | 5 | 5 | 5 | 5 | 450 |
| | 12 | 15 | 20 | 3 | 2 | 3 | 2 | 2 | 1 | 84 |
| | 13 | 20 | 20 | 5 | 5 | 5 | 5 | 5 | 5 | 450 |
| | 14 | 20 | 25 | 2 | 1 | 2 | 1 | 1 | 1 | 30 |
| | 15 | 25 | 25 | 2 | 2 | 2 | 2 | 2 | 2 | 72 |
| | 16 | 25 | 25 | 4 | 4 | 4 | 4 | 4 | 4 | 288 |
| | 17 | 25 | 30 | 1 | 1 | 1 | 1 | 1 | 1 | 18 |
| | 18 | 30 | 30 | 2 | 2 | 2 | 2 | 1 | 1 | 54 |
| | 19 | 30 | 30 | 2 | 2 | 2 | 2 | 2 | 2 | 72 |
| | 20 | 30 | 30 | 3 | 3 | 3 | 3 | 2 | 2 | 132 |
| | 21 | 30 | 30 | 4 | 4 | 4 | 4 | 4 | 4 | 288 |
| | 22 | 35 | 35 | 3 | 3 | 3 | 3 | 3 | 3 | 162 |
| | 23 | 40 | 40 | 3 | 3 | 3 | 3 | 2 | 2 | 132 |
| | 24 | 40 | 40 | 3 | 3 | 3 | 3 | 3 | 3 | 162 |
| | 25 | 50 | 50 | 2 | 2 | 2 | 2 | 1 | 1 | 54 |

- To determine how many segments were more suitable and what size is better for them, these different situations were compared.

- It was observed that DL in situation 21, SVM and C5.0 in situation 19 had the best performance with the accuracy rate of 95.3%, 92.8% and 90.2%, respectively.

- The confusion matrix for these algorithms after ignoring the confusion between different types of facial expressions was observed . It means that for example, both types of happiness type 1 and 2 are happiness and it is not important which type of happiness is recognized. Their overall averages were also in the same rank.

# MISIDENTIFICATIONS

- As far as misidentifications produced by these methods are concerned, most of them arose from confusion between similar motion vector location and directions.
- Only the motion vectors direction in area number two and three could distinguish angry type 1 from disgust, causing misidentification of them in the case where the motion vector directions were not recognized precisely in these areas.
- Other high rate misclassifications happened between fear and happiness. Since both types of happiness and both types of fear had the same motion vectors in the lower part of face, in about five percent, fear was classified as happiness.
- The most important misclassifications are summarized in the Table shown, sorted according to the misclassification rate.

Table 9. The most important misclassifications

| Algorithm name | Emotion | Misclassify as | Rate of misclassification |
|---|---|---|---|
| C5.0 | Sadness | Happiness | 7.8 |
| C5.0 | Fear | Surprise | 7.6 |
| SVM | Fear | Happiness | 7.1 |
| SVM | Anger | Sadness | 7.0 |
| C5.0 | Anger | Sadness | 6.9 |
| DL | Anger | Disgust | 6.3 |
| DL | Fear | Happiness | 4.8 |

# ANALYSIS OF BEST ALGORITHM AND BEST SEGMENTATION

- What must be considered is which algorithm can classify emotion vectors
  - With higher accuracy
  - With lower features.

- Three best situations of each algorithm are shown in Table. In three algorithms CRT, SVM and C5.0, situation 19 resulted in the best.
- It had the second rank in DL. Meanwhile, situation 19 had the best overall accuracy among other situations.
- In situation 19, the face is divided into four sections in each area. There were two subsections in each direction with equal width and height of 30 pixels. So, it was the best segmentation for facial expression recognition.

Table 10. Three best situations for each algorithm

| DL | CRT | QUEST | CHAID | SVM | C5.0 | Discriminant | Overall AVG |
|---|---|---|---|---|---|---|---|
| 21 | 19 | 16 | 11 | 19 | 19 | 12 | 19 |
| 19 | 20 | 10 | 24 | 17 | 3 | 25 | 18 |
| 23 | 16 | 22 | 6 | 9 | 25 | 15 | 25 |

# CONCLUSION

- In this research, some of the most famous classification algorithms were used upon changes in the position of facial points. These points were tracked in an image sequence of a frontal view of the face. The best methods were chosen. They were DL, SVM and C5.0, with the accuracy rate of 95.3%, 92.8% and 90.2%, respectively.

- The most distinguishing changes were found to be deformation in Y direction, in the upper and lower areas of the face. Meanwhile, some more changes of face during facial expression were investigated in this research. It shows that six more changes can be identified in addition to six basic changes happens in face during representation of facial expression.

- This paper not only provided a basic understanding of how facial points could change during a facial expression, but also it tried to classify these deformations.

- Future work on this issue aims at investigating automatic facial expression while head has rigid motions. Meanwhile, the performance of the method must be invariant to occlusions like glasses and facial hair. In addition, the method must perform well independently of the changes in the illumination intensity while image sequence is created.

# MY VIEWS AND LEARNINGS

- Automating the analysis of facial changes, especially from a frontal view, is important to advance the studies on automatic facial expression recognition, design human-machine interfaces.
- Facial Expression Analysis would  boost  applications in various areas such as security, medicine, animations and education.
- Different changes in  parts of face were analyzed to address what exactly happened in face when a person shows an emotion. Changes happening in the face because of facial expression were introduced in this research for the first time.
- Introduced to some of the most state-of-the-art classification algorithms such as C5.0, CRT,  QUEST, CHAID, Deep Learning (DL), SVM and Discriminant algorithms which were used to classify the extracted motion vectors.