FACIAL EXPRESSION RECOGNITION BASED ON FACIAL MUSCLES BEHAVIOR ESTIMATION

Saki Morita and Kuniaki Uehara
Department of Computer and Systems Engineering, Kobe University
1-1 Rokko-dai, Nada, Kobe 657-8501, Japan

Keywords: Facial expression analysis, finite element method, angular metrics for shape similarity, emotion classification.

Abstract: Recent development in multimedia urges the need for an engineering study of the human face in communication media and man-machine interface. In this paper, we introduce a method not only for recognizing facial expression and human emotion, but for extracting rules from them as well. Facial data can be obtained by considering the relative position of each feature point in time series. Our approach estimates the behavior of muscles of facial expression from these data, and evaluates it to recognize facial expressions. In the recognition process, essential parameters that cause visible change of the face are extracted by estimating the force vectors of points on the face. The force vectors are calculated from displacements of points on the face by using FEM (Finite Element Method). To compare the multi-streams of force vectors of each facial expression effectively, A new similarity metric AMSS (Angular Metrics for Shape Similarity) is proposed. Finally, experiments of recognition of facial expressions shows that usable results are achieved even with few testees in our approach and variable rule corresponding AUs can be detected.

1 INTRODUCTION

Recent development in multimedia urges a need for an engineering study of the human face in communication media and man-machine interface. The human face is full of nonverbal information that is used for communication and to indentify ourselves from one another. Therefore, the number of researchers who study on the recognition of facial expression is increasing.

Currently, a lot of facial expression recognition systems have been devised and different approaches have been introduced. There are two types of approaches: 2D (image-based) method and 3D (model-based) method. In some approaches of 2D method, motion information such as optical flow is extracted from sequential images. Then, the expression is described by some specific parameters representing optical flow and six facial expressions as well as eye blinking are recognized (Yacoob and Davis, 1994). In other approaches of 2D method, the presence or absence of furrows and wrinkles is observed and the magnitude as well as the direction of the motion of the face's parts are represented. Then, the expression is classified using perceptron with these parameters (li Tian et al., 2001).

However these approaches need lots of information since they use planar information of the whole face or parts of it. In contrast, 3D method recognizes facial expressions by analyzing three-dimensional information of several points on a face. Our approach is based on 3D method and estimates the movements of the muscles of a facial expression by considering the relationship among points on a face. Then, the facial expression is classified based on the movement of the muscles of the facial expression.

This method is inspired by the idea of FACS (Facial Action Coding System) (Ekman and Friesen, 1978). FACS is a method for measuring and describing facial behaviors in the field of psychology. It divides facial movement into 44 kinds of basic units, AU (Action Unit).

In order to obtain positions of points on the face, we use a motion capture system. Data obtained by the motion capture system have the following characteristics:

- three dimensional time series data:
  Three dimensional coordinates of the markers placed on the face are obtained as time series data. Such data are called a stream.

Morita S. and Uehara K. (2006).
FACIAL EXPRESSION RECOGNITION BASED ON FACIAL MUSCLES BEHAVIOR ESTIMATION.
In *Proceedings of the First International Conference on Computer Vision Theory and Applications*, pages 48-55
DOI: 10.5220/0001372400480055
Copyright © SciTePress

- multi-stream data:
  Since a large number of markers are placed on the face, it is necessary to process the multiple streams by considering the interrelation between them.

These data have information about the position of each part of the face with good accuracy.

Since these data obtained by motion capture system have just the positions of the markers, they need to be transformed to parameters representing the features of each facial expression notably. We present a method for extracting essential parameters that cause visible change of the face by estimating the force vector of each point on the face. The force vectors need to be calculated from displacements of points on the face. That is, the forces acting on the skin of the face are obtained by the movments of each point on the face. Therefore, our approach uses a method of analyzing inverse problem using FEM (Cook, 1995) to estimate the force vectors. FEM is a very powerful tool to obtain stress and strain when outside forces act on an object. Since FEM can be easily applied to various engineering problems and handles complex loading, the method for analyzing inverse problem using FEM is the best tool as a method to estimate the force vectors from the displacemants.

Then, the force vectors must be compared correctly to recognize facial expressions. The comparison of force vectors by Euclidean distance does not have essential significance, since the force vectors have two elements: direction and length. Thus, we propose a new similarity metric AMSS (Angular Metrics for Shape Similarity) for an effective evaluation of force parameters. A similarity of force vectors by AMSS is calculated from the difference of the length and the angle between the two vectors. The comparison method using AMSS can achieve an exact evaluation using elements of vectors effectively. Expression recognition is done by estimating force vector in DTW (Dynamic Time Warping) (Sankoff and Kruskal, 1983) using AMSS.

This paper is organized as follows: Section 2 describes the motion capture system and the data for facial expression. Section 3 and Section 4 describe the technique for extracting parameters representing forces acting on points on a face using FEM, and the method of facial expression recognition using the force vectors. Section 5 performs an experiment on expression recognition and discusses the results. We conclude in Section 6 by summarising the paper and suggesting future research directions.

2 FACS

In this paper, expressions are recognized based on the idea of FACS which is widely used in the field of facial expression analysis. Since a facial expression is indeed the combination of movements of the points on the face, the data of the whole surface of the face is not needed for the recognition of facial expressions. All you need to recognize the facial expressions relies on the information of some points on a face. We also use the optical motion capture system to obtain the data of 35 points on the face.

FACS is a method for measuring and describing facial behaviors proposed by Ekman and Friesen. FACS is widely used in the field of the research about the recognition of facial expression (Essa and Pentland, 1997) (Lien et al., 1998). Ekman and Friesen developed the original FACS by determining how the contraction of each facial muscle (singly and in combination with other muscles) changes the appearance of the face.

In FACS, an expression is described by a basic unit called AU. AU is a primitive unit of the expression movement that can be identified visually, and there are 44 kinds of AUs in total. Human expressions are described by these combinations. For example, $Sadness$ is described as “1+4+15+23” since it is composed of four AUs, such as 1, 4, 15 and 23.

The 17 AUs used to express basic facial expressions are shown in Table 1. For instance, AU 1 describes the movement of raising the inner corner of the eyebrow and AU 4 describes the movement of puckering up one’s brows. Each AU is influenced by a specific expression muscle respectively.

Table 1: Examples of AU(Action Unit).

No.	Name	No.	Name
1	$InnerBrowRaise$	14	$Dimpler$
2	$OuterBrowRaise$	15	$LipCornerDepress$
4	$BrowLower$	16	$LowerLipDepress$
5	$UpperLidRaise$	17	$ChinRaise$
6	$CheekRaise$	20	$LipStretch$
7	$LidTight$	23	$LipTight$
9	$NoseWrinkle$	25	$LipsPart$
10	$UpperLipRaise$	26	$JawDrop$
12	$LipConerPull$		

Six basic expressions proposed by Ekman are widely used in classification of a human expression. The six basic expressions are as follows: happiness, sadness, surprise, disgust, fear and anger. Table 2 shows the combination and the strength of AUs to represent the six basic expressions. The numerical value in parentheses describes strength. It is 0 when the AU is invisible, and it is 100 when the AU is apparent. For instance, $Anger$ is composed of eight AUs, such as 2, 4, 7, 9, 10, 12, 15 and 26, because features of $Anger$ are puckering up one’s brows, staring, applying, and clenching teeth.

In this paper, expression data are obtained by using an optical motion capture system. These data consist of x, y, z coordinates for each point on the testee’s

Table 2: Combination of AU parameters.

Expression	AU and its Percentage
Happiness	1(60), 6(60), 10(100), 12(50), 14(60), 20(40)
Sadness	1(100), 4(100), 15(50), 23(100)
Surprise	1(100), 2(40), 5(100), 10(70), 12(40), 16(100), 26(100)
Disgust	2(100), 4(100), 9(100), 15(50), 17(100)
Fear	1(100), 2(40), 4(100), 5(70), 12(30), 15(70), 26(60)
Anger	2(70), 4(100), 7(60), 9(100), 10(100), 12(40), 15(50), 26(60)

face. The position of the marker is determined based on FACS. The markers are sticked in the appropriate places where the movement of each AU is observed. The positions of markers and those names are shown in Figure 1. The total number of the markers used in this research is 35. In Figure 1, we link markers on the face with black lines to describe the places where AUs happen. For example, AU 1 and AU 2 can be detected by observing the markers marked as number 6 and 7 respectively, but some AUs like AU4 should be detected by observing the line between the markers numbered 6 and 7.

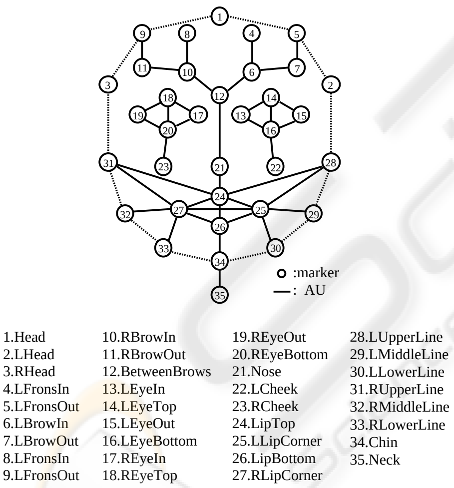

Figure 1: The markers and Action Units.

3 FEATURE EXTRACTION BY FEM

The behavior of muscles of facial expression is estimated and evaluated to classify the expression data in our approach. First, displacement for each coordinate between corresponding points on the face in two continuous frames are calculated according to their position values. Thereby, these data are three-dimensional time series and multi-stream. Next, parameters representing forces acting on points on the face are calculated from values of displacements by the method of inverse problem using FEM.

FEM is a very powerful technique to obtain the numerical solution of a wide range of engineering problems. FEM is based on the concept that a body or structure may be devided into smaller elements of finite dimensions. The original body or structure is discretized to these elements which are related to each other through nodes. Application of the governing equations, loading and boundary conditions results in a system of equations that could be solved to find an approximate solution. The main features of FEM are as follows:

- It can readily handle very complex geometry.
- It can handle a wide variety of engineering problems.
- It can handle complex restrains.
- It can handle complex loading.
- It obtains approximate solutions.

Since visible change of the face is caused by complex effecting of the muscles of facial expression, FEM is appropriate to handle facial expression recognition problems.

FEM can achieve the displacements of nodes when outside forces act on an object. But using facial expression data as an input, we must take an approach that calculates the force vector from the displacement vector. This method is called a method for the inverse problem using FEM.

Figure 2 shows the example of the plane board model with nine nodes and eight elements. The number of materials and the number of restraint conditions are given as input data. Here, assume that the number of materials is one and the number of restraint conditions is three. The calculation procedure of reverse problem analysis using FEM is as follows:

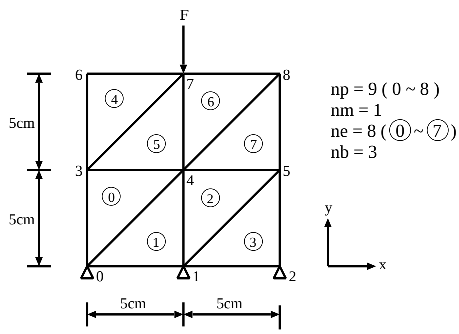

Figure 2: An example of reverse problem.

1. Preparation for input: The structure that needs to be analyzed is divided into elements, and nodes and elements are numbered.
2. Input data: The coordinates of nodes, composition of the elements, physical properties of materials and fixed conditions are inputted. Additionally, information on nodes and elements decided by step 1 are given.
3. Making an element rigidity matrix: Element rigidity matrix $[EK]$ of each element is made by using input data. At this time, it is necessary to calculate plane strain matrix $[B]$ and stress strain matrix $[D]$ (for the plane strain problem). $[B]$ and $[D]$ are represented as the following expressions:

$$[B] = \frac{1}{2S}\left[\begin{array}{cccccc} y_j - y_k & 0 & y_k - y_i & 0 & y_i - y_j & 0 \\ 0 & x_k - x_j & 0 & x_i - x_k & 0 & x_j - x_i \\ x_k - x_j & y_j - y_k & x_i - x_k & y_k - y_i & x_j - x_i & y_i - y_j \end{array}\right] \quad (1)$$

$$[D] = \frac{E}{(1+\nu)(1-2\nu)}\left[\begin{array}{ccc} 1-\nu & \nu & 0 \\ \nu & 1-\nu & 0 \\ 0 & 0 & \frac{1-2\nu}{2} \end{array}\right] \quad (2)$$

$$S = \frac{1}{2}\left|\begin{array}{ccc} 1 & x_i & y_i \\ 1 & x_j & y_j \\ 1 & x_k & y_k \end{array}\right| \quad (3)$$

$$[EK] = tS[B]^T[D][B] \quad (4)$$

Note that x_n and y_n are the x-coordinate and y-coordinate of the node n of each element respectively. E is the Young's modulus and ν is the Poisson's ratio. t is a board thickness.
4. Assembling the whole rigidity matrix: Whole rigidity matrix $[TK]$ is made by combining each element of the rigidity matrix $[EK]$.
5. Processing the force condition and the condition of constraint: The nodal force vectors $\{F\}$ is obtained by solving the following simultaneous linear equation. $\{d\}$ is displacements of the nodes.

$$\{F\} = [TK]\{d\} \quad (5)$$

Then, we describe the method of applying the reverse problem analysis of FEM to expression data. When expression data are obtained, the positions of markers are decided by considering the expression muscle, as is shown in Figure 1. To apply FEM, it is necessary to divide the testee's face into some elements. Figure 3 shows the division of the face into elements. Then, the markers' positions are considered to be the nodes in this structure. 56 elements are divided by 35 nodes. Each element is formed by three nodes. For example, nodes marked as number 1, 4 and 5 in Figure 1 form the element marked as number 1 in Figure 3. The elements can be defined arbitrarily and they are determined to be left-right symmetrical. The reason is that human face is almost left-right symmetrical and so does the behavior of muscles of facial expressions in most cases.

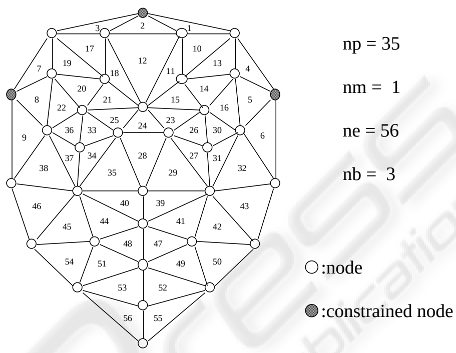

Figure 3: Dividing face into elements.

The number of materials is one because the skin overlying the face is assumed to be uniform. The number of restraint conditions is three since three points are fixed. These three points are Head, LHead, RHead, shown in Figure 1. The Young's modulus E and Poisson's ratio ν are $0.14[MPa]$ and 0.45 respectively.

After these input data are given, the force parameters of the nodes in each frame are calculated by the above method for analyzing inverse problem using FEM. Since the displacement of z-axis in Figure 6 is small, it has little influence on the result. Therefore, the node displacement as well as the node force in this direction are ignored.

Each force parameter of a node in each frame obtained by the above process is described by two real values. One means the strength of the force which acts on in the vertical direction, the other means the strength of the force in the horizontal direction. The length and the angle of the node force vector are calculated from these two real values of the force parameter. The expressions are as follows:

$$Length = \sqrt{x^2 + y^2} \quad (6)$$

$$Angle = \arctan(y/x) \quad (7)$$

Note that x is the strength of the force in the horizontal direction and y is strength of the force in the vertical direction. If x is 0, the angle is $2/\pi$ when y is plus and it is $-2/\pi$ when y is minus.

4 SIMILARITY BETWEEN MULTI-STREAMS

Facial expressions are recognized by evaluating force vectors obtained in Section 3. The force vectors must be compared exactly. Although the Euclidean distance is commonly used as the metric of distance, the comparison of force vectors by Euclidean distance does not have essential significance because the force vector has two elements: length and direction. Thus, we propose AMSS, a new similarity metric for an effective evaluation of force vectors.

AMSS measure is inspired by the similarity measure based on LCSS (Longest Common SubSequence) model (Vlachos et al., 2002) (Gunopoulos, 2002). LCSS is proposed for measuring the similarity of model object's trajectory. LCSS based measure is a method of analyzing object trajectories in two or three dimensional space. It ignores dissimilar segments of trajectories and calculates the similarity from only the similar segments. The advantage of this measure is that it minimizes the effects of the location where data are captured. However, LCSS based metric is not sufficient to measure force vectors. We present how AMSS measure is calculated: assume that $\overrightarrow{v_1}$ and $\overrightarrow{v_2}$ are force vectors.

$$SimA(\overrightarrow{v_1},\overrightarrow{v_2}) = 1 - Dist(\overrightarrow{v_1},\overrightarrow{v_2}) \quad (8)$$

$$Dist(\overrightarrow{v_1},\overrightarrow{v_2}) = \frac{1}{2}(Dist_a(\overrightarrow{v_1},\overrightarrow{v_2}) + Dist_l(\overrightarrow{v_1},\overrightarrow{v_2})) \quad (9)$$

$$Dist_a(\overrightarrow{v_1},\overrightarrow{v_2}) = \frac{\theta}{\pi/2} \quad (10)$$

$$Dist_l(\overrightarrow{v_1},\overrightarrow{v_2}) = \frac{||\overrightarrow{v_1}| - |\overrightarrow{v_2}||}{Max(|\overrightarrow{v_1}|,|\overrightarrow{v_2}|)} \quad (11)$$

Note that θ is the angle between $\overrightarrow{v_1}$ and $\overrightarrow{v_2}$ (Figure 4(a)). $SimA(\overrightarrow{v_1},\overrightarrow{v_2})$ and $Dist(\overrightarrow{v_1},\overrightarrow{v_2})$ are the similarity and the distance between $\overrightarrow{v_1}$ and $\overrightarrow{v_2}$ respectively. $Dist_a(\overrightarrow{v_1},\overrightarrow{v_2})$ is the angle between the two vectors (Figure 4(b)), and $Dist_l(\overrightarrow{v_1},\overrightarrow{v_2})$ means the difference of the length of the vctors (Figure 4(c)). As a result, similarity can be appropriately calculated from both viewpoints: the direction and the length of the vector.

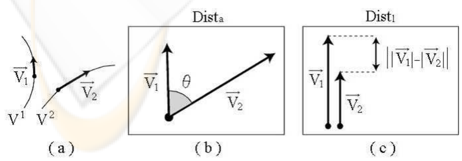

Figure 4: How to measure the similarity between two vectors.

For example, the similarities between two vectors in Figure 5 are calculated. In Figure 5(a), $|\overrightarrow{a_1}|$ and $|\overrightarrow{a_2}|$ are 1 and $\sqrt{2}$ respectively, and the angle between $\overrightarrow{a_1}$ and $\overrightarrow{a_2}$ is $\pi/4$. Since $Dist_a(\overrightarrow{a_1},\overrightarrow{a_2})$ is 0.50 and $Dist_l(\overrightarrow{a_1},\overrightarrow{a_2})$ is 0.29, $Dist(\overrightarrow{a_1},\overrightarrow{a_2})$ is 0.40 and $SimA(\overrightarrow{a_1},\overrightarrow{a_2})$ becomes 0.60. On the other hand, the similarity between $\overrightarrow{b_1}$ and $\overrightarrow{b_2}$ in Figure 5(b) is calculated and $SimA(\overrightarrow{b_1},\overrightarrow{b_2})$ becomes 0.42. This result shows that $\overrightarrow{a_1}$ and $\overrightarrow{a_2}$ are more similar than $\overrightarrow{b_1}$ and $\overrightarrow{b_2}$. These similarities are based on the angle and the length of the vectors, and we can see the effect of this metrics according to this result.

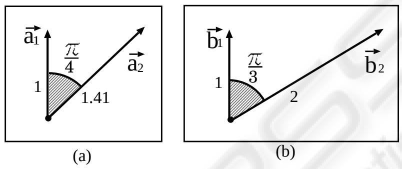

Figure 5: How to measure the similarity between two vectors in AMSS.

DTW is generally used as a method of measuring the similarity between waveforms. DTW is an algorithm that measures the distance between two time series. Therefore, the similarity between the data of a same expression with different timings can be correctly measured. In this research, the distance metric, AMSS, is used as an alternative to the Euclidean distance and the similarity between streams is measured by DTW.

For instance, suppose that the similarity between stream A and B needs to be evaluated. The similarity of stream A (frame 0 to frame i), and stream B (frame 1 to frame j) can be given in the next equation:

$$\begin{aligned} SimB(A,B) &= D(i,j) \\ &= Max[D(i-1,j-1) + 2SimA(i,j), \\ &D(i-2,j-1) + SimA(i-1,j) + SimA(i,j), \\ &D(i-1,j-2) + SimA(i,j-1) + SimA(i,j)] \end{aligned} \quad (12)$$

A single facial expression data consists of 35 force vector streams as described in Section 2, so that we consider it is a multi-stream data. Since three nodes marked as number 1, 2 and 3 are considered as fixed points, the similarities of their force vector streams are not calculated. That is, the similarity between two facial expressions is calculated from 32 streams. Finally, the similarity of the multi-streams is calculated as an accumulation of the similarity of each feature points. The similarity between two expression data $E1$ and $E2$ is calculated as follows:

$$Similarity(E1,E2) = \sum_{n=4}^{35} SimB(E1_n, E2_n) \quad (13)$$

Note that n is the node number and $E1_n$ and $E2_n$ represent the n-th force vector streams of $E1$ and $E2$ respectively. The similarity between two multi-streams is calculated as the accumulation of the similarities of all pairs of single streams. That is, all forces acting on any points on the face are treated equally. Naturely, irrelevant streams may decrease the recognition's accuracy. Thus, we need more consideration about selection the streams that lead to be the best performance. This will be discussed in the following section.

Then, we describe how these multi-streams of expression data are classified into the categories of basic facial expression. First, the similarity between one multi-stream of unknown expression data and each multi-stream fallen into a certain category is calculated in equation 13. By calculating the mean value of these similarities between multi-streams, the similarity between unknown data and this category are obtained. Second, the similarities between unknown data and other categories are obtained as well. Finally, unknown data is classified into the category whose similarity is the highest. The result of recognition $result$ is represented with the following expressions:

$$SimC(c) = \frac{\sum_n Similarity(unkown, E_n^c)}{total_c} \quad (14)$$

$$result = \arg\max_c SimC(c) \quad (15)$$

Note that c reprensents the category and $SimC(c)$ is the similarity between unknown data and the category c. E_n^c is n-th multi-stream in the category c, $total_c$ is the number of data in the category c.

5 EVALUATION

5.1 Experiment with All Streams

We evaluate our system by considering the classification performance for four types of emotions. Expression data are obtained by using the optical motion capture system HiRES (4 cameras) made by Motion Analysis company. The overview of the motion capture system is indicated in Figure 6. The sampling frequency of the expression data is 60 $[Hz]$.

The facial motion capture markers are used to obtain the data. The total number of markers is 35 and they are sticked in the places determined in Figure 1 on the testee's face (Figure 7).

All tests are performed with 12 testees. Each testee is tested with four emotions; $Anger$, $Sadness$, $Happiness$, $Surprise$. The length of each sequence is five seconds, in another words, 300 frames, starting and ending with the neutral expression. The experiments have been carried out using four types of expression data set of one person as test data, and the remainder as training data. This trial was repeated 12 times, each time leaving a different person out (leave one out cross varidation). The results are shown in Table 3 and Table 4.

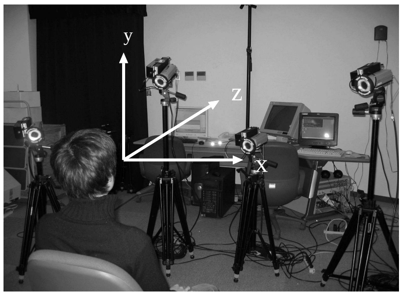

Figure 6: The optical motion capture system.

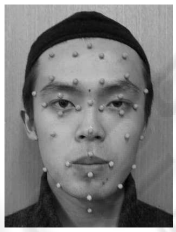

Figure 7: Facial motion capture markers.

Table 3: Recognition accuracy for person-indepent classification.

Emotion	Percent correct
$Surprise$	100.0%
$Anger$	33.3%
$Happiness$	50.0%
$Sadness$	83.3%
Total	66.7%

Table 3 gives the percentage of correctly classified testing data for each basic emotion and the overall recognition accuracy. Table 4 gives the confusion matrix for this trial. It can be seen from this result that while $Surprise$ and $Sadness$ are classified correctly, $Anger$ and $Happiness$ are misclassified frequently. Especially for $Anger$, it is more frequently misclassified than others. The reason of this result is that the forces acting on the face in $Anger$ are smaller than those of other expressions. In addition, $Anger$ has features in common with other facial expressions. For example, $Anger$ and $Surprise$ share some AUs, such as AU 2, AU 10 and AU 12. Additionally, $Anger$ and $Sadness$ share some AUs, such as AU 4 and AU 15. Therefore, the data of $Anger$ are misclassified into $Surprise$ and $Sadness$.

Table 4: Person-independent confusion matrix.

Input	Output			
	$Sur.$	$Ang.$	$Hap.$	$Sad.$
$Surprise$	12	0	0	0
$Anger$	3	4	0	5
$Happiness$	1	0	6	5
$Sadness$	0	1	1	10

5.2 Stream Selection

Feature Selection is a process of choosing the most appropriate subset of the features and is commonly used in machine learning. By selecting the smallest relevant subset of the features which maintain the characteristics of the original data, the computation time to analyze data can be reduced. Moreover, it is possible that a prediction performance is improved by using only effective information subset in the prediction.

In our research, the positions of 35 markers are observed and 35 streams for each facial expression data are extracted. Since all streams representing the force vectors are treated equally to the facial recognition in Subsection 5.1, irrelevant streams could decrease the recognition's accuracy. It is necessary to select some relevant streams from the whole set for the sake of the classification. Therefore, the subset of relevant streams are attempted to be selected by certain processes. This process is based on the idea of feature selection and we call it $stream\ selection$. In this subsection, two approaches are used for stream selection: a wrapper approach (Kohavi and John, 1997) and a skimming approach.

The first method of stream selection is known as wrapper approach. This approach generates various subsets of features and evaluates them by measuring the accuracy of the resulting classifier. Then, the subset of features given the highest mark is used to recognize expressions. The feature subsets are generated starting with a single feature and gradually adding a feature at a time. Since the face is left-right symmetrical, it is assumed that the forces which act on each point of the face are almost symmetric. Therefore, 19 streams on the right half of the face are targets of stream selection. The following two tables show the recognition results using two streams chosen by stream selection.

Table 5 and Table 6 give the recognition accuracy and the confusion matrix respectively. It can be seen that this method can recognize facial expressions with slightly decreased accuracy when only two streams are used. In the recognition using all streams, it is difficult to classify $Anger$ correctly since $Anger$ is composed of many AUs. Therefore, the recognition rate of $Anger$ is improved by using the relevant streams selected by the wrapper approach.

Table 5: Recognition accuracy using stream selection.

Emotion	Percent correct
$Surprise$	66.7%
$Anger$	50.0%
$Happiness$	66.7%
$Sadness$	50.0%
Total	58.3%

Table 6: Person-independent confusion matrix.

Input	Output			
	$Sur.$	$Ang.$	$Hap.$	$Sad.$
$Surprise$	8	0	3	1
$Anger$	2	6	3	1
$Happiness$	1	1	8	2
$Sadness$	3	0	3	6

Table 7 shows the subsets frequently selected by the wrapper approach. The experiment is performed changing the number of the selected streams from one to three. The markers' numbers of selected streams for each case are shown in Table 7. In each case, most-selected features are the subset of streams of markers marked as number 10 and 11 in Figure 1. It indicates that some discriminative information is embedded in the upper part of the face.

Table 7: The subset of streams chosen by stream selection.

frequency	The number of the streams		
	1	2	3
1st	{11}	{10,11}	{10,11,17}
2nd	{10}	{11,17}	{10,11,19}
3rd	{34}	{10,26}	{9,11,19}

The second experiment uses the skimming approach. This approach selects the subset of the features whose values change remarkably. In this trial, five streams of the markers which the powerful forces act on are chosen by the skimming approach. The recognition accuracy and the confusion matrix are shown in Table 8 and Table 9. The usable results can be achieved even reducing the dimensions from 35 to 5. While the total average recognition rate is superior to the experiment using the wrapper approach, the recognition accuracy of $Happiness$ decreases. The reason is that $Happiness$ has less movement in the chin and the eyebrows whereas they move obviously in other facial expressions.

Figure 8 shows the force vectors of relevant points on the face for the classification. Arrows represent the force vectors on the markers selected by the skimming approach. The direction and the length of the arrow represent the mean direction and length of the force vectors from the first frame to the peak frame. Some very short arrows are omitted. The result shows that force vectors obtained by our method are highly

Table 8: Recognition accuracy for person-indepent classification using stream selection.

Emotion	Percent correct
Surprise	91.7%
Anger	50.0%
Happiness	25.0%
Sadness	75.0%
Total	60.4%

Table 9: Person-independent confusion matrix.

Input	Output			
	Sur.	*Ang.*	*Hap.*	*Sad.*
Surprise	11	0	0	1
Anger	0	6	0	6
Happiness	3	1	3	5
Sadness	0	2	1	9

relevant with AUs. That is, by extracting essential parameters that cause visible change on the face, the behavior of muscles of a facial expression is estimated correctly. For instance, forces acting on the markers numbered as 6 and 10 correspond to AU 1, and forces acting on the markers numbered as 7 and 11 correspond to AU 2 in $Surprise$ (Figure 8(a)). Since these AUs are features of $Surprise$, it seems that our approach can detect variable rules and use them in the classification of facial expressions.

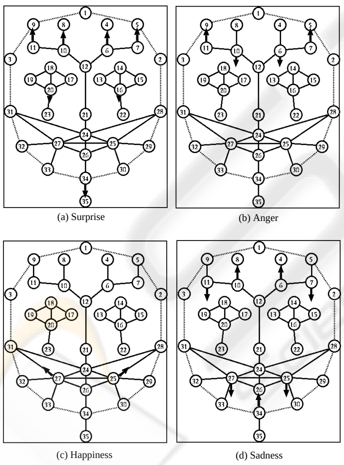

Figure 8: The force vectors of effective points for the classification.

6 CONCLUSION

In this paper, we presented an approach for recognizing facial expressions using three dimensional time series data from the point of view of data mining. This approach recognizes emotions by estimating and evaluating the behavior of muscles of facial expression. Three-dimensional positions of the points on the face are converted into force vectors using FEM. To compare the force vectors effectively, we proposed a new similarity measure AMSS. By applying AMSS to DTW, the similarity between streams of force vectors is appropriately calculated. In the experiment on expression recognition, usable results are achived with few testees.

Furthermore, the experiment of stream selection shows that this approach can recognize emotions using data of even two positions on the face. It can also find the points on the face that are effective in classification of facial expressions. The results indicates the possibility that stream selection method increases the classification's accuracy. More reliable system to recognize facial expression will be achieved by identifying the subset of streams that lead to be the best performance.

REFERENCES

Cook, R. D. (1995). *Finite Element Modeling for Stress Analysis*. Wiley.

Ekman, P. and Friesen, W. (1978). *The Facial Action Coding System*. Consulting Psychologists Press.

Essa, I. A. and Pentland, A. P. (1997). Coding, Analysis, Interpretation, and Recognition of Facial Expressions. *IEEE Trans. Pattern Anal. Mach. Intell.*, 19(7):757–763.

Gunopoulos, D. (2002). Discovering Similar Multidimensional Trajectories. In *Proc. of the 18th International Conference on Data Engineering*, pages 673–684.

Kohavi, R. and John, G. H. (1997). Wrappers for Feature Subset Selection. *Artificial Intelligence*, 97(1-2):273–324.

li Tian, Y., Kanade, T., and Cohn, J. F. (2001). Recognizing Action Units for Facial Expression Analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(2):97–115.

Lien, J. J., Kanade, T., Cohn, J. F., and Li, C. C. (1998). Automated Facial Expression Recognition Based on FACS Action Units. In *Proc. of the 3rd. International Conference on Face & Gesture Recognition*, pages 390–395.

Sankoff, D. and Kruskal, E. J. B. (1983). *Time Warps, String Edits, and Macromolecules: The Theory and Practice of Sequence Comparison*. Addison-Wesley.

Vlachos, M., Gunopulos, D., and Kollios, G. (2002). Robust Similarity Measures for Mobile Object Trajectories. In *Proc. of the 13th International Workshop on Database and Expert Systems Applications*, pages 721–728.

Yacoob, Y. and Davis, L. (1994). Computing Spatio-Temporal Representations of Human Faces. In *Proc. of Computer Vision and Pattern Recognition 94*, pages 70–75.