A NEW INVARIANT REPRESENTATION OF FACIAL EXPRESSIONS: DEFINITION AND APPLICATION TO BLENDED EXPRESSION RECOGNITION

Catherine Soladié
SUPELEC/IETR
Team SCEE, France

Nicolas Stoiber
Dynamixyz
Rennes, France

Renaud Séguièr
SUPELEC/IETR
Team SCEE, France

ABSTRACT

This paper proposes a novel method to perform accurate facial expression recognition by transforming the appearance space into an expression space. The expression space is computed from the person-independent organization of the facial expressions found out from data. The dimension of the expression space is reduced by the projection on a manifold compliant with the organization of the expressions. Experimental results on 14 different blended expressions show that the proposed organization based method improve the facial expression recognition performance compared to appearance based methods by 13%.

Index Terms—Facial expression recognition, Representation of facial expressions, Manifold of expressions

1. INTRODUCTION

A major goal of many new government programs is to find better ways that will allow the elderly to stay in their own homes longer, rather than moving to a care facility. One main domain for aging in place is to increase security and raise an alarm when some specific changes in behavior occur. Such systems have to be non intrusive, easy to use and agreeable. Among the ways of finding a change in behavior and feelings, facial expressions analysis play an important part, for they display human emotions and moods.

This paper explores a system based on a camera which can be included into a Set-top box. People in front of their TV set are most of the time mute and expressive. Such a system requires accuracy to analyze blended unknown expressions (as displayed in figure [1]).

Most facial expression analysis methods are user-independent and based on the learning of a huge amount of expressive faces [1]. They focus on the recognition of a small amount of facial expressions (usually the six "basic" expressions) and are not adapted to deal with blended unknown expressions. Some attempts have been made to deal with other expressions such as pain [2]. Few systems cope with blended expressions with varying intensities [3]. They then focus on one subject.

Until now, there exists just one data-based invariant representation to characterize uniquely an expression which is the FACS system [4]. The method is based on multiple separate action units, so that the natural correlation between multiple facial actions is ignored. To deal with morphological differences, some methods align different subjects’ expressions by subtracting the neural face assuming that all people have similar patterns of facial expression changing from neutral expression to a specific expression [5]. In pattern analysis, the choice of representation is known to influence the recognition performance. Psychologists are the first that have proposed a continuous space to organize the emotions [6]. Trying to align expression features to these high-level organizations lead to poor recognition results [7] because these organizations are not data driven. To answer these constraints, the organization of the emotions has been studied from appearance features which are deducted from the data [8] but no experiment has been performed on the expression recognition nor on the similarity of the space between subjects. Manifold learning methods have also been used for automatic facial expression analysis [9]. These methods need a high density of training data and usually fail at extrapolating to unknown expressions.

The main contribution of this article is the specification of a new invariant representation concerning the facial expressions. The originality of the approach is that we did not focus on the characteristics of one expression but on the organization of this expression with respect to the others. This organization has been found out from data and we have attested that this organization is person-independent by computing a similarity index. We used this invariant organization to transform the appearance space which is person-specific into the expression space which is person-independent. The recognition tasks are then processed in the expression space with a basic classifier. The experimental results compare an appearance based method (ABM) with our organization based method (OBM).

The remainder of this paper is organized as follows. In section [2] we justify the need of the transformation of the appearance space into the expression space. Section [3] describes the invariant representation used to perform the transformation and section [4] the individual steps of the transformation. Section [5] demonstrates the accuracy of the proposed system. Section [6] concludes the paper.

This research has been conducted with the support of Immemo (french ANR project) and IMMERSIVITE (Britanny Region PME project).
2. EXPRESSION RECOGNITION BY APPEARANCE BASED METHOD (ABM)

This section analyses the expression recognition results with an appearance based method (ABM). We use a person-independent active-appearance model (AAM) \cite{10}. The resulting appearance parameters are aligned between the subjects by subtracting the neutral face parameters for each subject. We use a voting algorithm on the nearest neighbor of each other subject for recognition rate computation.

As we will see in section 5, this method leads to poor recognition results. The comparison of distances between appearance parameters of different subjects’ similar expressions is not relevant. This can be justified by the fact that the appearance parameters carry information of both morphology and expression.

3. A PERSON-INDEPENDENT ORGANIZATION OF EXPRESSIONS

To overcome the limits of appearance based methods, we need an invariant representation of expressions that characterizes uniquely an expression, independently from the morphology. In this section, we will find out from data the organization of the facial expressions and will attest that this organization is person-independent by the computation of a similarity index.

To find out from data the organization of the facial expressions, we use $K$ known expressions, each one displayed by each of $P$ subjects. For each subject, we compute a person-specific AAM from the $K$ expressions and neutral face of the subject. Having a person-specific model increases accuracy, for it focuses on the facial distortions of the studied subject (that are the expressions) and not on the distortions due to morphological differences between subjects. To achieve comparability between the intensities of the expressions, we normalize the appearance parameters according to the parameters of the neutral face of the subject so that the expressions are located on an hypersphere centered on the neutral face.

We perform a $n$-Delaunay tessellation on $K + 1$ normalized parameters to get the organization of the $K$ expressions (the +1 is due to the neutral face). Different subjects’ organizations are presented in figure 2.

To compute the similarity index noted $S_i$ (where $i$ is the index of the subject), we compare the connections between the expressions, that is the sets of the edges of the $n$-Delaunay tessellation. For the following 3-Delaunay tessellation (where 1 is the label of the neutral face) we have

$$DT_i = \{\{1/2/3/4\}, \{1/2/3/7\}, \ldots\}$$

the sets of edges between expressions are

$$S_i = \{\{2/3\}, \{2/4\}, \{3/4\}, \{2/7\}, \{3/7\}, \ldots\}$$

We use Sorensen’s similarity index which is applied to presence/absence of edges between the Delaunay tessellations of two subjects. The similarity index between two organizations is defined by

$$Q(S_i, S_j) = \frac{2 |S_i \cap S_j|}{|S_i| + |S_j|}$$
where $|S_i|$ is the number of edges of the $n$-Delaunay tessellation of subject $i$ and $|S_i \cap S_j|$ is the number of edges in common of the $n$-Delaunay tessellation of both subject $i$ and $j$. The factor 2 allows to have an index between 0 and 1.

We define the similarity index of one organization with all the $P-1$ others by the mean value of the similarity indexes.

$$Q(S_i) = \frac{1}{P-1} \sum_{k=1,k\neq i}^{P} Q(S_i, S_k)$$

(4)

The person-independent organization of the facial expressions is defined as the organization $S_s$ corresponding to the higher similarity index.

$$Q(S_s) = \max_{i=1..P} Q(S_i)$$

(5)

An experiment was conducted on 8 expressions displayed by 17 subjects with a 3D Delaunay tessellation ($n = 3, K = 8, P = 17$). Choosing $n = 3$ allows to have sufficiently connected expressions without having all the expressions connected (around 18 out of 28 possible edges). Figure 4 shows in grey the distribution of the similarity index of the 17 subjects ($Q(S_i, S_s), i = 1..17$). The similarity index is between 0.82 and 1. The differences between the organizations $S_i$ and $S_s$ are most of the time due to substitutions of one edge whose configurations keep the same neighborhood (see figure 3). As comparison, one transposition of two neighboring vertices with 5 neighbors each would have lead to an index of 0.78. The organizations with an index between 0.8 and 1 can then be considered as similar to $S_s$. The distribution of the similarity index between $S_s$ and 10 000 organizations of 8 random parameters is shown in black. As the random organizations are very different from real ones and all the reals ones are similar to $S_s$ (similarity index over 0.8), the organization $S_s$ can then be considered as the person-independent organization.

4. EXPRESSION RECOGNITION BY ORGANIZATION BASED METHOD (OBM)

As similar expressions are organized in the same way from one person to another, one expression can be defined by its relative position to other expressions. We use this property to associate a person-independent signature to an expression (see figure 5).

To reduce the dimensionality, we first project the appearance parameters of the expression on a person-specific manifold compliant with the organization $S_s$. The person-specific manifold is the piece-wise linear manifold formed, in the appearance space, by the simplexes of the $n$-Delaunay tessellation $DT_s$ that corresponds to the person-independent organization $S_s$. In dimension 3 ($n = 3$), the manifold is a collection of connected tetrahedra, each tetrahedra being formed by the parameters of the neutral face and of three known expressions. Each new expression of the subject (named 'unknown expression') can be approximated by a point on this piece-wise linear manifold as follows.

As the manifold is piece-wise linear, each simplex behave locally as a linear space. For a new expression, we perform an orthogonal projection of the appearance parameters onto each simplex of the manifold. The kept projection is the one with the minimum of error.

The direction-intensity signature of an expression is computed from the projection on the manifold:

1. The direction is given by the barycentric coordinates on the outer surface of the manifold.
2. The intensity is given by the norm of the vector 'neutral-expression' normalized such as the outer surface of the manifold has an intensity of 1.

The direction is given by $K$ components formed by the $n$ barycentric coordinates and $K-n$ coordinates to null (where $K$ is the number of the known expressions used to compute the organization of expressions and $n$ is the dimension of the manifold). The intensity is given by 1 component. The direction-intensity signature is a $K+1$ vector (see figure 5).
5. EXPERIMENTAL RESULTS

Experiments have been performed on a new database. The database consists of 22 expressions made by 17 adults aged from 20 to 55 years old (see figure 1). About 30% females are included. Most of the subjects are Caucasian. Some subjects wear glasses and some have a beard. Subjects were shown 22 pictures of expressive face of the same person and were instructed to perform a series of 22 facial expressions by copying as closely as possible the displayed expressions.

This section compares the experimental results of 14 unknown blended expression recognition computed from shape parameters (ABM, cf. section 2) and from direction-intensity signatures (OBM, cf. section 4). 8 of the 22 expressions of the database are used to compute the manifolds in OBM ($K = 8$), the other 14 are used for recognition tasks ($U = 14$).

Figure 6 shows the confusion matrix on shape parameters (a) and on the direction-intensity signatures (b). The 14 unknown expressions are recognized with 47% rate in average with ABM and 60% with OBM (chance would lead to 7%). The direction-intensity signature (OBM) increase by 13% the recognition rate on unknown expressions. These results show that the position of one expression to the others is a more relevant information to qualify an expression than the absolute position of this expression and confirm that the expressions are organized in the same way between different subjects. Figure 6 (b) also shows that the confusion appears when the expression corresponds to the same kind of emotion.

6. CONCLUSION

This paper has presented a new invariant representation of facial expressions and hence a new approach for facial expression analysis. The representation is based on the organization of the facial expressions. A person-specific manifold of expressions compliant with the organization of the expressions has been proposed and used to define a direction-intensity signature of an expression. The proposed method shows better performance in expression recognition than an appearance based method. In the future, we will extend the method to unknown subjects’ expression analysis.

7. REFERENCES