ABSTRACT

This paper proposes a new method for the analysis of blended expressions with varying intensity. The method is based on an asymmetric bilinear model learned on a small amount of expressions. In the resulting expression space, a blended unknown expression has a signature, that can be interpreted as a mixture of the basic expressions used in the creation of the space. Three methods are compared: a traditional method based on active appearance vectors, the asymmetric bilinear model on person-independent appearance vectors and the asymmetric bilinear model on person-specific appearance vectors. Experimental results on the recognition of 14 blended unknown expressions show the relevance of the bilinear models compared to appearance-based methods and the robustness of the person-specific models according to the types of parameters (shape and/or texture).

Index Terms— Facial expression space, Bilinear decomposition, Blended expressions, Person-specific models

1. INTRODUCTION

Facial expression analysis has grown significantly these last decades, especially for the recognition of emotions. Indeed, the work of Ekman [1] showed that the six basic emotions (joy, anger, surprise, fear, sadness and disgust) were universally expressed by six corresponding facial expressions. The recognition of these 6 prototypic expressions on known subjects is now taken for granted, as FERA 2011 challenge shows [2]: the person-specific best performance was 100%. Unfortunately, those prototypic expressions are rarely displayed in everyday life, but blended expressions with variable intensity are (as displayed in figure [1]). This is also the case of emotions, that are often more subtle that those 6 categories. That is why systems evolved towards a dimensional representation of emotions (activation, valence, evaluation). AVEC 2012 challenge [3] aimed at detecting the variations of four affect dimensions (valence, power, expectancy and arousal) of subjects speaking to an emotional agent. In AVEC 2012 challenge, the winners’ average correlation between ground truth and prediction only reached 0.45, which shows that real life emotion prediction is still a challenging problem. With such representations, we do not have a direct link between expression and emotion anymore. A two step process (facial feature extraction and emotion detection) is no more adequate. An intermediate step has to be added: expression space representation. In this paper, we focus on this key step. The expression space is large and continuous, thus it is not possible to learn all expressions. For this reason, we propose a system, that has a restricted learning database but yet deals with the expressions not included in any training database (called blended expressions). In out method, we assume that 8 expressions of each subject are known. As we need to deal with subtle differences between expressions, we use appearance features based on Active-Appearance Models (AAMs) [4]. AAMs are known to provide important spatial information of key facial landmarks. In this paper, we use manual annotations of the faces to avoid problems due to AAM tracking. The last step of the system (emotion prediction) is not addressed in this paper, but a basic recognition algorithm is implemented instead to measure the relevancy of the expression space representation.

It is commonly admitted that the expressions are similar across people and that the identity is person-specific. The most popular invariant representation to characterize uniquely an expression is the FACS system [5]. In FACS, one expression is characterized by a combination of action units (AU). In systems using AUs as the intermediate step for expression representation, the question of non-prototypic expressions is rarely addressed, taking the FACS combinations for granted. Moreover, AUs detection is still a challenging problem as FERA 2011 challenge illustrates (best recognition rate of 62%) [2]. Liu and Wu [6] showed that, with their method, training AU6 and AU12 simultaneously performed better for smile deceit detection than training AU6 and AU12 separately, which makes think that AU description is not adapted to Computer Vision.

To overcome morphological differences between subjects in a more unified manner, neutral face subtraction on facial features is often performed. Cheon & Kim [7] aligned different subjects’ expressions by using Diff-AAM before manifold learning and recognition tasks. Diff-AAM features are the difference of active appearance model (AAM) features between the input face image and the reference face image (neutral face). They represent the main distortions of the face. These methods assume that all people have similar patterns of facial expression changing from neutral expression to a specific expression.
Manifold learning methods have also been used for facial expression representation. The facial features are projected on a low dimensional subspace which is optimal for low cost and accurate classification. Such manifolds take into account the mixture of expressions and the notion of intensity. Stoiber et al. [8] embedded AAM features into a disc. The resulting space leads to promising results for animation purpose but is manually labeled and dedicated to a subject. [9] addressed the question of blended expressions, that are not included in the training database. They created a person-specific manifold for each subject by Lipschitz embedding applied on video sequences displaying transitions from the neutral face to one of the 6 basic expressions. The transitions between the neutral face and the six basic expressions represented 6 paths on the manifold. Blended expressions with varying intensities lied between those paths. Their experiments were limited to only five subjects and they did not measure the adequacy of the representation of the blended expressions across subjects. Moreover, embedding methods need a high density of training data to compute a manifold that properly approximates the organization of the expressions.

Finally, another way of extracting expression features from the facial features is to separate expression and identity via bilinear models. The advantage of bilinear methods is that they need few training images. Wang and Ahuja [10] used HOSVD (High Order Singular Value Description) to decompose appearance features similar to AAM features into a person subspace and an expression subspace. They used the resulting model to synthesize facial expressions of a new subject and to recognize simultaneously face and expression. Ab-boud and Davoine [11] used bilinear symmetric and asymmetric models on AAM features to perform both facial expression recognition and synthesis. Mpiperis et al. [12] decoupled face and facial expression by bilinear models applied on elastically deformable model. They used the resulting model to perform face and expression recognition simultaneously. For each of these methods, the recognition was performed on the 6 basic expressions and neutral face but it was never mentioned how the system performs with blended expressions.

AUs representation, widely used by psychologists, seem to be not adapted to Computer Vision, for AUs recognition still lead to poor results, and mixed AUs are still a matter of concern. Representations of the expressions in a unified manner seem more performing. In this paper, we propose a representation that uses the distortions of the whole face. Separating expression and identity is then the key problem. Manifold learning methods need a high density of training data to cover the whole expression space. That is why we preferred a method based on bilinear models. Even if bilinear models have been already used for expression analysis, tests were limited to the 6 basic expressions. Here we propose to analyze how the system behaviors with expressions not included in the training database, and improve bilinear models by using person-specific appearance vectors. Finally, contrary to the other methods that deal with blended expressions, we measure the relevancy of the representation we propose by computing the recognition rate of blended unknown expressions.

The main contribution is the application of asymmetric bilinear decomposition on person-specific appearance parameters to uniquely characterize non-prototypic expressions. The originality of the approach is to determine a unique signature of a blended expression via a bilinear model learned on a limited number of expressions and to give meaning to the components of this signature. Another important contribution of this work is the discussion about the impact of texture for expression characterization. The particularity of the approach is the use of person-specific appearance models that limits this impact.

The rest of the paper is organized as follows. In Section 2 we briefly justify the importance of moving from an general appearance space to an expression space. Section 3 presents the learning phase of the asymmetric bilinear models. Section 4 describes the signature of a blended expression. Section 5 demonstrates the relevance of the system and section 6 concludes the paper.

2. APPEARANCE BASED METHODS (ABM ET DABM)

This section justifies the importance of moving from an appearance space to an expression space by analyzing the results of blended expression identification with two traditional methods. The first one (Appearance Based Method ABM) is based directly on the active appearance vectors. The second one (Differential Appearance Based Method DABM [7]) is based on differential active appearance vectors, that is to say, the appearance vectors of the subjects are aligned by subtracting the neutral expression vector of the subject.

To obtain comparable appearance vectors for the expressions of different subjects, we use a generic active appearance (AAM) [4], learned on a few expressions displayed by the subjects of the test set. The testing phase is performed on the appearance vectors of blended expressions, different from the ones used for the learning phase of the model (see process in figure 2).

To analyze the relevance of these vectors, the recognition task is performed by a basic algorithm:

1. we compute for each expression of a given subject, the closest expression (nearest neighbor) of each of the other subjects (leave-one-subject-out method) ;
2. we determine the recognized blended expression by a voting algorithm ;
3. we then compute the recognition rate.

As we will see in section 5 the ABM method gives low recognition rate, showing that the appearance vectors are not directly comparable across different subjects. Indeed, they carry both information about the expression and about the
subject’s morphology. The DABM method gives better results. Its purpose is to remove the influence of a face’s morphology from its appearance vectors, thus making the appearance space a purely expression description space. As intuition suggests, removing non-expressional components considerably improves recognition results. In the following, we propose a new parameterization that further improves the separation of identity and expression.

### 3. ASYMMETRIC BILINEAR DECOMPOSITION

To separate the expression and the morphology, we propose to use asymmetric bilinear decomposition on appearance parameters [13]. Given a corpus of \( E \) known facial expressions of \( P \) subjects, we want to decompose the appearance parameters \( y^{pe} \) of dimension \( K \) into a signature of the expression \( b^e \) of dimension \( E \) that is common to all the subjects and a linear mapping specific to the subject \( W^p \) of dimension \( K \cdot E \):

\[
y^{pe} = W^p.b^e
\]

or in matrix form:

\[
Y = W.B
\]

where \( Y \) is of the form

\[
Y = \begin{pmatrix}
y^{11} & y^{12} & \cdots & y^{1E} \\
y^{21} & y^{22} & \cdots & y^{2E} \\
\vdots & \vdots & \ddots & \vdots \\
y^{P1} & y^{P2} & \cdots & y^{PE}
\end{pmatrix}
\]

and \( W \) and \( B \):

\[
W.B = \begin{pmatrix}
W^1 & W^2 & \cdots & W^P
\end{pmatrix} \cdot \begin{pmatrix}
b^1 \\
b^2 \\
\vdots \\
b^E
\end{pmatrix}
\]

The singular value decomposition (SVD) of the matrix \( Y \) can solve this problem. By SDV:

\[
Y = U.S.V^t
\]

\( W \) is then given by U.S and B by \( V^t \). Each vector \( b^e, e = 1..E \) of B is the signature of the expression \( y^{pe} \). It is the same for all subjects in the training set (\( p = 1..P \)).

### 4. BLENDED UNKNOWN EXPRESSION ANALYSIS BY BILINEAR DECOMPOSITION (BDM AND SBDM)

This section describes the process used to compute the signature of an unknown blended expression of a known subject.

**Learning phase**

The learning phase is achieved by the asymmetric bilinear decomposition presented in section 3 on \( E \) known similar facial expressions of \( P \) subjects.

**Analysis phase**

For one subject \( p_i \) of the learning base, we want to analyze an unknown blended expression \( y^{p_i,e} \) (different from one of the \( E \) expressions of the learning phase). The signature \( b^e \) of the expression \( y^{p_i,e} \) is estimated with the matrix \( W^{p_i} \) of the learning model.

\[
b^e = (W^{p_i}.W^{p_i})^{-1}.W^{p_i}.y^{p_i,e}
\]

**The bilinear methods**

When the computation of the learning and analysis phase is made on the appearance parameters of a generic AAM, we call the method BDM (Bilinear Decomposition Method). When it is performed on the diff-aam parameters, we call the method BDDM (Bilinear Decomposition Method on Differential parameters). As we only use the linearity property of the expressions, we can apply the asymmetric bilinear decomposition on person-specific appearance parameters. For each subject, we then calculate a specific appearance model (AAM) learned on some expressions of the subject (the same expressions as for the other methods). When the computation is made on the appearance parameters of person-specific AAM, we call the method SBDM (Bilinear Decomposition on person-specific parameters). When it is performed on the diff-aam parameters, we call the method SBDDM (Bilinear
Fig. 2. Overview of the methods compared in this paper (ABM: appearance based method, DABM: differential appearance based method, BDM: bilinear decomposition, BDDM: bilinear decomposition on differential parameters, SBDM: bilinear decomposition on person-specific parameters, SBDDM: bilinear decomposition on person-specific differential parameters).

<table>
<thead>
<tr>
<th>Method</th>
<th>Shape</th>
<th>Shape+Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABM</td>
<td>36.2</td>
<td>25.4</td>
</tr>
<tr>
<td>DABM</td>
<td>54.1</td>
<td>34.0</td>
</tr>
<tr>
<td>BDM</td>
<td>57.9</td>
<td>32.1</td>
</tr>
<tr>
<td>BDDM</td>
<td>58.7</td>
<td>32.5</td>
</tr>
<tr>
<td>SBDM</td>
<td>57.7</td>
<td>59.9</td>
</tr>
<tr>
<td>SBDDM</td>
<td>59.1</td>
<td>57.3</td>
</tr>
</tbody>
</table>

Table 1. Mean recognition rate of 14 unknown expressions on known subjects according to the type of appearance parameters (shape or shape+texture).

Decomposition on person-Specific Differential parameters). Figure 2 shows the link between the methods.

A meaningful signature

The matrix \( B \) is by construction a matrix with orthonormal vectors and can therefore be interpreted as an orthonormal basis. Each signature can then be characterized in this basis on the form of a relative signature \( b' \) with respect to the \( E \) expressions used in the learning phase. By rotation, we have:

\[
b' = B^T b
\]

(7)

Thus, the relative parameters \( b' \) can be interpreted as a mixture between several known expressions.

5. EXPERIMENTAL RESULTS

This section compares the signatures of 14 blended expressions on appearance based methods ABM and DABM (section 2) and on bilinear decomposition methods BDM, BDDM, SBDM and SBDDM (section 4).

Database

The experiments were performed on a public database\(^1\) containing 22 blended similar expressions plus neutral on 17 subjects aged between 20 and 55 (see figure 1). The database contains about 30% of female. Most subjects are Caucasian. Some subjects wear glasses and others have beards. They were showed 22 pictures of expressive faces and were asked to reproduce as closely as possible the displayed expressions. Having 22 expressions allows to have enough expressions to build the expression model and to test it. The expressions have also been manually annotated with one of the 6 basic emotions.

Learning and recognition phase

The models were trained on 8 expressions plus neutral face of the 17 subjects. The other 14 remaining blended expressions were used to test the models (see figure 1). To measure and compare the relevance of the signatures, the same basic recognition algorithm (described in section 2) is used for all the methods.

Comparison of the methods

Table 1 shows the best recognition rates for each method on shape vectors. As expected, the ABM method gives the worst results. Indeed, the appearance parameters include expression information but also subjects’ morphology. Bilinear decomposition methods give better results than DABM. This can be interpreted by the fact that the differential appearance vectors still contain a significant amount of information about identity, such as how the subject smiles (crescent smile or flat smile). In contrast, bilinear models learn from the expressions used in the learning phase, how each subject achieves the expression. This type of deformation is then specific to the subject and is used for blended expression recognition. We note that the bilinear decomposition applied to the differential appearance vectors gives slightly better results, but this difference is not significant enough to be relevant.

Shape vs. texture

The tests were performed on shape vectors only and on shape and texture vectors. Figure 3 shows, for each method, the recognition rate depending on the size of the vectors and table 1 the best recognition rate. We note that for all methods performed on vectors from person-independent appearance models, adding texture information significantly reduces the

\(^1\)Database available at [http://www.rennes.supelec.fr/immemo/](http://www.rennes.supelec.fr/immemo/)
recognition rates. DABM and ABM results can be compared to the work of Cheon & Kim [7]. Contrary to our results, they found an improvement in recognition rate when the texture information was added. This can be explained by the fact that the database we used contains a greater variety of subjects than theirs (which contained only Korean people) and texture vectors thus contain too much information of identity to be relevant and comparable between subjects. One might expect that the bilinear methods would eliminate this difference and would give similar results whether applied to shape vectors or shape and texture vectors. This is not what is observed for bilinear methods on person-independent appearance vectors. In contrast, the use of bilinear models on person-specific appearance vectors gives similar good results. Indeed, the texture variations are mainly due to identity variations in person-independent models, whereas they are due to expression variations in person-specific models.

**Expression space dimension**
Methods based on bilinear decomposition determine in advance the size of the expression signature (here 8 or 9), contrary to the appearance based methods whose results often require a large number of parameters. Although in our tests, the best recognition rate with DABM method is obtained for a parameter size of 10, it is more common to have up to 20 or 30 to get satisfying results [11], this value depends greatly on the learning database.

**Interpretation of the signature of an expression**
Figure 4 shows the relative signatures of different expressions obtained by equation 7. In each case, the first two lines show the signatures of expressions of the learning phase, the last line shows the signature of a blended expression of the test phase (either a mixture of different expressions, or an expression with a different intensity). We can notice that the observed mixture in the expression is reflected in the parameters of the signature (highest components).

**Confusion**
Figure 5 shows the confusion matrix on the signatures with DABM. We can notice that the confusion mostly appears
when the expression corresponds to the same kind of emotion. For some expressions, we experienced the classical confusion between Anger & Disgust and between Fear & Surprise.

6. CONCLUSION

This paper has presented methods for blended expression analysis based on asymmetric bilinear decomposition, that efficiently separates expression and identity. For each blended unknown expression, the method computes a signature, that characterizes uniquely the expression. This signature can thus be used to recognize blended expressions with varying intensity. Contrary to appearance based methods, the dimension of the signature is small (8 parameters in our tests) and could thus be used for real-time application. Moreover, bilinear models compute a signature that can be interpreted: the value of the components of a blended expression indicates the expressions of the learning phase that are mixed. Finally, the bilinear method applied on person-specific appearance models is robust to the type of features (shape and/or texture), because person-specific models reduce the variations of texture due to the subjects’ morphology. The recognition rate of 59% may seem low and with no practical use. Yet, in a complete system, the expression is computed for each image of a video sequence and is integrated in time with a minimum period of 1 second (that is around 30 images). The information is then accurate enough to detect the emotion of the subject. In the future, we will extend the method to unknown subjects.

7. REFERENCES


