Smart Cameras: 2D Affine Models for Determining Subject Facial Expressions

Ioana Bacivarov, *Student Member,* IEEE, Peter Corcoran, *Senior Member*, IEEE, and Mircea Ionita, *Student Member,* IEEE

Abstract — *In this paper we consider how next generation of smart cameras will move beyond face detection and tracking to develop more useful analysis tools for determining the expressions and emotions of the subjects in images. An initial proof-of-concept using an extension of an Active Appearance Model (AAM) to measure facial parameters and a set of classifiers to determine facial states are described. Preliminary findings are encouraging and suggest that the use of the component-based AAM models and SVM classifiers will enable new applications and opportunities in smart cameras*[1].

Index Terms — smart cameras, affective imaging, expression classification and recognition, AAM, SVM.

I. INTRODUCTION

In the last year or so face-tracking technology has become a common feature of consumer digital cameras. Such camera subsystems are typically derived from Haar classifier based face detection algorithms, a technique originally pioneered by Viola and Jones [8]. The most recent implementations feature hardware coded face tracking in an IP-core [29].

Applications to date have been limited to optimizing the exposure and acquisition parameters of a final image [13]. Yet there are many additional applications which have even greater potential to enrich the user experience. The rapid deployment of face tracking technology in cameras suggests that other more advanced face analysis techniques will soon become feasible and begin to offer even more sophisticated capabilities in such consumer devices.

The detailed analysis of facial expression is one such technique which can offer a wide range of new consumer applications for mobile and embedded devices. In the context of managing our personal image collections it is useful to be able to sort images according to the people in those images [12]. It would be even more useful if it were possible to determine their emotions and thus enable images to be further sorted and categorized according to the emotions of the subjects in an image.

Other consumer device applications also can gain from such capabilities. Many such devices now feature a camera facing the user, e.g. most mobile smart-phones, and thus user-interfaces could respond directly to our facial expressions. An e-learning system could match its level of difficulty to the degree of puzzlement on the student's face. A home health system could monitor the level of pain from an elderly person's facial expression. As the underlying expression recognition technologies improve in accuracy, the range of applications will grow further.

In this article, initial studies and results on facial expression modeling are presented. Parameters describing human faces are extracted using Active Appearance Models (AAM) [1]. These models are designed to handle small to medium head pose variations and to be robust to illumination changes. Our goal is to provide a general model, sufficiently complex to be able to compensate for variations in image acquisition conditions, but simple from the point of view of algorithmic complexity and speed of execution. The parameters are classified using Nearest Neighbors based classifiers (NN) and Support Vector Machines (SVM), in order to decode facial expressions. The ultimate goal is to abstract a model which is sufficiently robust to justify implementation as an IP-core hardware module.

This paper is organized as follows. Section II briefly reviews the literature and explains the main ideas and the original aspects of our approach. Section III and IV describe the facial features extraction algorithm and its optimization to expression decoding. An overview of our system is given in Section V. In Section VI, the preliminary results are summarized together with some other applications of our model, e.g., expression tracking. In the last section we draw some conclusions and present our future work.

II. STATE-OF-THE-ART

Facial expressions are generated by contractions of facial muscles, which result in temporally deformed facial features, such as eyelids, eyebrows, nose, lips, and skin texture, e.g., bulges, blushing, expression wrinkles. Various approaches try to transfer this knowledge to computers.

Specialized literature is rich in automatic algorithms dedicated to facial expression recognition, analysis, and synthesis in static images or in video sequences. However none of the proposed methods is able to fully satisfy the requests of an ideal system. To attain a high level of recognition performance, most current expression recognition approaches require some control over the imaging conditions. Such control is achieved in terms of illumination conditions, keeping frontal non-occluded faces etc. However, this need for precise control over imaging conditions restricts the

[1] Peter Corcoran is with the Dept. Electronic Engineering, National University of Ireland, Galway (e-mail: peter.corcoran@nuigalway.ie).

Ioana Bacivarov and Mircea Ionita are with the Dept. Electronic Engineering, Natuonal University of Ireland, Galway (e-mail: ioana@wuzwuz.nuigalway.ie and mircea@wuzwuz.nuigalway.ie).

deployment of expression recognition systems and reduces their effectiveness. In particular, consumer electronic applications require operational flexibility and the ability to operate on real-life images acquired in non-ideal conditions. However, few studies have systematically investigated the robustness of automatic expression recognition under real conditions.

Comprehensive surveys reviewing facial expression analysis exist in literature [19, 21]. To sum up, the approaches differ from point of view of the feature extraction methods, the employed types of features, or the classification algorithms. Several types of features can be extracted, i.e., *geometric* features [19], *appearance* features [20], and *hybrids* of the two [2-7]. Geometric features describe the shape of facial components and their locations. They can be computationally more expensive, but they are more robust to variations in face position, scale, size, and head orientations. Appearance features describe the appearance, i.e., skin texture, and the facial changes, e.g., wrinkles, furrows, bulges. The appearance features can be extracted on either the whole-face or on specific facial regions. The appearance-feature extraction is typically performed using Gabor wavelets [20]. Hybrid features are described by both geometric and appearance features. AAM belongs to the hybrid feature extraction category.

AAM is a popular technique to parameterize faces, as the algorithm separates geometric modeling from texture information. The method is recommended by its wide applicability, compact representation of allowable variations of the model, speed, accuracy, and robustness. It can be relatively easily implemented. Facial features are detected quickly and accurately; AAM yields good results on difficult and noisy data. The AAM approach also has the advantage that we can control the amount of details, or variance, of the model by changing its number of modes of variations, i.e., number of eigenvectors. There are many methods described in the literature which adapt AAM techniques to expression analysis. 3-D AAM deformable models [2], 2-D conventional AAMs [3-5], or extensions of AAM separating expressions from other types of variations [6-7], have been proposed.

Now, facial expression analysis can be done considering faces as a whole or applied to specific facial regions as sub-models. Instead of using a holistic representation, Zalewski and Gong [22] define expressions as a combination of intrinsic functionalities of the subcomponents, organized in a hierarchical decomposition. The idea of a hierarchical model is also employed in [23], so that major appearance modes of each sub-facial model encode the major variation for the corresponding area. For example, the highest mode of variation in the left-eye model encodes a blink. In [24], a multi-view shape model with expression variations is proposed. A component shape model, such as a mouth shape and an eye-contour shape, is used, in addition to a global shape model, to achieve more powerful representation for face components, under complex poses and expression variations.

In our work, we extend the AAM approach for facial feature extraction. Although AAM is a powerful tool for image interpretation, the conventional algorithm presents several drawbacks when used as a global appearance model for general conditions, e.g., consumer pictures. Also, when dealing with facial expressions, local features are responsible for most of the facial variation. With AAM, faces can be too closely constrained to global variations learned during the training phase. Unseen individual variations of facial features might not be perceived by the model. Component-based AAM [18] is presented as a solution to these particular drawbacks. It combines the global model with a series of sub-models. These sub-models are typically component parts of the object to be modeled. This approach benefits from both the generality of a global AAM model and the local optimizations provided by its sub-models.

We propose the component-based AAM approach in the context of facial expression decoding. We adjust it so to make it robust to small to medium pose variations and to directional illumination changes. We then demonstrate the benefits of our system, with respect to expression classification and recognition, while trying to take into account the needs of the system that is due to perform in real-time and which will be suitable for incorporation in future consumer imaging devices.

III. The Extraction of Facial Features using AAM

A. Statistical models of appearance (AAM)

AAM was proposed by Cootes et al. [9] in 1998 as a deformable model, capable of interpreting and synthesizing new images of the object of interest. Statistical Models of Appearance represent both shape and texture variations and correlations between them. The desired shape to be modeled is annotated by a number of landmark points. The shape is defined by the number of landmarks chosen to best depict the contour of the object of interest, in our case the eye region. A shape vector is given by the concatenated coordinates of all landmark points and may be formally written as, $s=(x1,x2,\ldots,xL,y1,y2,\ldots,yL)^T$, where L is the number of landmark points.

The shape model is obtained by applying Principal Component Analysis (PCA) on the set of aligned shapes:

$$s = \bar{s} + \varphi_s b_s \quad , \qquad (1)$$

where $\bar{s} = \frac{1}{N_s}\sum_{i=1}^{N_s} s_i$ is the mean shape vector, and N_S is the number of shape observations; φ_s is the matrix having the eigenvectors as its columns; b_s defines the set of parameters of the shape model.

The texture, defined as the pixel values across the object of interest, is also statistically modeled. Face patches are first warped into the mean shape based on a triangulation algorithm.

Then a texture vector $t=(t1,t2,\ldots,tp)^T$ is built for each training image by sampling the values across the warped, i.e. shape normalized patches, with p being the number of texture samples.

The texture model is also derived by means of PCA on the texture vectors:

$$t = \bar{t} + \varphi_t b_t \quad , \qquad (2)$$

where $\bar{t} = \frac{1}{N}\sum_{i=1}^{N_t} t_i$ is the mean texture vector, with N_t as the number of texture observations; φ_t is the matrix of eigenvectors, and b_t the texture parameters.

The sets of shape and texture parameters $c = \begin{pmatrix} W_s b_s \\ b_t \end{pmatrix}$ are used to describe the overall appearance variability of the modeled object, where W_s is a vector of weights used to compensate the differences in units between shape and texture parameters.

After a statistical model of appearance is created, an AAM algorithm can be employed to fit the model to a new image. AAM is a fast technique used to interpret unseen images using the appearance model, by finding the best match of the model to the image. Therefore, the algorithm allows us to find the parameters of the model which generate a synthetic image as close as possible to the target image. The whole problem is treated as an optimization problem in which we want to minimize the difference between the new image and the image synthesized by the appearance model.

B. Relevant AAM Parameters for illustrating emotions

AAM extracts two types of features: geometric features and features depicting the facial appearance. *Geometric features* describe shapes, deformations and locations of facial components, and poses variations. These features are highly affected when expressing emotions. For example, when surprised, eyes and mouth open widely, the latter resulting in an elongated chin; when sad, we often blink; when angry, eyebrows tend to be drawn together. FACS originally developed by Ekman and Friesen in 1976 [15], describes all deformations of facial features involved in facial expressions. *The appearance features* describe skin texture changes, e.g., furrows and bulges, blushing, expression wrinkles, illumination variations. Both types of features are important when fitting the AAM model to unseen pictures.

The performances of shape, texture, shape and texture modeled independently, or combined, are tested in our experiments. We have searched for the best compromise between a reduced number of parameters and high expression decoding accuracy. Subsequent to these tests, the shape parameters proved the most valuable in terms of facial expression decoding accuracy and a reduced number of parameters. The results for the shape parameters are comparable with the results obtained when applying a combined shape and texture model, but the number of the shape parameters is less. The shape parameters superiority is also confirmed in [22].

Furthermore, both shape and texture are employed by AAM to extract facial features, but only shape information will be used in expression analysis. Shape parameters present the advantage that, to a certain extent, they separate between the main sources of variations, e.g., poses, individual characteristics, facial expression variations. This theory implies that not all shape parameters are relevant for expression classification. A number of parameters are redundant and some can even deteriorate the system accuracy.

TABLE I-CLASSIFICATION ACCURACIES (%), WHEN USING SVM ON FEEDTUM [9] , FOR SHAPE PARAMETERS, TEXTURE PARAMETERS, SHAPE AND TEXTURED PARAMETERS CONCATENATED INDEPENDENTLY, AND A JOINT COMBINATION OF THE TWO.

Classifier	Shape	Texture	Sh. & T. Independent	Sh. & T. Combined
H/Non-H	92.7	80	97.4	86.7
D/Non-D	69.4	64	67.4	67.4
Su/Non-Su	72	64.7	82	71.4
A/Non-A	90	76	45.4	79.4
Sa/Non-Sa	68.7	45.4	66.7	67.4
F/Non-F	72	74	78	77.4
N/ Non-N	61.4	64	70	69.4
Average	75.2	66.9	72.5	74.2

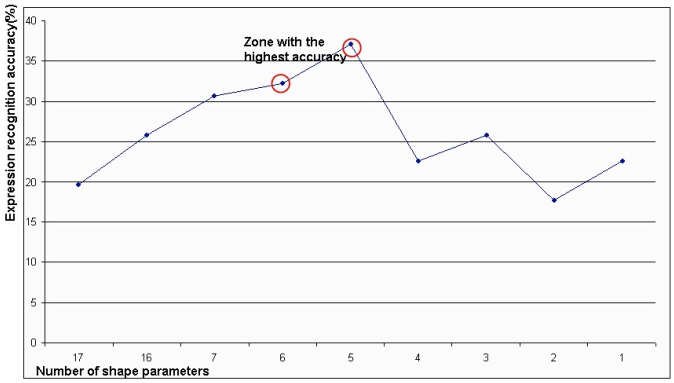

Fig. 1. Expression recognition accuracies versus the number of shape parameters.

An educated choice of parameters affects positively the system performances. We empirically study the parameters best depicting expression variations throughout AAM shape parameters. The results are presented in Fig. 1, where the NN expression recognition rate is plotted versus the employed number of parameters, 17 in total. Then, the number of parameters is reduced, eliminating the ones corresponding to the smallest variation. In our case, it is proven empirically that the first 8 parameters, excluding the first one, provide higher expression classification and recognition accuracies.

IV. REPRESENTATION OF FACE GEOMETRY TO OPTIMIZE EMOTION DECODING

A. Relevant facial features to indicate expressivity

Facial expressions are defined by the dynamics of the individual facial features. Psychophysical experiments [26] indicate that the eyes and the mouth are the most relevant facial features in terms of facial expressions. Experiments show that, in some cases, one individual facial region can entirely represent an expression. In other cases, the interaction of more than one facial area is needed to clarify the

expression. We conduct a study in order to verify the amount of information individually brought by facial features in an automatic expression recognition system.

A thorough description of the eye area is done using an AAM model, as described in our previous work [16]. A detailed AAM lip model, including a hue filtering is proposed by us for the lips area in [17]. Each of these independent models is used now for expression recognition.

The results confirm the fact that the eyes and the mouth are decisive features in facial expressions. In average, the scores obtained for the eyes shape represent 70% of the information contained by the entire face shape. Naturally, the mouth is also an important emotion feature carrier, especially for some emotions, e.g., surprise, with around 60% independent contribution. However, when combined and processed together with holistic facial information, the emotion decoding accuracies increase. Results of expression recognition for eyes, lips, and faces are summarized in Tables II-III (in Section VI.B).

B. Component-based AAM representation of facial expressivity

Component-based AAM [18] is an approach that benefits from both the generality of a global AAM model and the local optimizations provided by its sub-models. These sub-models are typically component parts of the object to be modeled. In addition to our global model, we use component models such as the mouth model, and two eye-models to achieve a more powerful representation for face components under complex poses and expression variations. In these situations, when aligning faces globally, facial features are not simultaneously and accurately aligned. This dramatically affects the performances of the expressions decoder.

In summary, the component-based algorithm is as follows. Two sub-models are built, one for the eye region and one for the lips. The eye sub-model is then derived in a left and, respectively, a right eye model. At each iteration, the sub-models are inferred from the global model. Their optimums are detected by an AAM fitting procedure. Then the fitted sub-models are projected back into the global model.

In a first step, the global AAM is fitted, roughly locating the two eyes and the mouth. The points of the global model which correspond to the sub-models are first extracted to form the three shape vectors, providing a good initialization for each sub-model which is next fitted using the AAM method. As a second step we reintegrate the fitted points of the three independent sub-models back into the global model. Another projection of the global shape vector onto its principal components space is required. This step is necessary in order to constrain the independently obtained eyes and mouth such that they remain within limits specified by the global model. The fitting error for this refined global model is compared with the original global model fitting error and a decision is taken to use the global model with the least error.

We prove that component-based AAM is extremely beneficial in expression modeling (Tables II-IV- in Section VI.B). The principal reason is that this approach improves the AAM fitting accuracy. More precision of the location and deformation of facial feature shape are extremely valuable in expression analysis, where the information is based on local variability.

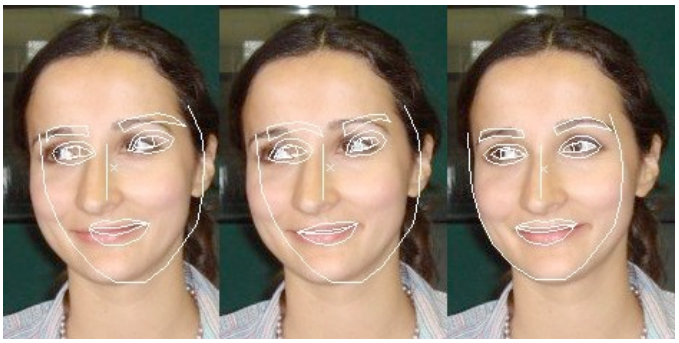

Fig. 2. Example of shape fitting on an unseen picture with expression and pose. The first is the result of a conventional AAM; the second picture represents the fitting of the AAM sub-models, while the third picture depicts the component-based result.

C. Robustness to Pose

In our paper we also address the problem of face posing in expression classification, inherent in consumer pictures which usually illustrate real-life conditions. In order to cope with face orientations, we introduce into the training set annotated image examples containing pose variations. The interval of variation is from small to medium, i.e. semi-profile. One draw-back in our research is the lack of specialized databases containing facial expressions in a range of pose variations. Pictures are taken especially for this experiment, emulating consumer images. Some examples are shown in Fig. 3.

Fig. 3. Examples of consumer pictures containing both small to medium pose variations and expression (from left to right, happy, sad, surprised, and disgusted).

After performing PCA on shape parameters, the information on the out-of-plane head rotation is mainly encoded in the first shape parameter. This is explained by the fact that the variation caused by pose changes is significantly more important than the variation caused by differences between individuals or between facial expressions.

The introduction of pose variation is very important in the AAM fitting and in the facial feature extraction stages. The elimination of the pose parameters from the expression classification stage increases though its accuracy. This results in improvements of the expression recognition rates. In Fig. 4, different modes of shape variations are exemplified. In the first row, the pose variation is showed. The second row depicts examples of expressions and gaze, while the next parameters representing expression variations are visualized in the last row.

Fig. 4. Modes of variation for the face model: pose parameters (first row) and variation of expressions, including gaze, and identity (last rows).

D. Robustness to illumination

In order to deal with illumination problems, we use a technique that we originally proposed in [11]. This approach is used along with the conventional AAM method to build a statistical model. The texture space is separated into two subspaces, one for inter-individual variability and another for variations caused by directional changes of lighting conditions. The model is built in the following manner:

- a shape model and a texture model are built from different individuals at constant frontal illumination
- a set of images with various directional lighting conditions are filtered by projecting the corresponding textures on the previously built space; the residues are used to build a second subspace for directional lighting.

The main advantage of this representation is that two sets of parameters are used to control separately inter-individual variation and intra-individual variation due to changes in illumination conditions. It significantly improves the fitting stage, improvements transmitted to the expression decoding stage.

V. SYSTEM OVERVIEW

An ideal facial expression analysis system, as defined in [14], should be able to handle various lighting conditions, plane and out-of-plane head rotations, occlusions, different image resolutions, etc. Also, if designed for consumer applications, the system has to be fully automatic, able to perform in real-time. AAM has the advantage that many of the algorithm steps can be performed off-line, e.g. training stage, reducing the computational time.

Facial expression recognition is a complex task as it involves multiple steps, i.e. face acquisition, features extraction, and facial expression classification. The computer model, see Fig. 5, automatically finds and registers faces in images, it extracts visual features, it takes binary decisions about the presence of each of seven expressions, i.e., happiness, sadness, fear, disgust, anger, surprise, neutral, and then it takes a multiple class decision.

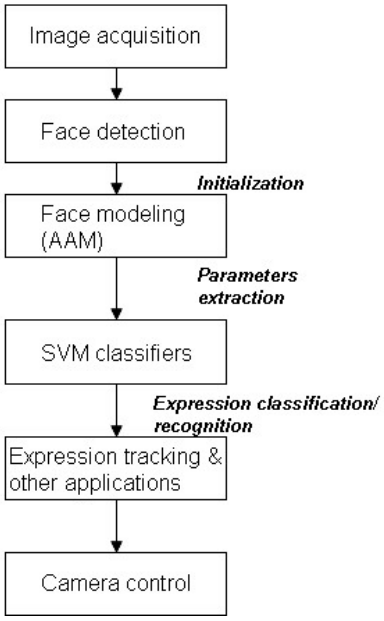

Fig. 5. System overview.

A. Face Detection & Facial Features extraction

The Viola-Jones face detection algorithm [8] is firstly applied on the current image. This algorithm is the state-of-the-art face detector. It achieves a detection rate equivalent to the best published results and it is known for its speed. The algorithm, based on AdaBoost, is a widely used method for real-time object detection. It minimizes computational time, while achieving high detection accuracies.

The face detection step not only detects whether there is a face in the picture, but it also gives a good initialization point for next steps, such as feature extraction. A good initialization point insures, to a certain extent, the convergence of AAM. A statistical relation between the estimates for face position and size of the detected face rectangle, and position and size of the reference shape inside the image frame is initially learnt from the training set of images during the face detection step. This relation is going to be used as starting point for the AAM search.

Facial feature extraction follows; it can be viewed as the task of finding some conceptual and relevant facial information from raw data, in our case the detected face. Facial model parameters are calculated from the extracted facial area, i.e., certain shape parameters describing the face, emphasizing on certain facial features.

B. Expression Classification and Recognition

The last step in an expression decoding system is expression classification and recognition. Two classifiers are compared: SVM and NN. This choice is based on their positive performances in the literature. As we designed them, the classifiers use as input relevant AAM parameters and they present at output the choice between two facial expressions. When dealing with poses, the pose parameters are discarded as described previously.

Despite its simplicity, the NN-based classifier is well-suited

for applications where there are many classes and a reduced number of examples per class, such as in facial expression recognition case. NN is tested based on Euclidean and cosine distances. A comparison is made between the test image and a template corresponding to each universal emotion. These templates are calculated as the mean between all the pictures corresponding to each class. Also, a second option for obtaining the emotion templates is the use of the median.

The SVM technique has been chosen among other learning methods because of its accuracy against other classifiers. The method is well known for its good generalization properties even in cases of high-dimensional nonlinear separable classification tasks. Also, its classification process is instantaneous, thus making it ideal for real-time applications. We trained SVMs with linear, polynomial and RBF (Radial Basis Function) kernels in order to compare performances and chose the best alternative. The optimal kernel function proved to be RBF, with δ fixed on small values.

A comparison between the two classifiers is performed. Experiments prove that the SVM outperforms the NN classifier, especially in more general cases where the training set and the testing set differ. One of our goals is to recognize all the six universal emotional expressions, plus the neutral expression. When applying the classifiers to expression recognition, multi-class classification schemes are proposed. We opt for multi-class SVMs, i.e. one-against-one and cascaded SVM classifiers, and NN-based recognition schemes. Using particular shape parameters, we train a cascade of SVM classifiers. Each binary SVM classifier in the cascade is trained to act as an expression detector, outputting a positive response if its expression is present and a negative response otherwise. The same parameters are used as input for the NN classifier. The results are presented in Section VI.B.

C. Towards Real-Time Implementation

For a real-time system, many factors have to be considered. Our system is not developed with the explicit purpose of performing in real-time, but it takes into account some important real-time aspects, such as a reduced number of input features or a fast modeling algorithm. Ideally it should not require a significant training phase, or else this stage should be performed off-line, on desktop computers. The size of the model is an important issue; to a certain extent it can be reduced by reducing either the size of the training set or the size of the model parameters; still satisfying accuracies can be obtained in these cases.

Some preliminary tests have been performed in order to evaluate the computational requirements involved by the proposed techniques. Although this aspect is not the main focus of this research, and exact figures are not provided, it is believed, and also supported by our preliminary tests, that the full algorithm described can be implemented very efficiently.

Nevertheless, the speed of the algorithm is very important for real-time applications. The speed can be improved either by reducing the computational time per iteration or by reducing the number of iterations. Many SVMs with a large number of support vectors seriously increase the response time. Intelligently selecting a number of features and parameters not only improves the system accuracy, but also can provide affordable computational time. By selecting only representative numbers of parameters, we can significantly save classification computational time and consequently improve the system quality in response speed.

Finally we remark that in a full hardware implementation, additional performance benefits can be achieved through parallelization. The SVM techniques are well suited to such optimizations.

VI. Results and Applications Using Our Model

A. Expression Tracking

Most of the studies on facial expression analysis restrict their research to still images. Still images provide us enough information to accurately recognize facial expressions. Additionally, they are preferable from the point of view of the simplicity of their description, manipulation, and analysis. However, facial expressions are dynamic actions. Moreover there are some expressions, like agreement and disagreement, which require particular motions, e.g., head motion, that cannot be depicted in a single picture.

In our work we simplify the spatio-temporal problem of facial expressivity, by dividing a video sequence into a collection of static snapshots. One may argue that this assumption does not bring any additional information to a static-based system. We consider that by analyzing a series of consecutives frames we learn more about:

- Time segmentation, i.e., automatic detection of the time boundaries of expressive sequences.
- Emotion intensities, i.e., neutral, onset, offset, and apex states. For each of the universal emotion we can define at least two level of intensity: happiness-ecstasy, anger-rage, fear-terror, sad-grief, disgust-loathing, and surprise-startle.
- The algorithm speed is increased, as the AAM fitting is only being applied on the first frame. For the consecutive ones, only parameter updates should be performed.

In expression tracking it is more indicated to use a person-specific model. The model is specialized on the face that we aim to track. Firstly, all parameters, extracted from the face detector, need to be optimized, i.e. the shape parameters, including head poses, and the texture parameters. This stage is the most computationally expensive and it is to be performed over several frames. Afterwards, the shape parameters encoding head poses and expressions variations, as well as position, 2D rotation and scale parameters are only updated. As the computational requirements are substantially reduced, this operation is now performed on each separate frame. Using the updated shape parameters, the expression variation can be tracked and interactively recognized.

An important issue in facial expression analysis is to develop easy-to-train, efficient, and robust tracking algorithms, which have real-time potential implementation. We conduct some initial experiments on the FEEDTUM database. These initial tests show that the system is able to

track a previously unseen person in a subjectively accurate way. The component-based AAM partially solves the out-of-plane variations. The quality of sub-models fitting can improve the overall face model fitting results. The parameters of a partially occluded eye could be inferred from other correct fitted sub-models parameters, by using a relation between the global model and its sub-models.

An important observation is that the tracking algorithm works better if the frame rate is higher. Because of the parameters update stage, the tracker is not able to follow fast and large moves. If the frame rate is higher, then the motion between each frame is smaller. In consequence fewer iterations are needed per frame, so less time is spend per frame. When processing frames captured at a low frame rate, the algorithm suffers from being greedy and sometimes it gets stuck in a local optimum. Around 30 frames per second seem to be sufficient for handling normal head motion.

B. Summary of Results for facial expression decoding

The expression classification and recognition methods are tested on the specialized databases FEEDTUM and MMI [25], while their accuracy against pose variations is tested on pictures collected especially for this experiment (Fig. 4). The results obtained from our tests, suggest that the system is robust in dealing with subjects of both genders. Also, it is independent of race and age. Firstly, the result of eliminating the pose parameter is evaluated. The experiments verify our hypothesis that eliminating the head pose parameters helps decoding facial expressions.

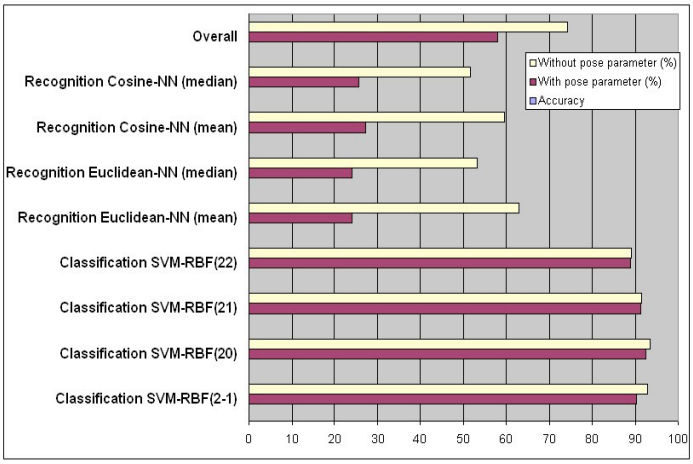

Fig.6. Comparison of classification and recognition performances when using / ignoring the pose parameter.

Then, the AAM sub-models contributions to expression analysis are evaluated. The results are compared with the results of a holistic AAM face model and a component-based AAM facial representation. Table II summarizes the classification results for a mouth model, two-eye model, a conventional AAM face model, and a component-based face AAM when using NN. Their expression recognition equivalent is exposed in Table III. Tables IV and V underline the superiority of a component-based representation, this time when using a SVM.

We consider our results as partial successes, as they are mainly achieved in constraint conditions, hosted by specialized databases. Also, we have to underline that our results correspond to models created on particular databases and tested on the same databases by dividing the data set into training and testing sets.

TABLE II- SUMMARY OF THE SYSTEM ACCURACY (%) FOR CLASSIFYING EMOTIONS WITH EUCLIDEAN-NN (NOTATIONS: SU-SURPRISE, D-DISGUST, H-HAPPINESS, SA-SADNESS)

Classifier	FEEDTUM				MMI			
	Eyes	Mouth	Face	Comp based	Eye	Mouth	Face	Comp based
Su/ D	44.4	55.9	85.9	83.4	73.2	76.7	85.7	78.4
Su/ H	28.8	69.4	88.4	83.4	68.8	75	86.3	79.2
H/Sa	41.9	46.3	68.4	82.3	72.5	50	73.2	77.5
Su/ Sa	33.2	51.7	93.4	83.4	69.4	52.5	77	73.4
Overall	37.1	55.8	83.95	83.1	71	63.6	80.5	77.1

TABLE III-SUMMARY OF THE SYSTEM ACCURACIES (%) FOR RECOGNIZING EMOTIONS WITH EUCLIDEAN-NN

Database	Eyes	Mouth	Face	Component -based
FEEDTUM	27.37	27.37	35.71	54.28
MMI	34.1	26.95	40.88	60
Our database	22.57	27.01	22.66	28.19
Overall	28.01	27.11	33.08	47.49

TABLE IV-ACCURACIES OF THE SVM CLASSIFIERS (%) WHEN USING A CONVENTIONAL AAM

Emotion	H	F	Sa	Su	N	D	A
Happy	70	63	70	66.7	70	70	73.4
Scared		79.4	63.4	70	73.4	53.4	70
Sad			72	66.7	56.7	80	53.4
Surprised				92.7	73.4	60	66.7
Neutral					79.4	53.4	80
Disgusted						68.7	76.7
Angry							71.4

TABLE V- ACCURACIES OF THE SVM CLASSIFIERS (%) WHEN USING A COMPONENT-BASED AAM

Emotion	H	F	Sa	Su	N	D	A
Happy	75	75	75	68	75	79.4	68.7
Scared		73.4	70	75	75	68	71.4
Sad			70	73.4	68.7	83.4	63
Surprised				93.4	75	75	66.7
Neutral					68.7	63	83.4
Disgusted						63.4	75
Angry							73.4

VII. CONCLUSIONS & FUTURE WORK

The main contributions of our research are as follows: firstly, we developed a detailed face representation based on facial sub-region modeling and we selected the relevant AAM features for expression decoding. We present a system which can be used for locating and tracking facial features, coding, and recognizing expressions. We report results on a series of experiments comparing feature selection methods and recognition engines. Our results are promising, even though variations in 3D pose, lighting, and other conditions are allowed, demonstrating potential for use in real-life applications.

Using the AAM approach such as the feature extraction method has proven successful, even with a simple NN Euclidean-distance classification scheme. The capability of AAM to model both shape and texture of faces makes it a strong tool to derive feature sets for emotion-based expression classification. A more sophisticated classifier, such as SVM, proved to be able to provide even more accurate results. In our work, we investigated the effect of reducing the number of features for classification, focusing only on expression specific shape parameters. Having irrelevant features in the vector increases complexity for the classifier, and thereby directly decreases performance in most cases.

The component-based AAM that we applied on faces permitted us to have a more exact description of facial details, hypothesis confirmed by the positive results of the expression recognition system. Our system is not developed with the explicit purpose of performing in real-time, but it takes into account some important real-time aspects, such as a reduced number of input features and a fast modeling algorithm.

An important drawback is the fact that there are no approved databases permitting to test the model in spontaneous, unconstrained, real-life situations for both still images and video sequences. If random consumer pictures are used, they have to be previously tagged as depicting specific facial expressions. For our experiments, we have developed our own collection of still images containing different individuals with variations in poses and facial expressions. As a result, the facial expressions depicted by the subjects are not spontaneous, but posed.

The realization of a complete database, containing still and video sequences, but also variations in pose and illumination, which also incorporates occlusions, should be a topic for future work. Another important future direction would be the classification of genuine expressions from posed ones and the recognition of combinations of expression.

Our model can be also incorporated into a facial recognition system robust to expression variations. Also, a separation of the expression parameters would permit an accurate expression transfer and synthesizer.

In conclusion our work demonstrates the practical realization of a facial expression recognizer which is suitable for further development and incorporation into next-generation consumer imaging appliances.

References

[1] T. F. Cootes, G. J. Edwards, and C. J. Taylor, “Active appearance models”, Lecture Notes in Computer Science, vol. 1407, pp. 484–, 1998.

[2] M.D. Cordea, E.M. Petriu, T.E. Whalen, “A 3D-anthropometric-muscle-based active appearance model”, in the IEEE Symposium on Virtual Environments, Human-Computer Interfaces and Measurement Systems, (VECIMS), pp. 88-93, 2004.

[3] H. Choi, S. Oh, “Real-time Recognition of Facial Expression using Active Appearance Model with Second Order Minimization and Neural Network”, International IEEE Conference on Systems, Man and Cybernetics, SMC 06, vol. 2, pp.1559 – 1564.

[4] Y. Saatci, C. Town, “Cascaded Classification of Gender and Facial Expression using Active Appearance Models”, in the 7th International Conference on Automatic Face and Gesture Recognition, FGR 2006.

[5] H. Li, H. Lin, G. Yang, “A New Facial Expression Analysis System Based on Warp Image”, in the Proceedings of the 6th World Congress on Intelligent Control and Automation, June 21 - 23, 2006, Dalian, China.

[6] H. Lee, D. Kim, “Facial Expression Transformations for Expression-Invariant Face Recognition”, Advances in Visual Computing, Springer, vol. 4291/2006.

[7] D.Datcu, L. Rothkrantz, “Facial Expression Recognition in still pictures and videos using Active Appearance Models. A comparison approach”, International Conference on Computer Systems and Technologies - CompSysTech’07.

[8] P. A. Viola, M. J. Jones, “Robust real-time face detection”, International Journal of Computer Vision, vol. 57, no. 2, pp. 137–154, 2004.

[9] F. Wallhoff, “The Facial Expressions and Emotions Database Homepage (FEEDTUM),” www.mmk.ei.tum.de/˜waf/fgnet/feedtum.html, Sept. 2005.

[10] M.J. Black and Y. Yacoob, “Recognizing facial expressions in image sequences using local parameterized models of image motion”, International Conference on Computer Vision, 1995, pp 23-48.

[11] M.C Ionita, I Bacivarov, P Corcoran, “Separating Directional Lighting Variability in Statistical Face Modeling Based on Texture Space Decomposition”, 15th International Conference on Digital Signal Processing, Cardiff, Wales, 2007.

[12] M. Yang, Y. Wu; J. Crenshaw, B. Augustine, R. Mareachen, “Face detection for automatic exposure control in handheld camera”, Computer Vision Systems, 2006 ICVS '06, pp.17 – 25.

[13] P. Corcoran, G. Costache, “Automated sorting of consumer image collections using face and peripheral region image classifiers”, in IEEE Transactions on Consumer Electronics, Vol.51, Issue 3, pp. 747- 754.

[14] Y.Tian, T. Kanade, J. F. Cohn, Facial Expression Analysis, Book Chapter in Handbook of face recognition, S.Z. Li & A.K. Jain, ed., Springer, October, 2003.

[15] Ekman, P. and W. Friesen, Facial Action Coding System: A Technique for the Measurement of Facial Movement. Consulting Psychologists Press, Palo Alto, 1978.

[16] I. Bacivarov, M. Ionita, P. Corcoran, Statistical Models of Appearance for Eye Tracking and Eye-Blink Detection and Measurement. IEEE Transactions on Consumer Electronics, August 2008.

[17] I. Bacivarov, M.C. Ionita, and P. Corcoran, A Combined Approach to Feature Extraction for Mouth Characterization and Tracking, in ISSC, Galway, Ireland, 2008.

[18] Zhang and F.S. Cohen Component-based Active Appearance Models for face Modeling, in International Conference of Advances in Biometrics, ICB, Hong Kong, China, January 5-7 2006.

[19] M. Pantic, and L.J.M. Rothkrantz, System for automatic analysis of facial expressions. Image and Vision Computing vol. 18, p. 881-905, 2000.

[20] M. Lyons, S. Akamasku, M. Kamachi, J. Gyoba, Coding facial expressions with gabor wavelets, in the Proceedings of International Conference on Face and Gesture Recognition, 1998.

[21] B. Fasel, and J. Luettin, Automatic facial expression analysis: A survey, in the journal of Pattern Recognition, 2003.

[22] L. Zalewski and S. Gong, 2D statistical models of facial expressions for realistic 3D avatar animation, in Computer Vision and Pattern Recognition, CVPR, 20-25 June 2005.

[23] D. Marshall, D. Cosker, P. L. Rosin, Y. Hicks, Speech and Expression Driven Animation of a Video-Realistic Appearance Based Hierarchical Facial Model, in Workshop in conjunction with IEEE CVPR of Learning, Representation and Context for Human Sensing in Video, June 22nd 2006, New York, USA.

[24] S. Xin, and H. Ai, Face Alignment under Various Poses and Expressions. Affective Computing and Intelligent Interaction, Springer, 2005.

[25] M. Pantic, M. F. Valstar, R. Rademaker, L. Maat, Web-based database for facial expression analysis, IEEE International Conference on Multimedia and Expo (ICME’05), http://www.mmifacedb.com, 2005.

[26] M. Nusseck, D. W. Cunningham, C. Wallraven, H. H. Bülthoff, The contribution of different facial regions to the recognition of conversational expressions, Journal of Vision, 8(8):1, pp. 1–23, 2008.

[27] P.J. Phillips, H. Wechsler, J. Huang, P. Rauss, The FERET database and evaluation procedure for face recognition algorithms, Image and Vision Computing J, Vol. 16, No. 5, pp. 295-306, 1998.

[28] P.J. Phillips, H. Moon, S.A. Rizvi, P.J. Rauss, The FERET Evaluation Methodology for Face Recognition Algorithms, IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 22, pp. 1090-1104, 2000.
[29] http://www.embedded-computing.com/news/db/?10084

Ioana Bacivarov received the M. Eng. Degree in Signal Processing from the National Polytechnic Institute of Grenoble, France and from the University "Politehnica" of Bucharest, Romania in a double diploma agreement in 2006, 2007 respectively. She is currently pursuing a Ph.D. degree in Image Processing & Computer Vision at NUI, Galway. Her research interests include signal processing and pattern recognition with applications in face recognition.

Peter Corcoran received the BAI (Electronic Engineering) and BA (Math's) degrees from Trinity College Dublin in 1984. He continued his studies at TCD and was awarded a Ph.D. for research work in the theory of Dielectric Liquids. In 1986 he was appointed to a lectureship in Electronic Engineering at NUI, Galway. His research interests include embedded systems, home networking, digital imaging and wireless networking technologies.

Mircea C. Ionita received the B.Sc. degree in Computer Science & Electrical Engineering from University "Politehnica" of Bucharest, Romania, in 2003. He received the M.Sc. degree in Images, Patterns & Artificial Intelligence from the same university in 2004. He is currently pursuing a Ph.D. degree in Image Processing & Computer Vision at National University of Galway, Ireland. His research interests include color image processing, statistical face modeling, and face recognition and tracking.