*Journal of Information Hiding and Multimedia Signal Processing* ©2010 ISSN 2073-4212
**Ubiquitous International** Volume **1**, Number **3**, July **2010**

# Statistical Analysis of Human Facial Expressions

Stelios Krinidis

Department of Informatics
Aristotle University of Thessaloniki
Box 451, 54124 Thessaloniki, Greece
stelios.krinidis@mycosmos.gr

Ioannis Pitas

Department of Informatics
Aristotle University of Thessaloniki
Box 451, 54124 Thessaloniki, Greece
pitas@aiia.csd.auth.gr
Informatics and Telematics Institute
CERTH, Greece

Received March 2010; revised June 2010

Abstract. *This paper presents a method for generalizing human facial expressions by means of a statistical analysis of human facial expressions coming from various persons. The data used for the statistical analysis are obtained by tracking a generic facial wireframe model in video sequences depicting the formation of the different human facial expressions, starting from a neutral state. Wireframe node tracking is performed by a pyramidal variant of the well-known Kanade-Lucas-Tomasi (KLT) tracker. The loss of tracked features is handled through a model deformation procedure that increases the robustness of the tracking algorithm. Tracking initialization is performed in a semi-automatic fashion, i.e., the facial wireframe model is fitted to an image representing a neutral facial expression, exploiting physics-based deformable shape modeling. The dynamic facial expression output model is MPEG-4 compliant. The method has been tested on a variety of sequences with very good results, including a database of video sequences representing human faces changing from the neutral state to the one that represents a fully formed human facial expression.*
**Keywords:** Facial expressions, facial expression analysis, tracking, deformable model, MPEG-4, statistical analysis.

1. **Introduction.** Facial analysis and synthesis have become two very important goals for human-centered interface applications [1, 2, 3]. Facial analysis refers to the extraction of information concerning head location, pose and facial feature movement, notably movement of the eyes and mouth from video sequences [4, 5]. Facial synthesis refers to the reverse process of animating a facial model using a set of high-level parameters that control facial pose, expression and gaze [6, 7]. Facial analysis would be useful for several applications, such as eye-tracking, facial expression recognition [8, 9, 10, 11] and visual speech understanding [12], whereas facial synthesis would be useful for animating virtual characters or digital actors [7, 13, 14, 15]. Together, facial analysis and synthesis *in tandem* would be useful for model-based coding applications, such as video email and video-teleconferencing, as well as for the representative human facial expressions [16, 17, 18, 19].

The human process of producing expressions is very complex. It involves subtle movements of facial features such as eyes, eyebrows, cheeks, the mouth, etc, driven by contraction and relaxation of muscles. Moreover, the human ability of perceiving facial expressions is so accurate that even a trivial displacement of a facial feature can be detected immediately. Therefore, it is very difficult to fully automatically analyze facial expressions.

The human facial expression analysis is based on the motion extraction, which can be classified in three categories: *dense optical flow*, *difference-images* and *feature point tracking*. Dense optical flow has been applied both locally and holistically:

- **Holistic dense optical flow** approaches allow for whole-face analysis. Lien [20] analyzed holistic face motion with the aid of wavelet-based, multi-resolution dense optical flow.
- **Local dense optical flow:** Region-based dense optical flow was used by Mase and Pentland [21] in order to estimate the activity of 12 of the totally 44 facial muscles. Otsuka and Ohya [22] estimated facial motion in local regions surrounding the eyes and the mouth. Feature vectors were obtained by taking 2D Fourier transforms of the vertical and horizontal optical flow fields. Yoneyama *et al.* [23] divided normalized test faces into $8 \times 10$ regions, where local dense optical flow was computed and quantified region-wise into ternary feature vectors (+1/0/-1), indicating upwards, none and downwards movements, while neglecting horizontal facial movements.

Apart from a certain vulnerability to image noise and non-uniform lighting, holistic dense optical flow methods often result in prodigious computational requirements and tend to be sensitive to motion discontinuities (iconic changes) as well as non-rigid motion. Optical flow analysis can also be done in conjunction with motion models that allow for increased stability and better interpretation of extracted facial motion, e.g. muscle activations:

- **Holistic motion models:** Terzopoulos and Waters [24] have been used eleven principal deformable contours (also known as "snakes") to track lip and facial features throughout image sequences with the aid of a force field, which is computed from gradients found in the images. Only frontal faces were allowed and some facial make-up was used to enhance contrast. DeCarlo and Metaxas [25] presented a formal methodology for the integration of optical flow and 3D deformable models and applied it to human face shapes and facial motion estimation.
- **Local motion models:** Black and Yacoob [26] as well as Yacoob and Davis [27] introduced *local parametric motion models* that allow, within local regions in space and time, to not only accurately model non-rigid facial motions, but to provide also a concise description of the motion associated with the edges of the mouth, nose, eyelids and eyebrows in terms of a small number of parameters.

In addition to low-level dense optical flow, there are also higher level variants that focus either on the movements of generic feature points, patterns or markers:

- **Feature point tracking:** Here, motion estimation is obtained only for a selected set of prominent facial features [28, 29]. In order to reduce the risk of tracking loss, feature points are places into areas of high contrast, preferably around intransient facial features. Hence, facial movements can be measured by tracking the displacement of corresponding feature points.
- **Marker tracking:** It is possible to determine facial actions with more reliability than with previously discussed methods, namely by measuring deformation in areas, where underlying muscles interact. Unfortunately, these are mostly skin region with relatively poor texture. Highlighting is necessary and can be done by either applying

color to salient facial features and skin [30] or by affixing colored plastic dots to predefined locations on the subject's face.

Yet another way of how to extract image motion are **difference-images:** Specifically for facial expression analysis, difference-images are mostly created by subtracting a given facial image from a previously registered reference image, containing a neutral face of the same subject. However, in comparison to optical flow approaches, no flow direction can be extracted, but only differences of image intensities. In addition, accurate face normalization procedures are necessary in order to align reference faces onto the test faces. Holistic difference-image based motion extraction was employed in [31, 32].

Furthermore, another way to analyze the facial expressions is to use the well-known principal component analysis (PCA) [33] or kernel discriminant analysis (KDA) [34]. Pan *et al.* [34] exploit a variation of the KDA in order to analyze the facial expression, while Li *et al.* [35] use a variation of the locality preserving projection (LPP) technique to extract and analyze the facial features.

Our approach was motivated by the lack of a facial expression analysis system that is able to perform and exploit statistical analysis of the dynamic human facial expressions, i.e., of their formation from the neutral state to the fully expressive one. Statistical analysis of the facial expressions offers the opportunity to personalize expressions on an individual, with respect to nationality and social class, instead of just transferring a person's expression to another person. Furthermore, in our approach, the aforementioned process is a dynamic one and takes into account the entire video sequence of each facial expression from the neutral state to the fully expressive one. It is not restricted to a single video frame of the expression of interest. All video frames are used and combined to achieve a better result. The statistical analysis is based on the use of displacement vectors. The idea of using displacement vectors was inspired by the Facial Animation Parameters (FAPs) of the MPEG-4 standard. Finally, the portability of the results is achieved through their casting in an MPEG-4 format. Hence, they can be used in any MPEG-4 application.

The introduced algorithm, initially fits and subsequently tracks a facial wireframe model [36] in video sequences containing the formation of a dynamic human facial expression from the neutral state to the fully expressive one [37]. The facial wireframe model fitting on the face depicted on the initial frame of the video sequences is performed using a semi-automatic algorithm opposed to other fully manual algorithms lie in the literature. This method needs only 5-8 manually selected correspondences between model and the facial features depicted on the image. These correspondences are used in combination with a deformable model to fit the rest of the wireframe model on the image. Following, the facial features are tracked in the video sequence using a variant of KLT tracker [38]. If needed, model deformations are performed by mesh fitting at the intermediate steps of the tracking algorithm. Such deformations provide robustness and tracking accuracy contrary to the methods exploiting only the KLT tracker [38]. The extracted dynamic mesh deformation data can be used to calculate representative (or average) dynamic facial expressions for groups of people (e.g. of the same nation and social class), who express themselves in a similar way.

The MPEG-4 standard [39] - [43] is exploited in order to describe the results of our method. It specifies a way of modeling facial expressions, which is strongly influenced by neurophysiological and psychological studies [44, 45]. It employs the Facial Animation Parameters (FAPs) operating on a set of FDP (Facial Definition Parameter) facial feature points. The FDPs define the three dimensional $[x, y, z]$ location of 84 points on a neutral face. In our work, we shall use only the $x$ and $y$ dimensions since our video data are

inherently 2D. However, the full 3D coordinate system can be used if we possess 3D data (e.g. through a range camera). FDPs usually correspond to facial features and, therefore, can roughly outline the shape of the face. The FAPs specify FDP displacements, which model actual facial feature movements that a realistic human face would make, covering many natural facial expressions, as well as exaggerated expressions to some extent (e.g. for cartoon characters). All FAPs involving translational movement are expressed in terms of Facial Animation Parameters Units (FAPUs). Thus, they correspond to fractions of distances between some essential facial features (e.g. eye distance).

Before proceeding to the analysis of our method, we should provide some necessary definitions. More specifically, a *dynamic facial expression model* is the set of frame facial expression models (grids) that describe the formation of a facial expression through time. Accordingly, the term *frame facial expression model* corresponds to a grid that is single instance of the dynamic model on a particular video frame.

The main novelty of this paper is the statistical analysis performed on the dynamic facial expression models coming from the video sequences of each of the six human facial expressions that depict each expression from the neutral state to the fully expressive one. Facial grid displacement vectors, inspired by the MPEG-4 standard, were used to perform the statistical analysis, thus avoiding the direct registration of the data (human facial wireframe models). Another novelty of this paper is the presentation and implementation of a MPEG-4 compliant system that is able to perform statistical analysis (facial grid location and dispersion estimation) of the dynamic human facial expressions. Furthermore, the extraction of the *generalized median dynamic facial expression models* for each of the six human facial expressions could form and show the differences on facial expressions for different racial groups. In other words, the generalized median dynamic facial expression models not only could help us to generalized the expressions of a racial group, but it could also point-out the expression differences among different racial groups. Also, this system could be used to develop avatars not only for teleconferencing, but also for more advanced systems, such as virtual environments.

The remainder of the paper is organized as follows. The method of fitting the wireframe facial model to a facial image is presented in Section 2. Section 3 introduces the tracking algorithm [38] exploiting a physics-based deformation method to compensate for lost features. The statistical facial expression analysis of the data is introduced in Section 4. Finally, experimental results are illustrated in Section 5, while conclusions are drawn in Section 6.

2. **Wireframe-Based Model Fitting.** In this Section, our goal is focused on fitting a facial wireframe model to a face image in a video frame. It is performed in a semi-automatic way for attaining speed, reliability and robustness of the fitting procedure. The facial wireframe model that is used throughout this paper, is the well-known Candide wireframe model [36, 46]. Candide is a parameterized face mask specifically developed for model-based coding of human faces. A frontal and a profile view of the model can be seen in Figure 1. The Candide model is superset of the MPEG-4 facial model pattern illustrated in Figure 2. The MPEG-4 facial features that are not comprised in Candide model (e.g. ear), are ignored, since they are of no particular significance for facial expression recognition.

The fitting procedure consists of the following steps. First, the facial model is randomly initialized on the face image. The model is assumed to be in its neutral state. As soon as the model is initialized, a number of point correspondences are manually selected, i.e., model nodes are manually matched against facial features in the actual face image. The model nodes of greater significance are chosen to be matched. It has been empirically

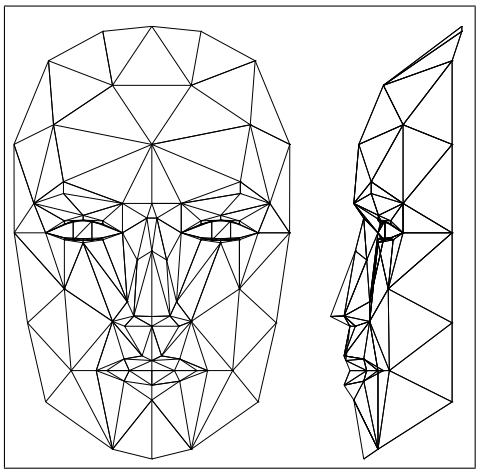

FIGURE 1. *Frontal and profile views of the Candide wireframe model.*

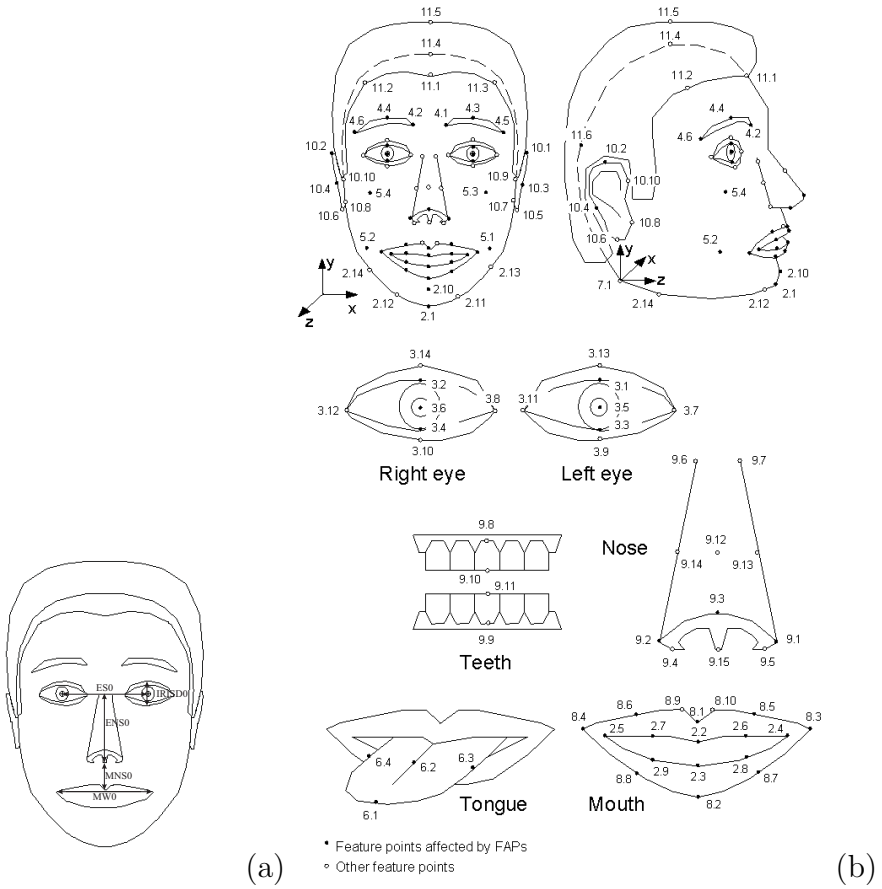


FIGURE 2. ***(a)*** *The MPEG-4 facial animation parameter units (FAPUs), and* ***(b)*** *the FDPs diagram.*

determined that 5-8 correspondences are enough for a good model fitting. These correspondences are used as the driving power which deforms the rest of the model and matches its nodes against face image points. The facial model is assumed to be a deformable 2D mesh model. The driving forces, needed to deform the model, are determined based on the point correspondences between the facial model nodes and the face image features. Each force is defined to be proportional to the Euclidean distance between the model nodes and their corresponding matched feature points on the face image.

Subsequently, the deformation algorithm deforms the facial model and translates the model nodes close to their true position on the face image. As previously mentioned, only a small number (about 5-8 pairs) of point correspondences, is enough to fit the model fairly well to a face image. If the scale difference between the used facial model and the face represented in the image is large, then the number of the required pairs of point correspondences increases to compensate for scale changes. For example, it has been experimentally found that if the model size is 1.5 times larger or smaller than that of the face image, 15 pair point correspondences are needed in order to produce good deformations. This problem can be solved by a preprocessing scaling procedure that scales the model to the size of the face image.

An example of the facial model fitting procedure on a face image is illustrated in Figure 3. Model fitting was performed using only 7 correspondences between model nodes and face image features. Figure 3a depicts the correspondences between the model nodes and image features, while Figure 3b illustrates the model after deformation. As can be clearly seen in Figure 3, the model nodes were fitted to image features with high accuracy, using few initial correspondences and exploiting the deformation process to accurately match the remaining model nodes to their corresponding image features. In case of any outlying nodes after deformations, their position can be manually changed to achieve the best possible model fitting.


FIGURE 3. *An example of mesh model fitting on a face image exploiting deformations.* ***(a)*** *The randomly initialized model on the face image, and the 7 manually defined correspondences between model nodes and face image features.* ***(b)*** *The model position after deformation.*

3. **Model Based Tracking.** In this Section, an algorithm is presented that is used for tracking the facial feature points of the wireframe model of interest in a video sequence. The algorithm is based on tracking a large number of previously selected feature points in the facial region. Although, feature points can be selected automatically [47, 48], in our case, the tracked feature points are the output of the previous process (Section 2), i.e., they are the nodes of the fitted face model. The test video sequences depict an initially neutral human face, which gradually deforms to produce a particular facial expression. The result of the tracking algorithm is the position of the facial model nodes at intermediate video frames.

Feature points are tracked using a pyramidal implementation of the well-known Kanade-Lucas-Tomasi (KLT) algorithm [38, 47, 48]. A modification of this algorithm that uses a pyramidal representation of the images of interest, is adopted in this paper [38].

As soon as the tracking algorithm computes the displacement of all the tracked features (i.e., the model nodes), the resulting configuration (containing the new positions of the model nodes) is deformed. All model nodes are feature points. The displacements of model

nodes, that have not been lost (after tracking), are assumed to be the driving forces of the model deformation, thereby providing an accurate and robust model based facial feature tracking method. This solves a major problem of feature-based tracking algorithms, the gradual elimination of features points with respect to time. In the modified tracking algorithm, the incorporation of the deformation step enables the tracking of features that would have been lost otherwise. Furthermore, feature displacements are enforced to have a uniform distribution considering the features with extreme displacement as lost and their displacement is handled by the deformation procedure. This part of the modified tracking algorithm is used due to that fact that a lot of features are located on plain skin. The tracking algorithm provides a *dynamic facial expression model* for each video sequence, which consists of a series of frame facial expression models, one for each video frame.

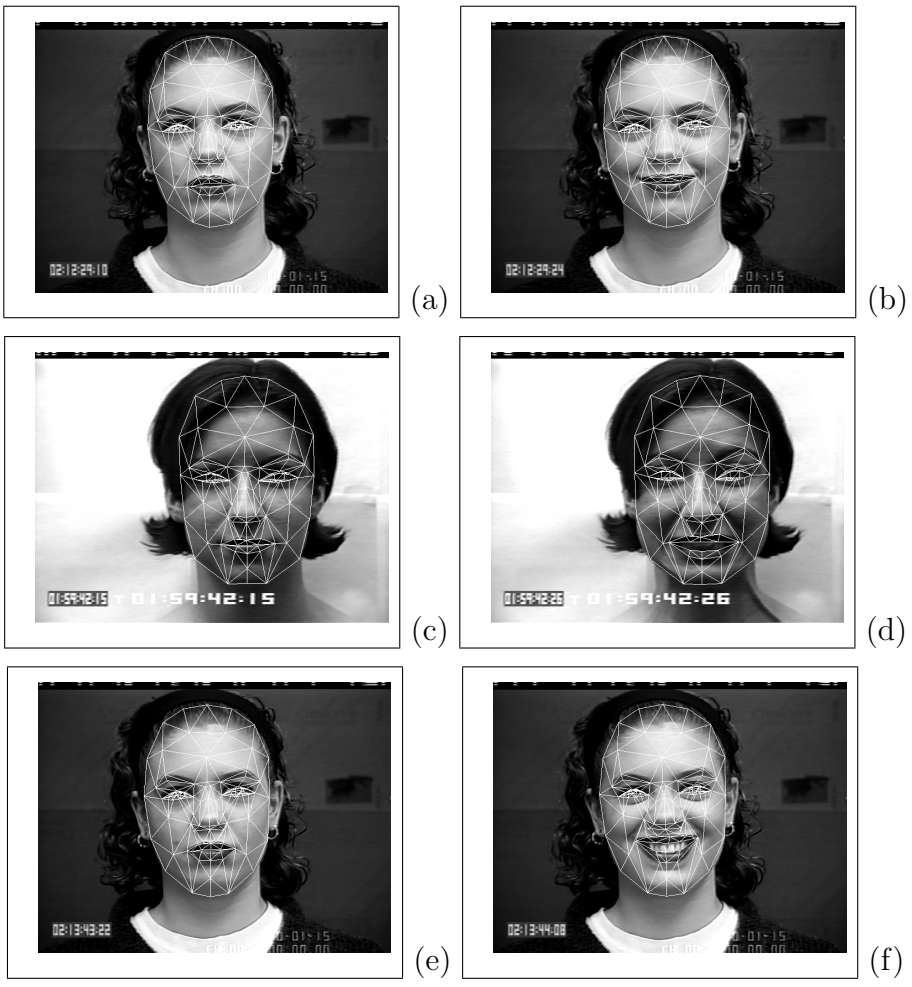


FIGURE 4. *Model-based facial feature tracking on various video sequences, corresponding to different facial expressions (one expression per row). The left column depicts the first frame of the video sequences, while the right column corresponds to the last frame.*

The proposed tracking algorithm has been applied to various video sequences, as illustrated in Figures 4 and 5. The test video sequences comprise of more than 10 frames and demonstrate a gradual change from a neutral state to a particular fully formed facial expression (i.e., laughter, anger, etc.). Figures 4a, 4c and 4e depict the model fitted to the neutral face image (i.e., the first frame of the video sequence), while Figures 4b, 4d and

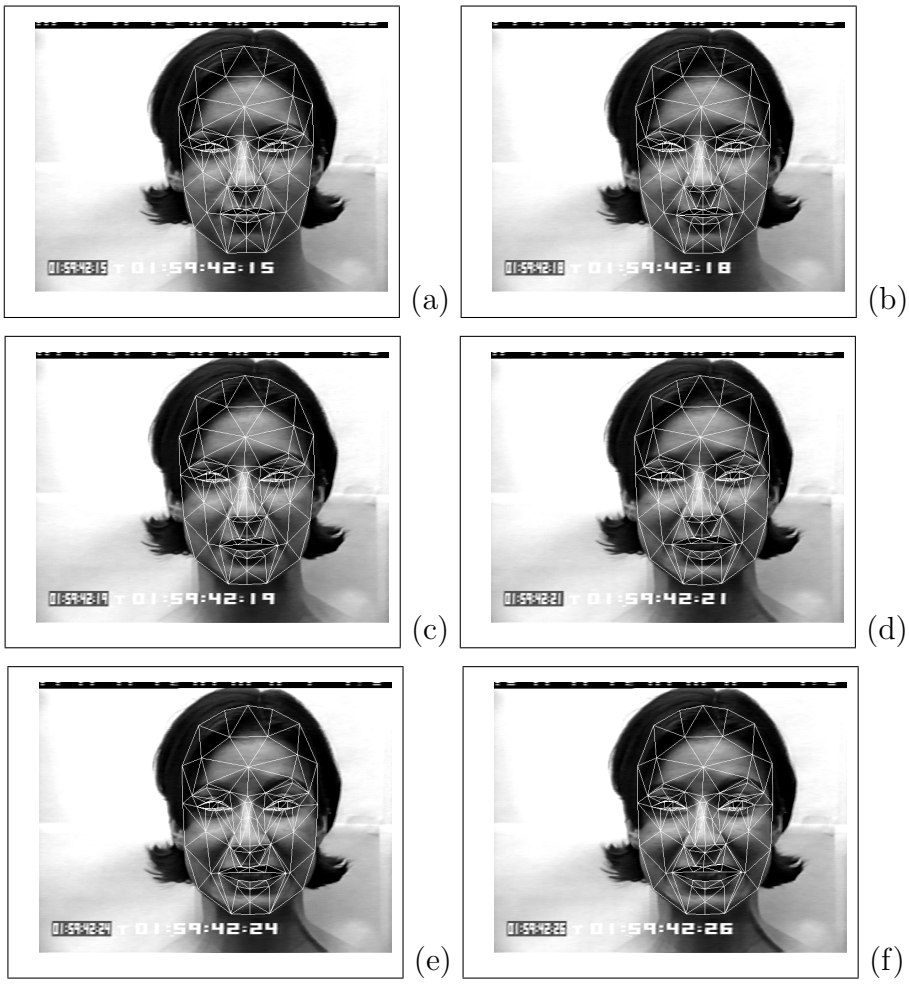


FIGURE 5. *Model-based tracking in six uniformly sampled frames of a video sequence showing the full formation of a facial expression.*

4f illustrate the result of model tracking at a fully formed facial expression (last frame of the video sequence). Moreover, Figure 5 illustrates the results of model tracking at six uniformly sampled frames of a video sequence that depicts a face from the neutral state to a fully formed facial expression.

Furthermore, we have performed an error analysis of the tracking method used by our algorithm. For this experiment we used video sequences containing 24 frames. The feature points (Candide model nodes) of interest were manually tracked for a number of video sequences. The error used in this experiment is defined as:

$$\epsilon = \frac{1}{N} \sum_{i=1}^{N} \sqrt{(x_i^m - x_i)^2 + (y_i^m - y_i)^2}, \tag{1}$$

where $N$ is the number of tracked feature points, $\mathbf{x}^m = [x^m, y^m]^T$ is the location of a manually tracked feature point (ground truth), while $\mathbf{x} = [x, y]^T$ is the location of the correspondent automatically tracked feature point. Figure 6 illustrates the errors performed by the tracking method used in our algorithm as well as the error occurred by the same tracking method without deforming the tracked feature points after each tracking iteration. It is clearly depicted that both errors are very low, proving the robustness of the tracking algorithm used. Furthermore, as it is obvious in Figure 6, the tracking method exploiting deformations after each iteration produces lower tracking errors. Furthermore,

the tracking without model deformation tends to produce error accumulation over time that can possibly cause a collapse of the entire tracking procedure.

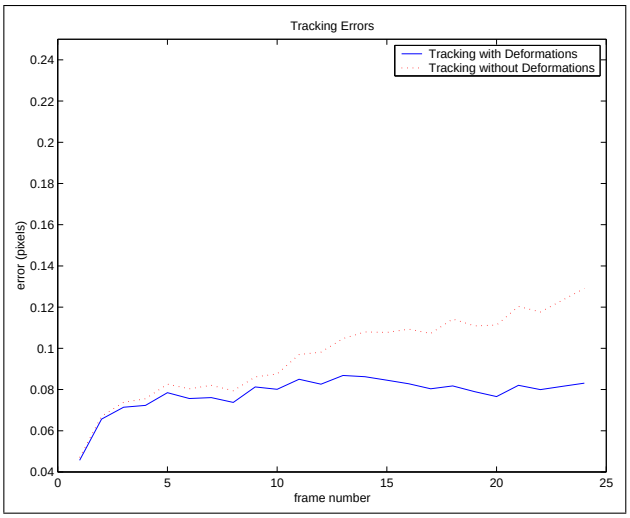


FIGURE 6. *Node position errors (in pixels) produced by the tracking procedure, with (solid line) and without (dotted line) deformations, in video sequences comprising of* 24 *frames.*

4. **Statistical Analysis of Facial Expression Grids.** In this Section, the statistical analysis of facial expression grids is exploited in order to derive a representative dynamic facial expression model for a set of different dynamic facial expression models corresponding to the formation of a specific facial expression on different persons. The dynamic models of interest are firstly normalized and subsequently subjected to an outlier rejection process. Finally, the representative dynamic facial expression models are extracted and their dispersion is studied. Both the outlier rejection process and the extraction of the representative dynamic model are performed in a variety of ways in order to determine the most accurate one, which achieves the best dynamic facial expression model representation.

4.1. **Dynamic Facial Expression Model Normalization.** A three-step normalization procedure is performed on the set of acquired dynamic facial expression models (grids), so that all dynamic models consist of the same number of frame models and have the same orientation position and size. The facial expression video sequences, and, as a result, the extracted dynamic facial expression models, do not necessarily have the same number of frames. It has been empirically determined that five frames are sufficient to describe an expression without loss of accuracy. In other words, five frames and their corresponding frame facial expression models accurately describe the gradual changes required to form the full facial expression, starting from the neutral state. Thus, each dynamic facial expression model is constrained to contain only five frame facial expression models, one per frame. If needed, more frame models could be acquired by interpolating the existing ones. This assumption was proven to be a valid one by interpolating the normalized dynamic models from the chosen five frame models, so as to contain the same number of frames as the original ones. Table 1 shows the interpolation error statistics of grid node position (in pixels) for different numbers of missing frames (first column of Table 1) between existing frames so that they get the same number of frames as the original ones. For example, if the original video sequence contains 17 frames, and we want to interpolate

the normalized dynamic facial expression model (consisting of five frames), so as to have the same number of frames as the original one, 3 frames are necessary to be interpolated for each pair of frames of the normalized dynamic model, whilst if the original video contains 25 frames, 5 frames should be interpolated. As can be seen, median and mean errors are less than 0.2 pixels, while maximum errors are slightly larger than 1.5 pixel, proving the validity of the aforementioned assumption.

TABLE 1. ***Interpolation error statistics*** *for facial frame models versus skipped frame number from the normalized dynamic models.*

| interp. frames | median | maximum | mean ± s. dev. |
|---|---|---|---|
| 0 | 0.0000 | 0.0000 | 0.0000 ± 0.0000 |
| 1 | 0.1049 | 0.7961 | 0.1397 ± 0.0334 |
| 2 | 0.0676 | 1.7784 | 0.1270 ± 0.1164 |
| 3 | 0.0625 | 1.5565 | 0.0910 ± 0.0850 |
| 4 | 0.1196 | 1.8802 | 0.1574 ± 0.1313 |
| 5 | 0.0835 | 1.8965 | 0.1194 ± 0.1400 |

The second step of the normalization procedure is to normalize the frame facial expression models with respect to orientation. This normalization is necessary, among others, for FAP file extraction. We assume that our videos have only frontal pose. More specifically, model rotation around the $y$ axis is performed in frame facial expression models normalization, so that all frame models are vertical [4]. The vertical bisector of the triangle formed by the pupils of eyes and the middle point of the horizontal mouth axis is used (Figure 7) to determine the rotation needed to bring the frame model to a vertical position. The mass center of the model is supposed to be the center of the rotation. Furthermore, the proposed procedure is translation invariant, since, as will be described below, it calculates and uses displacement vectors among the frame model nodes. Moreover, the frame model of the dynamic facial expression models are inherently registered by their acquisition way.

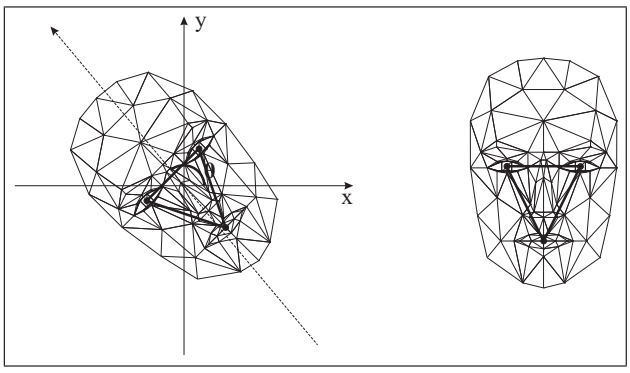


FIGURE 7. *Compensation of the orientation of the wireframe facial model. It is based on the triangular formed by the eyes and mouth centers.*

The last step of the normalization procedure is to normalize the frame facial expression models with respect to size, so that all models have the same FAPUs (Figure 2a). Hence, they are scaled in such a way that their facial animation parameter units (FAPUs) are equal. The chosen scaling factor along the $x$ axis is equal to the average of the scaling factors produced by ES0 and MW0 line segments shown in Figure 2a. We enforce them to be equal in all the frame models of all the dynamic facial expression models under

examination. We use the same procedure for scale normalization along the $y$ axis. The line segments IRISD0, ENS0 and MNS0 (Figure 2a) are used to this end.

4.2. **Definition of Facial Expression Model Distances.** As soon as all dynamic facial expression models are normalized, their statistical analysis can be performed. The aim of this study is to find a representative dynamic facial expression model for a set of dynamic facial expression models corresponding to different persons posing a specific facial expression. A direct model analysis would require geometric model registration, which may suffer from certain problems. For example, it is difficult to distinguish between scale changes due to camera position or due to the actual size of the head of a person. Instead, an indirect method is applied, which is based on the fact that dynamic facial expression is an "incremental" deformation of the face from the neutral state to the fully expressive one. Therefore, the displacements of the facial features from one video frame to the next one carry sufficient information to characterize a dynamic facial expression model. More specifically, model registration is by-passed by using the model node *displacement vectors*:

$$\boldsymbol{\tau}^i_m(k) = \overrightarrow{M^i_m(k)M^i_{m+1}(k)}, \tag{2}$$

where $M^i$ is the dynamic facial expression model under examination, $M^i_m$ is the $m$-th frame model of the dynamic model $M^i$ and $M^i_m(k)$ denotes the $k$-th node of the model $M^i_m$. The idea of using the displacement vectors is inspired by the MPEG-4 standard [39, 40, 42, 43] and FAP definition there in.

Let us, now, define the distance $d_v$ between two displacement vectors $\boldsymbol{\tau}^i_m(k)$ and $\boldsymbol{\tau}^j_m(k)$:

$$d_v(\boldsymbol{\tau}^i_m(k), \boldsymbol{\tau}^j_m(k)) \triangleq \|\boldsymbol{\tau}^i_m(k) - \boldsymbol{\tau}^j_m(k)\|, \tag{3}$$

where $\|\cdot\|$ denotes either the Euclidean distance or relevant distance metrics, such as the $L_1$ norm (*Manhattan distance*) or the *Chessboard distance* [49]. Furthermore, the distance $d_f$ between two frame facial expression models $M^i_m$ and $M^j_m$ is defined as:

$$d_f(M^i_m, M^j_m) \triangleq \frac{1}{N}\sum_{k=1}^{N} d_v(\boldsymbol{\tau}^i_m(k), \boldsymbol{\tau}^j_m(k)), \tag{4}$$

where $N$ is the number of facial model nodes. On the same basis, the distance $d$ between two dynamic facial expression models $M^i$ and $M^j$ is defined as:

$$d(M^i, M^j) = \frac{1}{N'-1}\sum_{m=1}^{N'-1} d_f(M^i_m, M^j_m), \tag{5}$$

where $N'-1$ is the number of the displacement vectors sets corresponding to the $N'$ frame facial expression models of each normalized dynamic model. It can be easily verified that the above defined distances (3), (4) and (5) satisfy the three fundamental distance properties, i.e., positivity, the permutational property and the triangle inequality.

4.3. **Outlying Facial Expression Model Rejection.** Before proceeding to the estimation of the location and dispersion of the facial expression models, an outlier rejection process has to be employed to remove outlying facial expression models. Let us assume that we have $N''$ dynamic facial expression models, one for each facial expression sequence. Facial model distance definitions can be used for outlying facial expression model rejection, before estimating their location and dispersion. The chosen outlier trimming process is performed as follows [50]. A distance $d_t(M^i)$ of one dynamic facial expression model $M^i$ from the rest of the models that belong in the same class is defined. Models of each class are sorted according to $d_t(M^i)$ and $a\%$ of them that possess the largest of

$d_t(M^i)$ are trimmed away, thereby freeing the model data from possible outliers. The trimming process can be applied in three different ways:

- Trimming at the dynamic facial expression model level, i.e., $\alpha\%$ of the dynamic facial expression models that are located too far away from the rest are removed. The distance between a dynamic facial model $M^i$ and the rest models that should be maximized is defined as:

$$d_t(M^i) = \frac{1}{N'' - 1} \sum_{\substack{j=1 \\ j \neq i}}^{N''} d(M^i, M^j), \tag{6}$$

  where $N''$ is the number of the available dynamic models.
- Trimming at the frame facial expression model level, i.e., $\alpha\%$ of the models that are located too far away from the rest are removed from the sets that are formed by the frame facial expression models which correspond to the same video frame. The distance between a frame facial expression model $M_m^i$ and the rest frame models corresponding to the video frame $m$, is defined as:

$$d_{tf}(M_m^i) = \frac{1}{N'' - 1} \sum_{\substack{j=1 \\ j \neq i}}^{N''} d_f(M_m^i, M_m^j). \tag{7}$$

- Trimming at the displacement vector level, i.e., $\alpha\%$ of the displacement vectors of each frame model node of the dynamic facial expression models that are located too far away from the rest are removed. The distance between a displacement vector $\boldsymbol{\tau}_m^i(k)$ and the rest of the displacement vectors of the corresponding nodes at location $k$ in the same frame $m$ is defined as:

$$d_{tv}(\boldsymbol{\tau}_m^i(k)) = \frac{1}{N'' - 1} \sum_{\substack{j=1 \\ j \neq i}}^{N''} d_v(\boldsymbol{\tau}_m^i(k), \boldsymbol{\tau}_m^j(k)). \tag{8}$$

4.4. **Estimation of Location and Dispersion of Dynamic Facial Expression Models.** Once the facial expression model data have been freed from outliers, we can proceed to the estimation of the location and dispersion of the facial expression model data. Location estimation is essentially the estimation of a representative dynamic facial expression model [51, 52, 53, 54] out of a set of dynamic facial expression models, e.g. of the six basic facial human expressions as expressed by various humans. The same procedure can be applied to find the representative dynamic facial expression of one particular expression (e.g. smile) for one person in case we have multiple video sequences of the same person while smiling. This can be done by finding the *generalized median dynamic facial expression model* [54, 55], which is defined as the dynamic model $M_{med} \in \mathbf{U}$ that minimizes the sum of distances to all dynamic models belonging to $\mathbf{M}$, i.e.,

$$M_{med} = \arg \min_{M \in \mathbf{U}} \sum_{M^i \in \mathbf{M}} d(M, M^i), \tag{9}$$

where $\mathbf{U}$ is the domain of all possible dynamic facial expression models, while $\mathbf{M}$ is the set of all dynamic facial expression models of a particular class (e.g. dynamic smiles of many persons, different smile sequences of the same person). The generalized median of graphs has been investigated and found useful in the field of statistical graph analysis [55]. The definition $d(M, M^i)$ depends on the distance metric used. It can be found using a greedy

search algorithm [54]. If the search is constrained to the given set $\mathbf{M}$, i.e., $\mathbf{U} = \mathbf{M}$, the resulting dynamic model:

$$\hat{M} = \arg \min_{M \in \mathbf{M}} \sum_{M^i \in \mathbf{M}} d(M, M^i), \tag{10}$$

is called *set dynamic median facial expression model* of $\mathbf{M}$.

Dynamic model dispersion can be measured by a facial expression model sample "variance" estimator defined appropriately in order to fit the dynamic facial expression model data:

$$\hat{\sigma}^2 \triangleq \frac{1}{N'' - 1} \sum_{i=1}^{N''} \left[ d(M^i, \bar{M}) \right]^2, \tag{11}$$

where $N''$ is the number of dynamic models under examination and $\bar{M}$ is the dynamic facial expression model for which each displacement vector $k$ of each frame model $m$ is defined as:

$$\boldsymbol{\tau}_m(k) = \text{arithmetic\_mean}\{\boldsymbol{\tau}_m^1(k), \boldsymbol{\tau}_m^2(k), \ldots, \boldsymbol{\tau}_m^{N''}(k)\}. \tag{12}$$

The median $med_M$ of the distances of each dynamic facial expression model from the rest $d_t(\cdot)$ could be another dispersion metric of the dynamic facial expression models belonging to the same class:

$$med_M \triangleq \text{median} \left\{ d_t(M^i) \right\}. \tag{13}$$

All the distances $d_t(M^i)$ are assumed to be ordered. As a third dispersion metric, we can use a variant of the Median of the Absolute Deviations (MAD) estimator [56] modified to fit the dynamic facial expression models:

$$\hat{\sigma}_{MAD} \triangleq \text{med} \left\{ d(M^1, M_{med}), \ldots, d(M^{N''}, M_{med}) \right\}, \tag{14}$$

where $M_{med}$ is the generalized median dynamic facial expression model [50, 56, 57, 58, 59]. The above mentioned dispersion estimation metrics indicate the dispersion, i.e. the dissimilarity of the various facial expression, e.g. among different persons. A flow chart of the overall algorithm is illustrated in Figure 8. At the end of the entire procedure, MPEG-4 FAP files describing the representative expressions and their progressive formation with respect to time, are extracted.

5. **Experimental Results.** To evaluate our method, we use the well-known Cohn-Kanade expression database [37] as input in our experiments. The database contains image sequences of over 200 subjects in the age range of 18 to 50 years. 69% of them are females, while 31% are males. 81% of the database subjects are Euro-Americans, 13% are Afro-Americans, and 6% belong to other racial groups. The motivation for the selection of this database originates from its content, i.e., the fact that image sequences depict the formation of human facial expressions from the neutral state to the fully expressive one.

The first set of experiments shows the efficiency of the proposed algorithm with respect to the extraction of representative dynamic facial expression models for each of the six basic human facial expressions. Figure 9 illustrates the last frame model (fully expressed) of the generalized dynamic facial expression median models for three expressions (anger, smile and sadness). The generalized median models shown in this Figure are extracted exploiting the Euclidian distance ($L_2$ norm) and all three trimming alternatives described in Section 4, namely, trimming at the dynamic model level (first column of Figure 9), trimming at the frame model level (second column of Figure 9) and trimming at the displacement vector level (third column of Figure 9). It can be noticed that there is no large difference between the generalized median models obtained by the three trimming

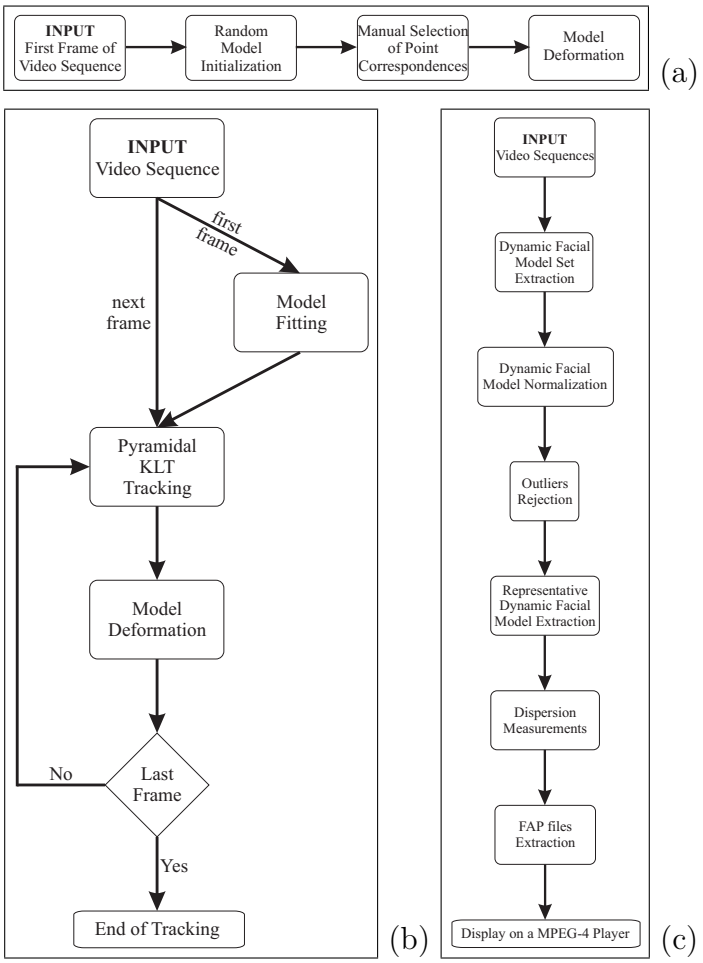


FIGURE 8. ***Flow diagrams***. ***(a)*** *Model fitting algorithm on a face image,* ***(b)*** *Facial feature tracking algorithm, and* ***(c)*** *Overall algorithm.*

TABLE 2. ***Dispersion*** $\hat{\sigma}^2$ *of the dynamic facial expression models for the six basic human facial expressions. The trimming cases are:* ***(***$1^{st}$ ***case)*** *without trimming,* ***(***$2^{nd}$ ***case)*** *trimming at the dynamic model level,* ***(***$3^{rd}$ ***case)*** *trimming at the frame model level, and* ***(***$4^{th}$ ***case)*** *trimming at the displacement vector level.*

| | **Anger** | **Disgust** | **Fear** | **Laughter** | **Sadness** | **Surprise** |
|---|---|---|---|---|---|---|
| $1^{st}$ case | 9,91 | 9,06 | 10,71 | 9,08 | 10,25 | 12,47 |
| $2^{nd}$ case | 9,56 | 8,86 | 10,34 | 8,74 | 10,01 | 11,94 |
| $3^{rd}$ case | 9,25 | 8,78 | 10,09 | 8,61 | 10,00 | 11,57 |
| $4^{th}$ case | **8,79** | **7,80** | **9,40** | **7,65** | **8,97** | **10,36** |

approaches. All the experiments described here are assumed to exploit the Euclidean distance. Furthermore, Tables 2, 3 and 4 show the dispersion of the dynamic models of the six basic human facial expressions for each trimming case as they defined in (11), (13) and (14). All the dispersion measures used correlate quite well. Laughter and disgust models possess the least dispersion, while surprise models show the largest dispersion among individuals in the Cohn-Kanade expression database [37]. This can be explained by the fact that surprise contains stronger node displacements. It is worth noticing that trimming reduces class dispersion, as expected. Furthermore, Figure 10 shows the dispersion of

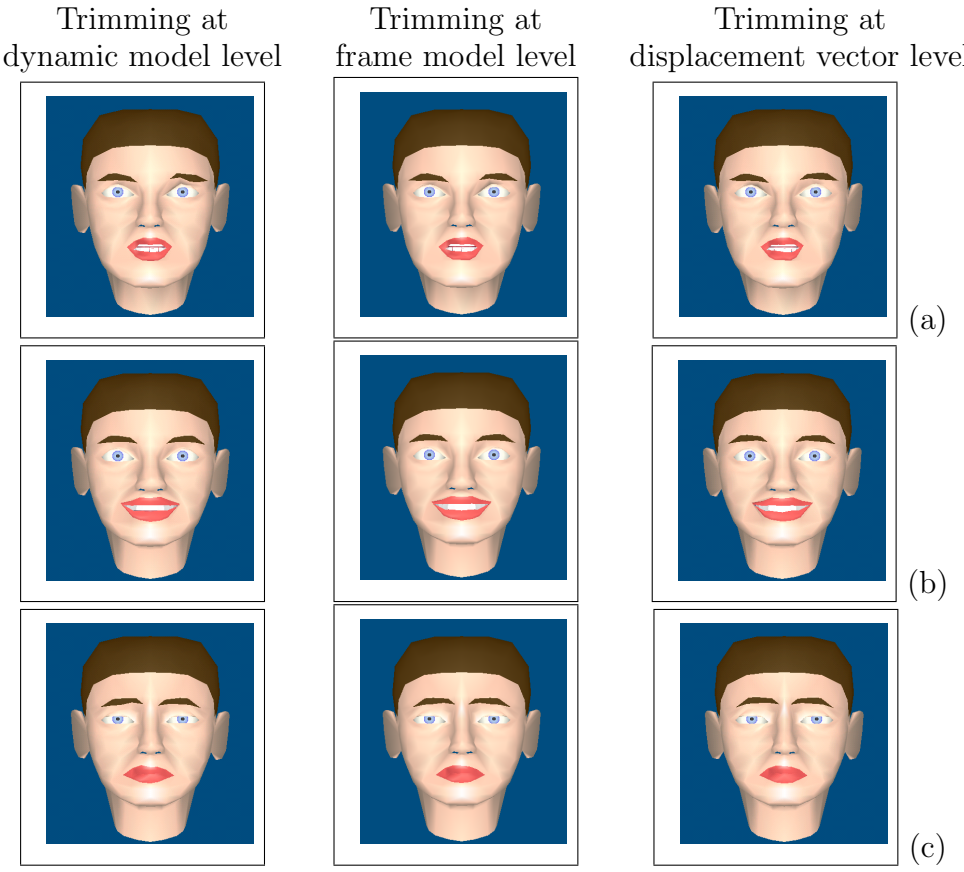


FIGURE 9. *Generalized median models using Euclidean distance visualized in an MPEG-4 player for* ***(a)*** *anger,* ***(b)*** *smile and* ***(c)*** *sadness, at three different trimming levels (dynamic model, frame model and displacement vector level).*

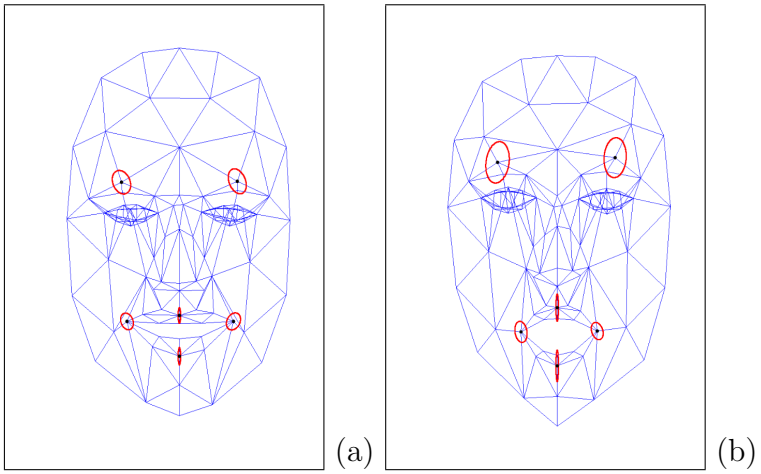


FIGURE 10. *The dispersion of some of the model nodes at full expression (for all corresponding tested video sequences) represented by ellipses. Dispersion for* ***(a)*** *laughter and* ***(b)*** *surprise.*

some of the model displacement vectors in the 2-D spatial space represented by ellipses at full expression. This Figure clearly indicates how the model displacement vectors are dispersed for laughter and surprise. As can be seen in Tables 2, 3 and 4 the minimum

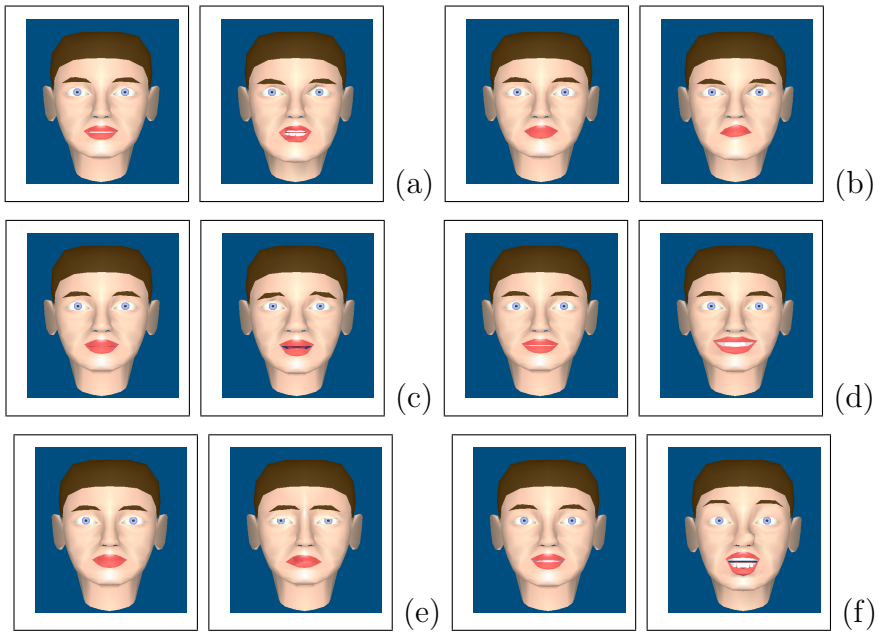


FIGURE 11. *Two frames (frame 3 and 5) of the representative dynamic facial expression model corresponding to* ***(a)*** *anger,* ***(b)*** *disgust,* ***(c)*** *fear,* ***(d)*** *laughter,* ***(e)*** *sadness, and* ***(f)*** *surprise.*

TABLE 3. *Median* $med_M$ *of the dynamic facial expression models for the six basic human facial expressions. The trimming cases are: (*$1^{st}$ ***case)*** *without trimming, (*$2^{nd}$ ***case)*** *trimming at the dynamic model level, (*$3^{rd}$ ***case)*** *trimming at the frame model level, and (*$4^{th}$ ***case)*** *trimming at the displacement vector level.*

| | **Anger** | **Disgust** | **Fear** | **Laughter** | **Sadness** | **Surprise** |
|---|---|---|---|---|---|---|
| $1^{st}$ case | 15,91 | 14,72 | 16,81 | 14,87 | 16,82 | 18,83 |
| $2^{nd}$ case | 15,49 | 14,52 | 16,36 | 14,57 | 16,65 | 17,39 |
| $3^{rd}$ case | 15,35 | 14,49 | 16,21 | 14,53 | 16,52 | 17,36 |
| $4^{th}$ case | **14,66** | **13,06** | **15,20** | **13,21** | **15,23** | **15,80** |

TABLE 4. ***Dispersion*** $\hat{\sigma}_{MAD}$ *of the dynamic facial expression models for the six basic human facial expressions. The trimming cases are: (*$1^{st}$ ***case)*** *without trimming, (*$2^{nd}$ ***case)*** *trimming at the dynamic model level, (*$3^{rd}$ ***case)*** *trimming at the frame model level, and (*$4^{th}$ ***case)*** *trimming at the displacement vector level.*

| | **Anger** | **Disgust** | **Fear** | **Laughter** | **Sadness** | **Surprise** |
|---|---|---|---|---|---|---|
| $1^{st}$ case | 12,91 | 11,18 | 13,21 | 13,72 | 15,34 | 15,82 |
| $2^{nd}$ case | 11,54 | 10,11 | 11,70 | 11,17 | 14,01 | 14,53 |
| $3^{rd}$ case | 11,13 | 9,80 | 11,39 | 10,89 | 13,46 | 13,57 |
| $4^{th}$ case | **11,08** | **9,65** | **10,85** | **10,05** | **11,35** | **13,25** |

model dispersion for all expressions under examination is achieved by the method extracting the generalized dynamic facial expression median model after trimming at the displacement vector level. However, a visual inspection of the full expressions, shown in

Figure 9, leads us to the conclusion that the generalized dynamic facial expression median model obtained after trimming at the frame model level achieves the best subjective ranking. Unfortunately, it is impossible to show the entire dynamic facial expression sequence played in an MPEG-4 player in the printed paper to justify this claim. The representative (generalized median) dynamic facial expression models have also been converted to MPEG-4 FAP files, so that they can be displayed by any MPEG-4 player. In our case, the facial model developed in the context of the European project ACTS MoMuSys [60] was used to this end. Figure 11a shows two frames (frame 3 and 5) of the representative dynamic facial expression model for anger, while Figures 11b, 11c, 11d, 11e and 11f illustrate the representative dynamic models for disgust, fear, laughter, sadness and surprise, respectively, obtained using the generalized dynamic facial expression median model after trimming at the frame model level. It has been found experimentally that this approach produces the best subjective facial expression results. This is best noticed when observing the entire MPEG-4 face animation sequence.

6. **Conclusion.** In this paper we presented a complete method for the statistical analysis of human facial expressions. The analysis was performed in the framework of facial expression analysis and produces results that are MPEG-4 compatible. The data used for the statistical analysis were obtained by tracking a generic facial wireframe model in video sequences depicting the formation of different human facial expressions, starting from a neutral state to a fully expressive one, using a pyramidal variant of the well-known Kanade-Lucas-Tomasi (KLT) tracker. Any loss of tracked features is handled through a physics-based deformation stage, after each single tracking step, providing accuracy and reliability. Tracking initialization is performed in a semi-automatic fashion. The facial wireframe model is fitted to a neutral facial image using a deformable shape modeling approach which is robust, fast and accurate. The output is MPEG-4 compliant and can be utilized in any MPEG-4 player. The method has been tested on a variety of sequences with very good results, including the Cohn-Kanade database of video sequences representing human facial expressions. It has been shown that the proposed method performs a number of intermediate steps in a reliable and accurate way and thus achieves a good statistical analysis of facial expressions. Furthermore, the extraction of the *generalized median dynamic facial expression models* for each of the six human facial expressions could lead us to generalization of the expressions of a specific racial group, as well as to locate the facial expression difference among different racial groups.

**Acknowledgement.** The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n 211471 (i3DPost).

## REFERENCES

[1] M. Pantic and L. Rothkrantz, Automatic analysis of facial expressions: The state of the art, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 22, no. 12, pp. 1424–1445, 2000.

[2] B. Fasel and J. Luttin, Automatic facial expression analysis: Survey, *Pattern Recognition*, vol. 36, no. 1, pp. 259–275, 2003.

[3] Z. Zeng, M. Pantic, G. Roisman, and T. Huang, A survey of affect recognition methods: Audio, visual, and spontaneous expressions, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 31, no. 1, pp. 39–58, January 2009.

[4] A. Nikolaidis and I. Pitas, Facial feature extraction and pose determination, *Pattern Recognition*, vol. 33, no. 11, pp. 1783–1791, November 2000.

[5] K. Sobottka and I. Pitas, A novel method for automatic face segmentation, facial feature extraction and tracking, *Image Communication*, vol. 12, no. 3, pp. 263–281, June 1998.

[6] A. Raouzaiou, N. Tsapatsoulis, K. Karpouzis, and S. Kollias, Parameterized facial expression synthesis based on MPEG-4, *EURASIP Journal on Applied Signal Processing*, vol. 10, pp. 1021–1038, 2002.
[7] Z. Ruttkay, J. Hendrix, P. Hagen, A. Lelievre, H. Noot, and B. Ruiter, A facial repertoire for avatars, *Proc. of Workshop Interacting Agents*, 2000.
[8] I. Buciu, I. Kotsia, and I. Pitas, Recognition of facial expressions in presence of partial occlusion, *Proc. of the 9th Panhellenic Conference on Informatics (PCI'03)*, November 2003.
[9] I. Buciu, C. Cotropoulos, and I. Pitas, ICA and Gabor representation for facial expression recognition, *Proc. of IEEE International Conference on Image Processing (ICIP'03)*, September 2003.
[10] E. Sifakis, I. Neverov, and R. Fedkiw, Automatic determination of facial muscle activations from sparse motion capture marker data, *ACM Trans. Graphics (TOG)*, vol. 24, no. 3, pp. 417–425, July 2006.
[11] Y. Chang, C. Hu, R. Feris, and M. Turk, Manifold based analysis of facial expression, *Journal of Image and Vision Computing*, vol. 24, no. 6, pp. 605–614, 2005.
[12] Z. Zeng, J. Tu, M. Liu, T. Huang, B. Pianfetti, D. Roth, and S. Levinson, Audio-visual affect recognition, *IEEE Trans. Multimedia*, vol. 9, no. 2, pp. 424–428, February 2007.
[13] P. Hong, Z. Wen, and T. Huang, IFACE: A 3D synthetic talking face, *International Journal of Image and Graphics*, vol. 1, no. 1, pp. 1–8, 2001.
[14] Y. Fu and N. Zheng, M-Face: An appearance-based photorealistic model for multiple facial attributes rendering, *IEEE Trans. Circuits and Systems for Video Technology*, vol. 16, no. 7, pp. 830–842, July 2006.
[15] Y. Fu, R. Li, T. Huang, and M. Danielsen, Real-time multimodal human-avatar interaction, *IEEE Trans. Circuits and Systems for Video Technology*, vol. 18, no. 4, pp. 467–477, 2008.
[16] Z. Deng, U. Neumann, J. Lewis, T. Kim, M. Bulut, and S. Narayanan, Expressive facial animation synthesis by learning speech co-articulation and expression spaces, *IEEE Trans. Visualization and Computer Graphics*, vol. 12, no. 6, 2006.
[17] S. Krinidis, I. Buciu, and I. Pitas, Facial expressions analysis and synthesis: A survey, *Proc. of International Conference on Human Computer Interaction (HCI'03)*, June 2003.
[18] Z. Wen and T. Huang, *3D Face Processing: Modeling, Analysis and Synthesis.* Kluwer Academic Publishers, 2004.
[19] Z. Deng and U. Neumann, eFACE: Expressive facial animation synthesis with phoneme-isomap controls, *Proc. of ACM SIGGRAPH/EG Symposium on Computer Animation*, 2006.
[20] J. Lien, Automatic recognition of facial expression using hidden markov models and estimation of expression intensity, Ph. D dissertation, The Robotics Institute, 1998, CMU.
[21] K. Mase and A. Pentland, Recognition of facial expression from optical flow, *IEICE Trans.*, vol. 74, no. 10, pp. 3474–3483, October 1991.
[22] T. Otsuka and J. Ohya, Spotting segments displaying facial expressions from image sequences using hmm, *Proc. of the 2nd International Conference on Automatic Face and Gesture Recognition (FG'98)*, pp. 442–447, April 1998.
[23] M. Yoneyama, Y. Iwano, A. Ohtake, and K. Shirai, Facial expression recognition using discrete hopfield neural networks, *Proc. of International Conference on Image Processing (ICIP'97)*, vol. 3, pp. 117–120, 1997.
[24] D. Terzopoulos and K. Waters, Analysis of facial images using physical and anatomical models, *Proc. of the 3rd International Conference on Computer Vision*, pp. 727–732, 1990.
[25] D. DeCarlo and D. Metaxas, The integration of optical flow and deformable models with applications to human face shape and motion estimation, *Proc. of International Conference on Computer Vision and Pattern Recognition (CVPR'96)*, pp. 231–238, 1996.
[26] M. Black and Y. Yacoob, Recognizing facial expressions in image sequences using local parameterized models of image motion, *International Journal of Computer Vision*, vol. 25, no. 1, pp. 23–48, 1997.
[27] Y. Yacoob and L. Davis, Recognizing human facial expression from long image sequences using optical flow, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 18, no. 6, pp. 636–642, 1996.
[28] Y. Tian, T. Kanade, and J. Cohn, Recognizing action units for facial expression analysis, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 23, no. 2, 2001.
[29] M. Wang, Y. Iwai, and M. Yachida, Expression recognition from time-sequential facial images by use of expression change model, *Proc. of the 2nd International Conference on Automatic Face and Gesture Recognition (FG'98)*, pp. 324–329, April 1998.

[30] B. Bascle and A. Blake, Separability of pose and expression in facial tracking and animation, *Proc. of International Conference on Computer Vision (ICCV'98)*, January 1998.
[31] G. Donato, S. Bartlett, C. Hager, P. Ekman, and J. Sejnowski, Classifying facial actions, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 21, no. 10, pp. 974–989, October 1999.
[32] B. Fasel and J. Luettin, Recognition of asymmetric facial action unit activities and intensities, *Proc. of International Conference on Pattern Recognition (ICPR'00)*, 2000.
[33] P. Belhumeur, J. Hespanha, and D. Kriegman, Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711–720, 1997.
[34] J.-S. Pan, J.-B. Li, and Z.-M. Lu, Adaptive quasiconformal kernel discriminant analysis, *Neurocomputing*, vol. 71, no. 13-15, pp. 2754–2760, August 2008.
[35] J.-B. Li, J.-S. Pan, and S.-C. Chu, Kernel class-wise locality preserving projection, *Information Sciences*, vol. 178, no. 7, pp. 1825–1835, April 2008.
[36] M. Rydfalk, CANDIDE: A parameterized face, *Ph. D dissertation*, Linkoping University, Linkoping, 1978.
[37] T. Kanade, J. Cohn, and Y. Tian, Comprehensive database for facial expression analysis, *Proc. of International Conference on Face and Gesture Recognition*, pp. 46–53, March 2000.
[38] J. Bouguet, Pyramidal implementation of the lucas kanade feature tracker, Intel Corporation, Microprocessor Research Labs, OpenCV Documents, Tech. Rep., 1999.
[39] MPEG, Coding of audio-visual objects - Part 2: Systems, *ISO/IEC 14496-1:2001*, 2nd Edition 2001.
[40] MPEG, Coding of audio-visual objects - Part 2: Visual, *ISO/IEC 14496-2:2001*, 2nd Edition 2001.
[41] P. Ekman and W. Friesen, *The Facial Action Coding System.* San Francisco, Calif, USA: Consulting Psychologists Press, 1978.
[42] J. Ostermann, *Face Animation in MPEG-4.* Chichester, U.K.: John Wiley & Sons, 2002.
[43] E. Jang and J. Ostermann, *The MPEG-4 Book.* Upper Saddle River, N.J.: Prentice Hall PTR, 2002.
[44] C. Whissel, *The dictionary of affect in language in Emotion: Theory, Research and Experience.* New York, USA: Academic Press, 1989.
[45] R. Plutchik, *A Psychoevolutionary Synthesis.* New York, USA: Harper and Row, 1980.
[46] J. Ahlberg, CANDIDE-3 – An updated parameterized face, Dept. of Electrical Engineering, Linkoping University, Sweden, Tech. Rep., 2001.
[47] C. Tomasi and T. Kanade, Shape and motion from image streams: A factorization method - part 3 detection and tracking of point features, Computer Science Department, Carnegie Mellon University, Tech. Rep., pp. 91-132, 1991.
[48] J. Shi and C. Tomasi, Good features to track, *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR'94)*, pp. 593–600, June 1994.
[49] N. Nikolaidis and I. Pitas, *3-D Image Processing Algorithms.* New York, USA: John Wiley & Sons, Inc., 2001.
[50] I. Pitas and A. Venetsanopoulos, *Nonlinear Digital Filters: Principles and Applications.* Norwell, MA: Kluwer Academic Publishers, January 1990.
[51] S. Loncaric, A survey of shape analysis techniques, *Pattern Recognition*, vol. 31, no. 8, pp. 983–1001, 1998.
[52] H. Bunke, Recent advances in structural pattern recognition with applications to visual form analysis, *Visual Form*, pp. 11–23, 2001.
[53] H. Bunke, X. Jiang, K. Abegglen, and A. Kandel, On the weighted mean of a pair of strings, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 5, no. 1, pp. 23–30, 2002.
[54] X. Jiang, A. Munger, and H. Bunke, On median graphs: Properties, algorithms and applications, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 23, no. 19, pp. 1144–1151, October 2001.
[55] X. Jiang, A. Munger, and H. Bunke, Synthesis od representative graphical symbols by computing generalized median graph, *Proc. of the 3rd IAPR International Workshop on Graphics Recognition (IAPR'99)*, pp. 187–194, 1999.
[56] A. Bors and I. Pitas, Median radial basis functions neural network, *IEEE Trans. Neural Networks*, vol. 7, no. 6, pp. 1351–1364, November 1996.
[57] G. Seber, *Multivariate Observations.* New York, USA: John Wiley, 1986.
[58] C. Cotropoulos, I. Pitas, and M. Gabbouj, Marginal median learning vector quantizer, *Proc. of European Signal Processing Conf. (EUSIPCO '94)*, pp. 1496–1499, August 1994.

[59] I. Pitas, C. Cotropoulos, N. Nikolaidis, R. Yang, and M. Gabbouj, A class of order statistics learning vector quantizers, *Proc. of IEEE International Symposium on Circuits and Systems*, vol. 4, pp. 387–390, 1994.

[60] G. Abrantes and F. Pereira, MPEG-4 facial animation technology: Survey, implementation and results, *IEEE Trans. Circuits and Systems for Video Technology*, vol. 9, no. 2, pp. 290–305, 1999.