Automatic Real Time Localization of Frowning and Smiling Faces under Controlled Head Rotations

YULIA GIZATDINOVA, JOUNI EROLA, AND VEIKKO SURAKKA
Department of Computer Sciences
University of Tampere
Kanslerinrinne 1, 33014
FINLAND
yulia.gizatdinova@cs.uta.fi

Abstract: - The aim of the present study was to develop a new method for fast localization of the face from streaming colour video. In the method, the face is located by analysing local properties of four facial landmarks, namely, regions of eyes, nose, and mouth. The method consists of three stages. First, the face-like skin-coloured image region is segmented from the background and transformed into the grey scale representation. Second, the cropped image is convolved with Sobel operator in order to extract local oriented edges at 16 orientations. The extracted edges are further grouped to form regions of interest representing candidates for facial landmarks. The orientation portraits, which define the distribution of local oriented edges inside the located region, are matched against the edge orientation model to verify the existence of the landmark in the image. The located landmarks are spatially arranged into the face–like constellations. Finally, the best face-like constellation of the landmark candidates is defined by a new scoring function. The test results showed that the proposed method located neutral, frowning, and smiling faces with high rates in real time from facial images under controlled head rotation variations.

Key-Words: - Face localization, Facial landmarks, Sobel edge detection, Frontal-view geometrical face model, Facial expressions, Head rotations

1 Introduction

Generally, face detection is considered as a specific case of object-class detection in which the task is to find the locations and sizes of all objects belonging to a certain class. Given an image or video frame, face detection is defined as a process of automatic finding of the true locations and sizes of the faces if there are any, or declaring about their absence [17,39]. Face localization is a simplified detection problem with the assumption that a face is shown in the image or video sequence. Given an image or video frame with a known number of faces (usually one), the task is to find the true face location and size. Automatic face detection has a wide range of applications. The face detection data are delivered as an input to various systems of automatic face analysis such as face recognition, facial expression recognition, and perceptual vision-based user interfaces. To ensure that these systems work in real time efficiently and robustly, face detection is aimed to provide high speed and accuracy of the detection process.

Human faces constitute a class of visually similar objects with a rigid structure that does not vary significantly from person to person (*e.g.* typically there are two eyes and the nose is located between eyes and mouth) [32]. This makes the task of distinguishing faces from objects of other classes relatively easy. However, within the face class there exist variations which make the detection of facial features a difficult task.

Facial appearance varies noticeably with changes in environmental conditions and between and within individuals [7,8]. Changes in the environmental conditions which typically impair a performance of face detectors are due to illumination, out- and in-plane head rotations, scene complexity, camera characteristics (*i.e.* image resolution, viewing distance, camera sensor noise), scale or image size, structural components like eye-glasses, and occlusions. The inner properties of the face determine between-individual variations which are due to race and gender, and within-subject variations which are due to facial expressions and age-specific face modifications [7,8]. Because the face changes noticeably in its appearance, the main challenge of automatic face detection is to find a face representation that is compact, distinctive, fast to compute, and remains robust with respect to various changes in facial appearance.

The aim of the present paper was to develop a new feature-based method for face localization from streaming colour video where facial expressions and head rotations were carefully controlled. In the method proposed, the face is located by analysing local properties of four facial landmarks, namely, regions of eyes, nose, and mouth. The multi-orientation edge representation of the facial image is used to compose candidates for facial landmarks. This is followed by spatial arrangement of the located landmark candidates. We present a new scoring function designed to define the best face-like constellation of the landmark candidates. The features to be located in the face are the centres of the eyes, nose tip, and the centre of the mouth. After facial landmarks are located, the location of the face in the image is also known.

The rest of the paper is organized as follows. Next section describes approaches to the existing face detection techniques. Section 3 describes our solution to the feature-based face detection. Section 4 describes the video database of faces under expression and head rotation variations used in this study. Section 5 presents the test results of the method. Section 6 discusses the results and concludes the paper.

2 Approaches to Face Detection

Many attempts have been undertaken to automatically detect faces and facial features from static images and videos [15,17,24,39]. Generally, the proposed methods can be classified as template-based, learning-based, and feature-based methods. Further, they can be classified as global and local face detectors. The global detectors are used to search for a pattern of the whole face. The methods from this category take as an input a global representation of the face. A distinctive property of these methods is that geometrical information on the facial pattern is preserved. In some cases, the use of a global representation has its own advantages. For example, preliminary face localization enables fast feature extraction because it discards those parts of the image that are irrelevant for the feature search [2,13,14,21,22]. However, global face representation has serious limitations. First, because the face is a highly non-rigid visual object, global representation is highly sensitive to occlusions and changes in facial expressions, head rotations, and lighting. For example, in a presence of between- and within-individual variations of facial expressions, it is difficult to enumerate all possible face models without a hierarchical decomposition of the problem. In addition, global representation requires processing of high dimensional data, which is a difficult problem.

The local face detectors utilize a local face representation and are applicable for the task of detecting independent facial features. Facial features to detect are typically those which are the most distinctive for humans - eyebrows, eyes, nose, and mouth [7,32]. The location of the face is further derived from the landmark locations. This approach has the advantage of producing the exact positions of facial landmarks which can be used in further face processing tasks.

In contrast to global face representation, appearances of single facial landmarks exhibit less variability. Facial features analysed locally usually vary less in pose, lighting, and facial expression changes than the pattern of the whole face. Therefore, by focusing first on facial parts, a local image representation makes it possible to handle a wider range of conditions than the global one. Generally, it has been shown that local face and feature detectors outperform global ones in a number of conditions including head rotations, changes in illumination, occlusions, and facial expressions [1,5,10,16,33]. In addition, local detectors operate in a small search space by reducing false or impossible feature constellations in the early stages of processing, resulting in efficient and fast computation.

2.1 Template-Based Approach

Facial templates are frequently used to encode facial information and further search for a correlation between facial templates and the input image or video frame. The structure of the facial templates can be predefined or deformable. The predefined templates are usually easy to construct and compute. It makes sense to use the predefined templates for face and feature detection under strictly controlled conditions (*i.e.* constant illumination, restricted head movements, simple background, neutral (or known) expression, *etc.*). However, if these conditions fail, the detection rates decrease significantly. The use of predefined templates locally for the detection of separate facial landmarks partly solves the problem, as separate facial parts exhibit less variability than the pattern of the whole face.

Recently, extensive work has focused on a deformable property of the face. Examples of deformable templates are active contour models or snakes [1,25], active appearance models (AAMs) [4], constrained local models [5], and active blobs [30]. Deformable templates such as snakes have been demonstrated to be robust regardless of

expressive facial deformations [1]. Thus, face and features can be successfully detected in spite of wide variations in expression when the eyes are closed or semi-closed and the mouth is open and the teeth are visible. A limitation of snake-based detectors is that they have to be initialized near the face or feature in the image. Otherwise they may fail to get a right start. Other deformable templates might suffer from strong facial expressions as AAA-based methods do. Marked changes in lighting, out-of-plane head rotations, and occlusions also greatly impair the performance of these methods. Therefore preliminary face detection is required to perform the feature detection in cluttered scenes. Another limitation of deformable templates is that they usually require high-resolution images and are not easily achievable in real time.

2.2 Learning-Based Approach

Another class of face detectors is represented by learning classifiers. These methods learn the appearance of the face and facial parts from examples (*i.e.* training set). Face or feature models are constructed from the training image set and further searched for in the image. This way, no knowledge of facial structure is required. The linear subspace methods assume that any face can be represented as a linear combination of all other faces from the face space. Examples of methods from this category are principal component analysis (PCA), Gaussian component analysis, independent component analysis, factor analysis, projection pursuit, and symplectic maps to name few. For a recent review on these methods see [6]. Principle Component Analysis (PCA) has been widely adopted for the purpose of face detection [14,38]. Generally, PCA-based methods can handle faces with nearly the same pose, constant illumination, and moderate facial expressions. These methods require the training data set to be normalized according to illumination and head pose.

Apart from PCA-based methods, there are other learning-based methods like boosted classifiers [35], support vector machines [23,40], and neural networks [27,28] which have to be trained on the representative sets of face and non-face images. The training set should cover a wide variability of facial appearances including faces with different illumination conditions, facial expressions, and view points. Usually, the bigger the training set, the better differentiation between face and non-face classes can be achieved. Because the whole image is fed to the system, learning classifiers need to deal with the problem of dimensionality reduction. They usually apply a preprocessing step to reduce a dimension of the input data and select optimal modal parameters in the received low-dimensional space.

2.3 Feature-Based Approach

In practice, feature-based face detection includes a selection of feature representation and a design of feature detector. These methods generally use the low-level facial features such as edges, lines, line intersections, points, and regions. The high-level processing further combines the low-level features into meaningful feature formations which represent specific visual patterns for the identification of corresponding facial structures between images. The high-level stage also performs a reduction of false detections received from the low-level stage. In order to judge whether or not the detected image parts constitute a face, the high-level stage uses knowledge of what constitutes a human face. A constellation analysis is applied to arrange the detected feature candidates into face-like formations, for example, by statistical probabilistic models [29]. Alternatively, only confidence measures received from individual low-level feature detectors are used. Many proposed face detectors utilize edges [9], grey-scale values [1], and their combinations [19,20,31]. A multi-resolution and multi-orientation representation of the image has been widely adopted for the purpose of landmark detection [37,40] and demonstrated to be effective in face detection under expression and small pose variations [11,12]. However, these methods are generally computationally expensive and, therefore, are not applicable for the task of real-time face and facial feature detection. In the subsequent sections we demonstrate how our proposed method successfully overcomes speed limitations of the feature-based face detection methods.

3 Feature-Based Face Localization

The method proposed for real-time face localization consists of three stages (Fig. 1). Given a facial video, the first stage finds a rough approximation of the face location in each video frame. The second stage forms and extracts all possible landmark candidates from the cropped facial region. The last stage decides whether constellations formed from the located landmark candidates meet requirements placed by the geometrical frontal-view facial model, and if so, defines the location of the best-scored face candidate. Below we explain each stage of the method in more detail.

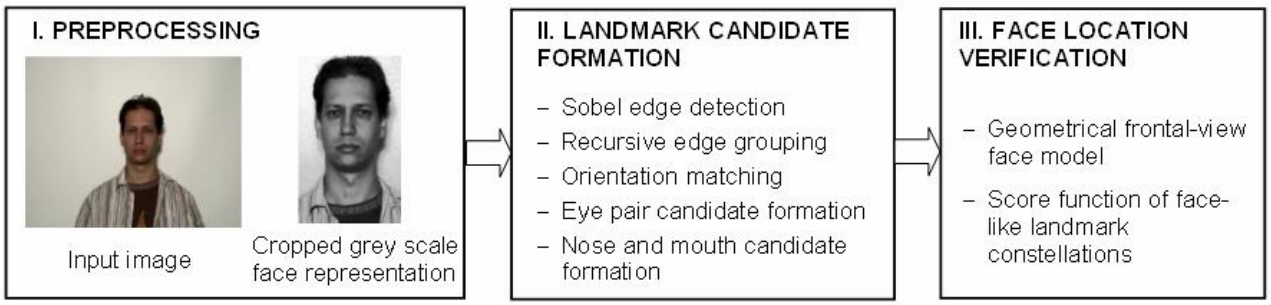

Figure 1. Data flow in our system of real-time face localization.

3.1 Preprocessing

On this stage, a face-like image region is segmented from the background. To do that we first apply the procedure of histogram equalization to each video frame for all three RGB colour channels. This calculation is not computationally expensive and is widely used to allow areas of low local contrast to gain a higher contrast without affecting the global contrast of the whole image [13]. After this, the original RGB image is transferred into YCbCr chromatic colour space as proposed in [22]:

$$Y_{ij} = 0.299R_{ij} + 0.587G_{ij} + 0.114B_{ij} \quad (1)$$

$$Cb_{ij} = -0.16874R_{ij} - 0.33126G_{ij} + 0.5B_{ij} + 128$$

$$Cr_{ij} = -0.5R_{ij} - 0.41869G_{ij} - 0.08131B_{ij} + 128$$

where R, G, and B are red, green , and blue components of RGB colour space; Y, Cb, and Cr are luminance, blue, and red components of YCbCr chromatic colour space.

Next, we segment the skin-coloured regions of the image similarly to the method proposed in [1]. This procedure allows for noisy regions of the image to be discarded at the early stage of the processing and, therefore, focuses the following stages of the method on those parts of the image in which the face is more likely to be located. The Gaussian-fitted skin colour model is used for this purpose. The idea that lies behind is that a distribution of skin colour for different people is clustered in the chromatic colour space and can be represented by a Gaussian model. The likelihood P of a skin colour for any pixel (x,y) of the image thus can be obtained with a Gaussian-fitted skin colour model:

$$P = e^{\left(-\frac{1}{2 \cdot Cv}\cdot\left((Cb-Cb_{mean})^2+(Cr-Cr_{mean})^2\right)\right)}, \quad (2)$$

where Cv is a covariance matrix; Cb is a blue chromatic value; Cr is a red chromatic value; Cb_{mean} is an average blue chromatic value; and Cr_{mean} is an average red chromatic value.

The skin-coloured regions are next segmented from the rest of the image through a thresholding process. The parameters for a thresholding are selected experimentally using a small image set from the database. Finally, the received regions are cropped from the background and the procedure of histogram equalization is applied for them. This stage outputs 8-bit grey scale image of the face. If a face is not extracted, the following steps of the method are applied to the whole image converted into 8-bit grey scale representation.

3.2 Landmark Candidate Formation

3.2.1 Sobel Edge Detection

On the next stage, local oriented edges of different orientations are extracted from the cropped facial image. Fig. 2 illustrates the edge orientation template used for edge detection. It consists of 16 orientations taken with a step of $22.5°$. In this case, orientation 0 is taken to mean that the direction of maximum contrast from light to dark runs from left

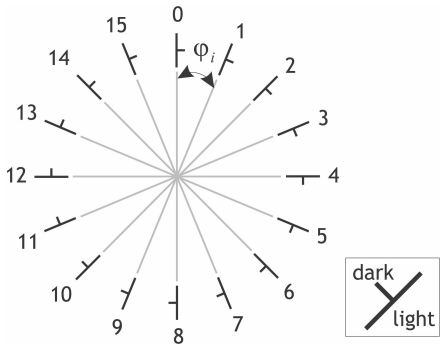

Figure 2. Orientation template for detection of local oriented edges, $\varphi = 22.5°$, $i = 0 \div 15$. Orientations $2 \div 6$ and $10 \div 14$ are used for orientation matching.

-1	0	+1
-2	0	+2
-1	0	+1

Gx

+1	+2	+1
0	0	0
-1	-2	-1

Gy

Figure 3. Convolution kernels of the Sobel edge detector.

to right on the image, and other angles are measured clockwise from this.

We used the Sobel operator [13,18] to detect local oriented edges. The choice of Sobel edge detector was motivated by the fact that it is fast, easy to implement, and well established in literature. Fig. 3 shows the kernels of the Sobel edge detector. The kernels give the maximal respond to vertical and horizontal local edges relatively to the pixel grid of the image. The kernels can be separately applied to the input image. This enables fast calculation of the separate measurements of the gradient component in each orientation (G_x and G_y respectively). The magnitude of the gradient at each point and at each orientation of that gradient is given by:

$$|G| = \sqrt{G_x^2 + G_y^2}\,, \tag{3}$$

where $|G|$ is the absolute magnitude of the gradient at the image point (x, y); G_x and G_y are the gradient components in vertical and horizontal orientations, respectively.

Typically, the absolute magnitude of the gradient is calculated using its approximation:

$$|G| = |G_x| + |G_y|\,, \tag{4}$$

where $|G_x|$ and $|G_y|$ are the absolute gradient components in vertical and horizontal orientations, respectively.

The angle of orientation of the edge (relative to the pixel grid) giving rise to the spatial gradient is given by:

$$\theta = \arctan\left(\frac{G_x}{G_y}\right), \tag{5}$$

The edge points are further recursively grouped together to form regions of interest which represent the candidates for facial landmarks.

The shapes of the bounding boxes which are placed over the located regions of interest are analyzed next. If a candidate is bounded by a box which has height much bigger than its width, the candidate is eliminated.

3.2.2 Edge Orientation Matching

After this, among the located candidates there still exist many noisy regions which have to be eliminated. In order to define regions which represent landmark candidates we analyze local properties of the located regions of interest. As it has been demonstrated earlier [11,12], regions of facial landmarks have a characteristic distribution of local oriented edges with two horizontal dominants (Fig. 4). On the other hand, non-landmark regions which are, for example, elements of face, hair, clothing, and decoration typically do not have a characteristic structure of the oriented edges. These regions demonstrate a random distribution of the oriented edges (Fig. 5). This local property of the located regions of interest allowed us to discard noisy regions while preserving regions which contain facial landmarks.

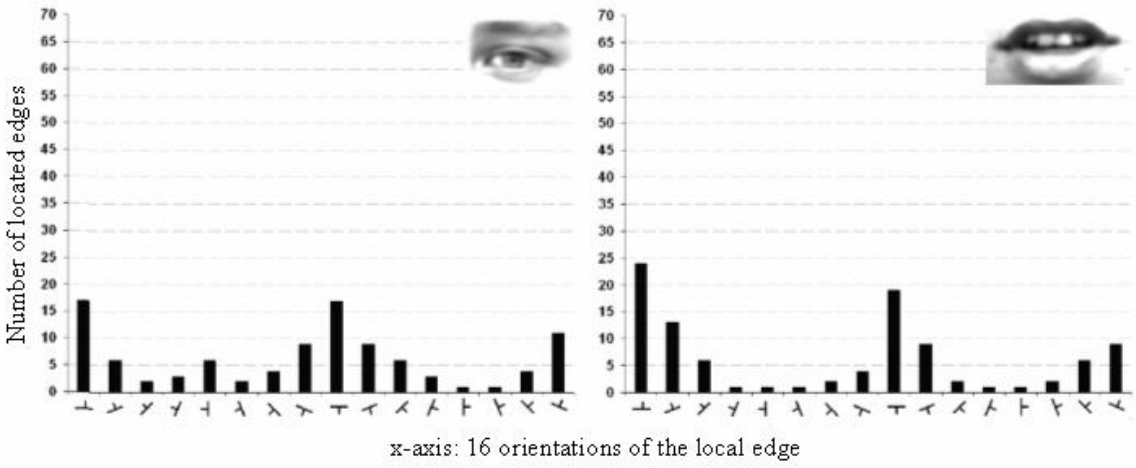

Figure 4. Orientation portraits of facial landmarks with the characteristic distribution of oriented edges.

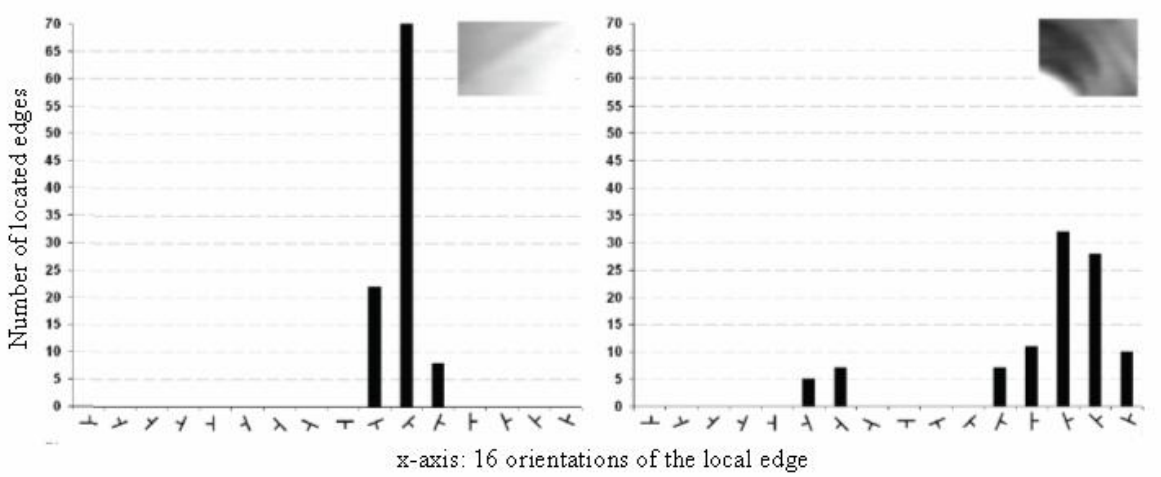

Figure 5. Orientation portraits of noisy regions with a random distribution of the oriented edges.

3.2.3 Geometrical face model

In order to classify the located candidates and find their proper spatial arrangement, the proposed method applies a set of verification rules which are based on face geometry. The knowledge on face geometry is taken from the anthropological study by Farkas [8]. This thorough study examined thousands of Caucasians, Chinese, and African-American subjects in order to determine characteristic measures and proportion indexes of the human face. We performed several tests to verify the existing facial measures, calculate new measures, and built a frontal-view geometrical face model depicted in Fig. 6.

The anthropometric facial features and measures of the model are described in Table 1. Centre points of facial landmarks are calculated as mass centres of the located edge regions. It has been demonstrated that facial measures from the table can slightly vary between subjects of different gender, age, and race [7,8]. Therefore, we define constrains for facial measures as intervals between minimum and maximum values for a given measure. All distance constrains from the table are defined as percentages of the interocular distance[1] $d(E_R, E_L)$. As our preliminary tests demonstrated, this set of facial measures and their corresponding constrains achieved good results in composing face-like constellations from the located landmark candidates.

The classification of the landmark candidates proceeds as follows. The eye pair candidates are found first as any two candidates aligned nearly

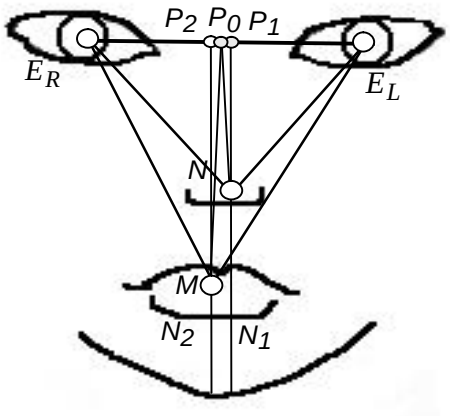

Figure 6. Frontal-view geometrical face model.

Table 1: Features and measures of the frontal-view geometrical face model (and their constrains).

Feature/measure	Feature description
E_R, E_L, N, M	Centres of the landmarks
$d(E_R, E_L)$	Interocular distance
N_1, N_2	Perpendiculars to $d(E_R, E_L)$
P_1, P_2	Cross points of $d(E_R, E_L)$ and N_1 and N_2, correspondently
P_0	Middle point between P_1 and P_2
$d(N, M)$	Nose-mouth distance (30-110% of $d(E_R, E_L)$)
$\angle NP_0M$	Nose-mouth angle (0-16°)
$\angle E_RNP_1$, $\angle E_LNP_1$, $\angle E_RMP_2$, $\angle E_LMP_2$	Eye-nose and mouth-eyes angles (0-13°)
$d(E_R, N)$, $d(E_L, N)$	Eye-nose distance (25-120% of $d(E_R, E_L)$)
$d(E_R, M)$, $d(E_L, M)$	Eye-mouth distance (60-160% of $d(E_R, E_L)$)

[1] The interocular distance is the distance between the centres of the eyes (Farkas, 1994).

horizontally. After this step, all possible nose and mouth candidates for each found eye pair candidate are independently searched for using constrains of the frontal-view geometrical face model.

3.3 Face Location Verification

On the last stage, the defined eye-nose and eye-mouth candidates are combined together into a complete face candidate so, that the eye pair is the same for both eye-nose and eye-mouth candidates. Each found face candidate consisting of four facial landmarks is given a score. A score is calculated as a sum of intermediate scores which show how well a face candidate performs verification tests. The verification tests are fuzzy rules defined as follows:

- *Test 1* checks the horizontality of the eye pair candidate. The eye pair candidate that has the most horizontal position in the image as compared to others gives the lowest score for a given face candidate.
- *Test 2* checks angles $\angle E_R NP_1$, $\angle E_L NP_1$, $\angle E_R MP_2$, and $\angle E_L MP_2$ - face candidate with small angles gets low scores, and vice versa.
- *Test 3* considers the result of the previous face localization. If the landmark centre points in the previous frame are nearly the same as compared to those in the current frame (face is not moving), a face candidate gets a score which is lower than in the opposite case.
- *Test 4* checks sizes of the located landmark candidates – the biggest values give the lowest score for a face candidate.
- *Test 5* checks widths of the landmark candidates in the facial configuration. It has been validated that mouth is usually wider than eyes and nose has nearly the same width as eyes have [8]. The closer face candidate satisfies to this criterion, the lower score it gets from the test.
- *Test 6* utilizes the property of face symmetry and checks sizes of both eyes. If eyes have the same size, a face candidate gets the lowest score from this test. Each verification test is also given a weight.

We performed a number of tests to define optimal weights for each test. This way, each test gives as its output a relative score for a given face candidate. In order to select the best-scored face-like constellation of the located landmarks, a new scoring function is introduced:

$$P_r = MAX - \frac{MAX}{P_{cand} / P_{\min}}, \quad (6)$$

where *MAX* is a maximum score for a given face candidate (we used 100); P_{cand} is a current score for a given face candidate; $P_{\min}$ is the lowest score achieved among all face candidates. This way, if we have several face-like constellations of the landmarks, we select the one that gives the highest score P_r.

A localization result is considered correct if a distance between manually annotated and automatically located landmark location met the requirement placed by a performance evaluation measure elaborated in [26]:

$$d_{eye} = \frac{\max\left(d(E_R, E'_R), d(E_L, E'_L)\right)}{d(E_R, E_L)} \quad (7)$$

where $d(a,b)$ is Euclidean distance between point locations *a* and *b*; E_R, E_L are manually annotated and E'_R, E'_L are automatically located positions of facial landmarks. A successful localization was considered if $d_{eye} < 0.25$ which corresponds approximately to 1/4 of the annotated interocular distance $d(E_R, E_L)$ (a half of the width of the eye). After the locations of the landmarks are known, the location of the face in the image is also known.

4 Test Data

It is self evident that any realistic vision-based application, which utilizes automatic face analyses, implies the presence of facial expressions and head rotations. In this study, the effect of facial expressions and head rotations on the performance of the developed face detection method was

Figure 7. Examples of head rotations included in the database.

systematically studied. For this purpose, we created our own video database with neutral, frowning, and smiling faces under three controlled head rotations with angles of rotation of 0°, 20°, and 30° using both right and left directions. Examples of facial images frames from the video database are shown in Fig. 7. The low-cost Canon Mini-DV camera with 720x568 pixel image resolution and 24-bit precision for colour values was used. The potential impact of illumination, background, facial hair or eye-glasses was controlled to some extent in all video sequences and therefore ignored. No face alignment was performed.

The database consisted of 10 Caucasian subjects (40% females) with average age of 30 years. Each video began with neutral face, proceeding gradually to the expressive face and ending up with neutral face. The level of the expression intensity varied among different subjects. In total, 150 video sequences were created with duration of about 7-8 seconds. The test data were annotated in advance by recording the true locations of the landmark centres in each frame for each test subject.

Table 2: Rates (%) of localization of neutral, frowning, and smiling faces under three head rotations.

	Neutral			Frowning			Smiling		
	0°	20°	30°	0°	20°	30°	0°	20°	30°
S1	69	91	97	79	84	67	96	72	93
S2	95	51	17	82	74	31	87	67	17
S3	55	17	30	45	10	37	27	34	31
S4	98	68	60	98	83	35	87	72	48
S5	95	62	36	89	91	17	82	56	29
S6	33	47	49	21	51	47	26	62	46
S7	93	88	30	85	79	13	24	49	46
S8	63	61	25	81	35	8	90	44	14
S9	92	95	98	62	93	71	85	77	37
S10	91	72	30	97	82	64	63	65	45

5 Results

The tests were run on the computer Dell Optiplex 745, Intel Core2 with 2133 MHz and 1 GB DDR2-memory in Win XP 2002 SP 2/Delphi-environment with DirectShow-interface. The speed of the method was about 20-24 frames per second.

Fig. 8 shows the average rates of face localization under three expressions and three head rotations. The raw test data are shown in Table 2. In order to analyze the effects of facial expression and head rotations to the functioning of the method, a statistical analysis was done by using a two-way *3* x *3* (expression×head rotation) repeated measures analysis of variance (ANOVA). The statistical analysis showed that head rotation had a statistically significant main effect on the face localization $F(2,18)=9.31$, $p<0.001$. Bonferroni corrected pairwise post-hoc comparisons showed that the detection percentage was significantly lower when head was rotated by 30° as compared with frontal head position $MD=30.81$, $p<0.05$. Difference between 20° and 30° head rotations was also statistically significant $MD=22.14$, $p<0.05$. There

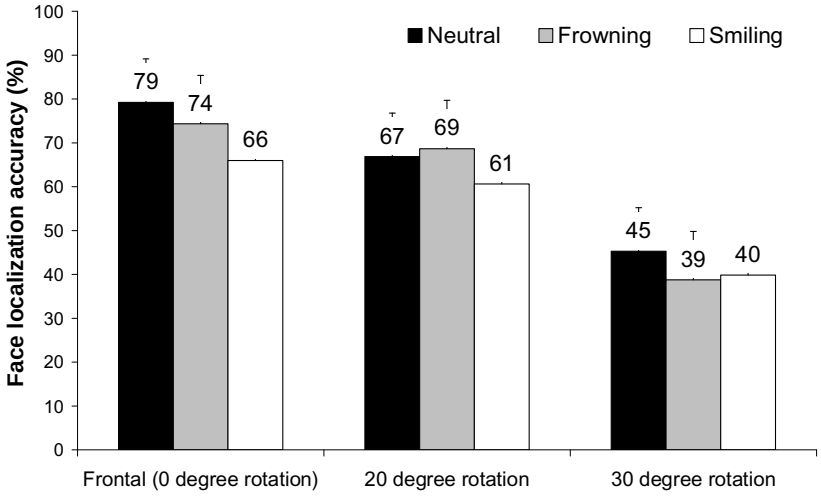

Figure 8. Average localization rates (%) of neutral, frowning, and smiling faces (all four landmarks were found) under three head rotations.

was no statistically significant difference between head rotation by 20° and frontal head position. ANOVA showed that there were no other significant main or interaction effects.

6 Discussion

The developed method demonstrated high speed performance in face localization from streaming colour video in real time. The speed of the method was about 20-24 frames per second which meets the requirements of real-time video processing defined in [34]. This way, the method is comparable to the best existing real-time face detectors [14,21,30,33,35,36] in terms of processing speed.

The local oriented edges served as basic features for expression-invariant representation of facial landmarks. The results confirmed that in the majority of expressive images a distribution of the local oriented edges had structure with two horizontal dominants as predefined in [11,12]. This property allowed discarding noisy regions and preserving regions of the landmarks. Thus, the method was able to locate landmarks from images with hair and shoulders. The use of frontal-view geometrical face model further improved the overall performance of the method. As Fig. 8 shows, the method was effective in locating faces with frontal and near-to-frontal head poses. However, head rotations by 30° significantly decreased face localization rates. This is explained by the fact that geometrical constrains from Table 1 were defined mainly for frontal-view geometrical face model. This resulted in the fact that the landmarks were located correctly in case of 30° head rotations, but failed to be composed into face-like constellations. The relaxation of geometrical constrains from Table 1 or development of new measures and constrains for a near-to-profile geometrical face model would improve the performance of the method in case of strong head rotations. It is noteworthy that facial expressions had no statistically significant effect on the detection rates.

In case of frontal and near-to-frontal head positions, the method demonstrated sufficiently high rates in locating faces with all three tested expressions - neutral, frowning, and smiling expressions. This gives similar performance of the method for these particular expressions as compared to the results of previous studies which use similar edge-based approach to facial landmark localization [11,12,31]. In these studies, the edge detector applied was based on a difference between two Gaussian operators with shifted kernels. This enabled rich edge structures to be detected meaning that a large number of edge points were detected by the Gaussian edge detector. It resulted in a situation in which the landmark orientation portrait consisted of a large number of points and provided a detailed description of the edge distribution inside the detected region. In present study, the Sobel edge detector was used providing less detailed information on the edge distribution inside the detected edge region. Because some edge orientations were missing in the landmark orientation portraits in this case, it resulted in fewer local similarity measures for the detected landmark candidates. Despite this, the Sobel edge detector proved to function in real time, which is a critical requirement for many applications.

In this study, we concentrated on face localization rather than on landmark localization meaning that all four facial landmarks needed to be correctly located in order to declare successful face localization. Face localization rate would be improved if we consider two or three correctly located landmarks as necessary and sufficient requirement for successful face localization, as it is done, for example, in [3]. This way, the method can also be applied for independent facial landmark localization, when it is allowed to miss some landmarks.

Further, the proposed detection scheme functioned well for the present work. However, it could fail if the test conditions include, for example, an image of a complex scene. In this case, the eyes might be incorrectly detected from the very beginning. Full occlusions of the eye(s) by hand or hair would also significantly impair the performance of the method. For such conditions, it would be beneficial to use more complex models to capture the spatial arrangement of the landmark candidates detected in the image and not assume in advance the correctness of any particular landmark location. It used subsets of all possible face-like constellations in order to validate the locations of other landmarks. The joint distribution of the landmark coordinates could be modelled with, for example, a single multi-variate Gaussian or a mixture of bi-variate Gaussians [29]. In this case the framework would become more robust with regard to false alarms and searching for missed or occluded landmarks at the expense of also being more computationally complex.

In summary, as compared to the existing feature-based methods of face localization, the method demonstrated similar or superior performance in terms of speed. Besides robustness to facial expressions and small out-of-plane head rotations, the developed method demonstrated robustness to

noise such as hair, and elements of clothing and decoration. Emphasizing simplicity, high speed, and low computation cost of the method, we conclude that it can be used in face localization as such and also in preliminary localization of regions of facial landmarks for their subsequent processing where coarse landmark localization is followed by fine feature detection.

7 Acknowledgement

This work was financially supported by the University of Tampere and Academy of Finland (project numbers 177857 and 1115997).

References:

[1] P. Campadelli, R. Lanzarotti, and G. Lipori, Automatic Facial Feature Extraction for Face Recognition, In *K. Delac and M. Grgic (Eds.), Face Rec.*, Vienna: I-Tech Education and Publishing, pp.31-58, 2007.

[2] H. Chang and U. Robles, Face Detection, *Project report, http://www-cs-students.stanford edu/robles/ee368/main.html*, (last accessed 10 April, 2008).

[3] D. Colbry, G. Stockman, and J. Anil, Detection of Anchor Points for 3D Face Verification, In Proc. *IEEE Comp. Society Int. Conf. on Comp. Vision and Pattern Rec. (CVPR)*, Vol.3, 2005, pp.118-126.

[4] T. Cootes, G. Edwards, and C. Taylor, Active Appearance Models, IEEE *Trans. on Pattern Analysis and Machine Intelligence*, Vol.23, No.6, 2001, pp.681-685.

[5] D. Cristinacce and T. Cootes, Feature Detection and Tracking with Constrained Local Models, In Proc. *British Machine Vision Conf. (BMVC)*, Vol.3, 2006, pp.929-938.

[6] G. Deco and D. Obradovic, *An Information-Theoretic Approach to Neural Computing*, New York: Springer-Verlag, 1996.

[7] P. Ekman and W. Friesen, *Facial Action Coding System (FACS): A Technique for the Measurement of Facial Actio*n, Palo Alto, California: Consulting Psychologists Press, 1978.

[8] L. Farkas, *Anthropometry of the Head and Face*, (2 ed), Raven: New York, 1994.

[9] B. Fröba and C. Küblbeck, Robust Face Detection at Video Frame Rate Based on Edge Orientation Features, In Proc. *IEEE Int. Conf. on Automatic Face and Gesture Rec. (FG)*, 2002, pp.342-347.

[10] T. Gernoth, R. Kricke, R.-R. Grigat, Mouth Localization for Appearance-Based Lip Motion Analysis, *WSEAS Trans. on Signal Processing*, Vol.3, No.3, 2007, pp.275-281.

[11] Y. Gizatdinova and V. Surakka, Feature-Based Detection of Facial Landmarks from Neutral and Expressive Facial Images, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol.28, No.1, 2006, pp.135-139.

[12] Y. Gizatdinova and V. Surakka, Automatic Detection of Facial Landmarks from AU-Coded Expressive Facial Images, In Proc. *Int. Conf. on Image Analysis and Processing (ICIAP)*, 2007, pp.419-424.

[13] R. Gonzalez and R. Woods, *Digital Image Processing*, Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 2001.

[14] R. Gottumukkal and V. Asari, Real Time Face Detection from Color Video Stream Based on PCA Method, In Proc. *Applied Imagery Pattern Rec. Workshop (AIPR)*, 2003, pp.146-150.

[15] B. Heisele, T. Poggio, and M. Pontil, Face Detection in Still Gray images, *A.I. memo AIM-1687, Artificial Intelligence Laboratory, MIT*, 2000.

[16] B. Heisele, T. Serre, M. Pontil, and T. Poggio, Component-Based Face Detection, In Proc. *IEEE Comp. Society Conf. on Comp. Vision and Pattern Rec.(CVPR'01)*, Vol.1, 2001, pp.657-662.

[17] E. Hjelmas and B. Low, Face Detection: A Survey, *Comp. Vision and Image Understanding*, Vol.83, 2001, pp.235-274.

[18] *Hypermedia Image Processing Lab (HIPR2)*, homepages.inf.ed.ac.uk/rbf/HIPR2/hipr_top.htm, (last accessed 10 April, 2008).

[19] T. Kawaguchi and M. Rizon, Iris Detection Using Intensity and Edge Information. *Pattern Recognition*, Vol.36, 2003, pp.549–562.

[20] J.-G. Ko, J.-H. Yu, K.-I. Jeong, Facial Feature Detection and Head Orientation based Gaze Tracking, *WSEAS Trans. on Signal Processing*, Vol.2, No.1, 2005, pp.209-215.

[21] S. Lee, S. Yoon, J. Jeong, The Effective Combination of Color and Motion Information for Tracking Face and Hands in Complex Background, *WSEAS Trans. on Signal Processing*, Vol.12, No.2, 2006, pp.1531-1536.

[22] J. Martinkauppi, M. Soriano, and M. Laaksonen, Behavior of Skin Color under Varying Illumination Seen by Dierent Cameras at Dierent Color Spaces, *Machine Vision in Industrial Inspection*, V.9, No.4301, 2001, pp.102-113.

[23] P. Michel and R. Kaliouby, Real Time Facial Expression Recognition in Video Using Support Vector Machines, *International conference on multimodal interfaces (ICMI)*, 2003, pp.258-264.
[24] M. Pantic and J. Rothkrantz, Automatic Analysis of Facial Expressions: The State of the Art. *IEEE Trans. on Pattern Analysis and Machine Intelligent*, Vol.22, No.12, 2000, pp.1424–1445.
[25] V. Perlibakas, Automatical Detection of Face Features and Exact Face Contour. *Pattern Recognition Letters*, Vol.24, No.16, 2003, pp.2977–2985.
[26] Y. Rodriguez, F. Cardinaux, S. Bengio, and J. Mariéthoz, Measuring the Performance of Face Localization Systems, *J. of Image and Vision Computing*, Vol.24, 2006, pp.882-893.
[27] H. Rowley, S. Baluja, and T. Kanade, Neural Network-Based Face Detection, *IEEE Trans. on Pattern Analysis and Machine Intelligent*, Vol.20, No.1, 1998, pp.23-38.
[28] M. Saaidia, S. Lelandais, V. Vigneron, E.-M. Bedda, Face Detection by Neural Network Trained With Zernike Moments**,** In Proc. *WSEAS Int. Conf. on Signal Processing, Robotics and Automation (ISPRA)*, 2007, pp.36-41.
[29] A. Salah, H. Çınar, L. Akarun, and B. Sankur, Robust Facial Landmarking for Registration. *Annals of Telecommunications*, Vol.62, No.1-2, 2007, pp.1608-1633.
[30] S. Sclaroff and J. Isidoro, Active Blobs: Region-Based, Deformable Appearance Models, *Comp. Vision and Image Understanding*, Vol.89, No.2–3, 2003, pp.197-225.
[31] D. Shaposhnikov, L. Podladchikova, and X. Gao, Classification of Images on the Basis of the Properties of Informative Regions, *Pattern Rec. and Image Analysis*, Vol.13, No.2, 2003, pp.349-352.
[32] P. Sinha, B. Balas, Y. Ostrovsky, and R. Russell, Face Recognition by Humans: Nineteen Results All Computer Vision Researchers Should Know About. In Proc. *IEEE*, Vol.94, No.11, 2006, pp.948-1962.
[33] Y. Tong, Y. Wang, Zh. Zhu, and Q. Ji, Robust Facial Feature Tracking Under Varying Face Pose and Facial Expression. *Pattern Recognition*, Vol.40, No.11, 2007, pp.3195-3208.
[34] M. Turk and M. Kölsch, Perceptual Interfaces, *In G. Medioni and S.B. Kang, (Eds), Emerging Topics in Computer Vision*, Prentice Hall, chapter 9, 2004, 45 p.
[35] P. Viola and M. Jones. 2004. Robust Real-Time Face Detection, *Int. J. of Comp. Vision*, Vol.57, No.2, pp.137–154.
[36] S. Vural, H. Yamauchi, Facial Analysis and Authentication for High-Speed Access Gates, *WSEAS Trans. on Signal Processing*, Vol.8, No.2, 2006, pp.1046-1052.
[37] L. Wiskott, J. Fellous, N. Krüger, and C. der Malsburg, Face Recognition by Elastic Bunch Graph Matching, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol.19, No.7, 1997, pp.775-779.
[38] M. Yang, N. Abuja, and D. Kriegman, Face Detection Using Mixtures of Linear Subspaces, In Proc. *Int. Conf. on Automatic Face and Gesture Rec. (FG)*, 2000, pp.70-76.
[39] M. Yang, D. Kriegman, and N. Ahuaja, Detecting Face in Images: A Survey, *IEEE Trans. on Pattern Analysis and Image Understanding*, Vol.24, 2002, pp.34-58.
[40] D. Xi and S. Lee, Face Detection and Facial Component Extraction by Wavelet Decomposition and Support Vector Machines, In Proc. *Audio- And Video-Based Biometric Person Authentication (AVBPA)*, 2003, pp.199-207.