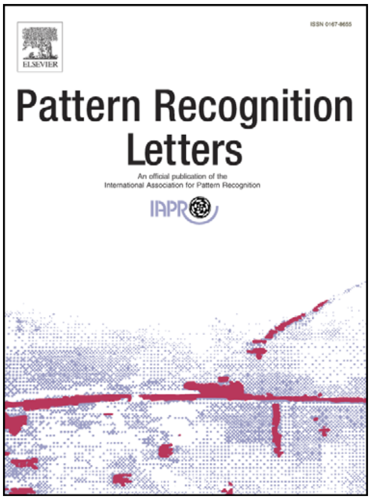
Post-print

Authors:	Gizatdinova Yulia, Surakka Veikko
Name of article:	Automatic edge-based localization of facial features from images with complex facial expressions
Year of publication:	2010
Name of journal:	Pattern Recognition Letters
Volume:	31
Number of issue:	15
Pages:	2436-2446
ISSN:	0167-8655
Discipline:	Natural sciences / Computer and information sciences
Language:	en

URN: http://urn.fi/urn:nbn:uta-3-987
DOI: http://dx.doi.org/10.1016/j.patrec.2010.07.020


ISSN 0167-8655
ELSEVIER
Pattern Recognition Letters
An official publication of the International Association for Pattern Recognition
IAPR

Automatic Edge-Based Localization of Facial Features from Images with Complex Facial Expressions

Yulia Gizatdinova, Veikko Surakka

PII: S0167-8655(10)00244-8
DOI: 10.1016/j.patrec.2010.07.020
Reference: PATREC 4932

To appear in: *Pattern Recognition Letters*

Received Date: 2 July 2009

Automatic Edge-Based Localization of Facial Features from Images with Complex Facial Expressions

Yulia GIZATDINOVA* and Veikko SURAKKA

Research Group for Emotions, Sociality, and Computing
Department of Computer Sciences
University of Tampere
Kanslerinrinne 1, 33014, Finland
{yulia.gizatdinova, veikko.surakka}@cs.uta.fi
**Corresponding author: yulia.gizatdinova@cs.uta.fi, tel: +358 (0)3 3551 4030,*
fax: +358 (0)3 3551 6070

***Abstract* –** Automatic localization of facial features is an essential step for many systems of face recognition, facial expression classification and intelligent vision-based human-computer interfaces. In this paper, we present automatic edge-based method of locating regions of prominent facial features from up-right facial images. The proposed localization scheme was tested on several public databases of complex facial expressions. The method demonstrated high localization rates when localization accuracy was evaluated by both a conventional point error measure and a new rectangular error measure that takes into account the location of the feature in the image and the true feature size.

***Classification codes* –** Image and signal processing, Image segmentation, Edge/line detection, Biometrics, Face and facial feature recognition.

***Keywords* –** Facial feature localization, local oriented edge, edge histogram, localization error, accuracy evaluation metrics, expression, action unit.

1. INTRODUCTION

Automatic localization of facial features is the first necessary step for many systems of face recognition, facial expression classification and intelligent vision-based human-computer interfaces. The localization of facial features is defined as a process of finding the true locations of prominent facial features (e.g. eyes, brows, nose, mouth, chin, etc.), given a facial image of a sole person. After the facial features have been located, the scale and position of the face in the image are also known. The description of some well known localization methods can be found in the survey papers by Hjelmas and Low (2001), Pantic and Rothkrantz (2000) and Yang et al. (2002). Typically, facial feature localization employs a reduction of the search space by first locating feature regions of interest (ROI). Following this, fine analysis of the located ROIs is performed in order to find characteristic feature points (e.g. mouth corners, eye centres, etc.). The latter is usually done by modelling texture information (Jesorsky et al., 2001; Vukadinovic and Pantic, 2005), shape information (Campadelli et al., 2005; Venkatesh at al., 2009), or combination of those (Campadelli et al., 2007; Cristinacce and Cootes, 2006). It has been shown (Lien et al., 2000; Tian et al., 2002; Tong et al., 2007) that facial expressions significantly deteriorate the performance of feature localization methods. Our work focuses on the development of effective strategies for feature ROI localization from images with complex facial expressions. Facial expressions originate from non-rigid facial movements which change feature shape and location in the face and may also result in the out-of-plane facial changes (e.g. showing the tongue). The localization of eyes and mouth is especially difficult, since these features are highly deformable and are subject to self-occlusion during expressive reactions. Facial expressions are frequently classified in terms of emotion-associated categories. The prototypical displays of neutral, happy, sad, fearful, angry, disgusted and surprised expressions have been included in the Pictures of Facial Affect (POFA) (Ekman and Friesen, 1976) and Japanese Female Facial Expression (JAFFE) (Lyons et al., 1998) databases. The Facial Action Coding System (FACS) (Ekman et al., 2002) is another way of classifying facial expressions without a direct reference to human emotions. The FACS defines action units (AU) as muscular activity that produces momentary changes in facial appearance. AU codes have been used in creation of the well known Cohn-Kanade AU-Coded Facial Expression (Cohn-Kanade) database (Kanade et al., 2000).

In order to locate feature ROIs from the facial image, a majority of the methods apply low-level image descriptors like colours (Cooray and O'Connor, 2004), edges (Song et al., 2006) and raw pixel intensities (Vukadinovic and Pantic, 2005). Another popular technique of feature ROI localization applies the learning-based methods such as support vector machines (Heisele et al., 2007), cascades of boosted classifiers (Wilson and Fernandez, 2006) and neural network (Campadelli et al., 2005). Wavelet image decomposition has been widely used for feature ROI localization when combined with neural networks (Feris et al., 2002) and support vector machines (Campadelli et al., 2007). In many cases, face detection is done to facilitate the feature ROI localization. Despite the fact that feature ROI localization has seen a lot of progress, the improvement of the existing localization schemes in terms of their accuracy, speed and robustness is still required in order to achieve true applicability of these methods in real-world situations. The majority of the feature ROI localization methods have demonstrated high performance when tested on databases of relatively expressionless faces, thus leaving out the question of expression invariance. We believe that more detailed and systematic approach to the development of expression-invariant feature ROI localizers is still required. The proposed approach aims at revealing specific facial behaviours which attenuate the performance of feature localizers and proposing effective strategies of eliminating the deteriorating effect of these behaviours.

Recently, edge-based method of locating ROIs of prominent facial features from up-right facial images has been introduced (Gizatdinova and Surakka, 2006). In the method, local oriented edges have been utilized in composing edge maps of the image at several levels of resolution. Facial feature candidates, which resulted from the step of edge map construction, have been further verified by edge orientation matching. The method has demonstrated promising results on the POFA database (Gizatdinova and Surakka, 2006). However, the following tests on the Cohn-Kanade database (Gizatdinova and Surakka, 2008) have revealed that feature ROI localization has been significantly deteriorated by a number of specific AUs. The detailed analysis has showed that nose and mouth ROI localization has been especially affected by the lower face AU 9 (nose wrinkler), AU 10 (upper lip raiser), AU 11 (nasolabial furrow deepener) and AU 12 (lip corner puller). All these AUs have frequently been associated with expressions of happiness and disgust. These AUs, when appeared alone or in combinations, have caused erroneous merging of nose and mouth ROIs into one region at the stage of edge map construction. Similarly, erroneous merging of eye and eyebrow ROIs has been frequently caused by the upper face AU 4 (brow lowerer), AU 6 (cheek raiser and lid compressor) and AU 7 (lid tightener) which are associated with expressions of anger, disgust, sadness and happiness. Taking into account the revealed facial behaviours, a number of improvements have been introduced to the method. Preliminary tests of the improved method with the Cohn-Kanade database (Gizatdinova and Surakka, 2007) have showed a significant improvement of feature ROI localization from images with AUs 9, 10, 11 and 12. The aim of the present study is to test the improved method on a wider range of facial expressions from the POFA, Cohn-Kanade and JAFFE databases. Another issue that we consider in this paper is an accuracy (or precision) evaluation of the feature ROI localization methods. Thus, a new error measure that takes into account both the feature location in the image and the true feature size is introduced.

The reminder of the paper is organized as follows. The method of feature ROI localization is described in Section 2. A new rectangular error measure is introduced in Section 3. Section 4 reveals the performance of the ROI localization when evaluated by the convenient and new error measures. Section 5 discusses the results and finally Section 6 concludes the paper.

2. EDGE-BASED FACIAL FEATURE LOCALIZATION

The method of facial feature ROI localization is illustrated in Figure 1. Theoretical implications of the stages of image pre-processing and feature ROI formation (Figures 2b-d) have been introduced in Gizatdinova and Surakka (2006) and the stage of face candidate formation (Figures 2e,f) have been briefly described in Gizatdinova and Surakka (2007). Practical implementations of these algorithms are given below.

The image is considered as a two dimensional array $I = \{b_{ij}\}$ of pixel size $X \times Y$. Each element b_{ij} of the array represents brightness *b* of the image pixel at location (i, j). On the pre-processing stage, the image is transformed to a grey-scale representation by summing three weighted colour components:

$$b_{ij} = 0.299 \cdot R_{ij} + 0.587 \cdot G_{ij} + 0.114 \cdot B_{ij} \tag{1}$$

where the weights assigned to colour components are normalized in the range [0,1] and exploit natural phenomenon of how human eye respond to light of different colour (Hurvich, 1981).

Further, the image is smoothed by a Gaussian operator (see also Algorithms 1.1 and 1.2):

$$b_{ij}^{(l)} = \sum_{p,q} a_{pq} b_{ij}^{l-1}, \quad b_{ij}^{(1)} = b_{ij} \tag{2}$$

where a_{pq} is a coefficient of the Gaussian convolution; *p* and *q* define the size of the smoothing filter; *i* and *j* denote a current pixel location ($i = 0 \div (X-1)$, $j = 0 \div (Y-1)$); $X \times Y$ is the size of the image in pixels; *l* defines the level of image smoothing (or resolution). The received smoothed image is referred to as low-resolution image and is used to explore all possible feature ROI candidates. The original non-smoothed image is referred to as high-resolution image and is utilized to analyze feature ROI candidates in detail.

Algorithm 1.1: Construction of the smoothing kernel*

Require: $Kernel$ is a kernel of the smoothing filter of size $K \times K$, $Mask$ is an array of size $K \times K$ that contains image area to be convolved with the smoothing filter.

```
$K = 5$, $\sigma = 5/6$, $Pi = 3.14159265358979323846$
for $i = 0$ to $K$ do { Iterate through each element of the kernel column }
    $d1 = (i\text{-}K/2)^2$ { Calculate coefficient 1 }
    for $j = 0$ to $K$ do { Iterate through each element of the kernel raw }
        $d2 = (j\text{-}K/2)^2 + d1$ { Calculate coefficient 2 }
        $Kernel[i \cdot K+j] = [1/(2 \cdot Pi \cdot \sigma^2)] \cdot \exp[-d2/(2 \cdot \sigma^2)]$ { Fill in the smoothing kernel matrix }
    end for
end for
```

* Runs only once.

Algorithm 1.2: Convolution of the image with the smoothing kernel

Require: A gray scale image $I = \{b_{ij}\}$ of pixel size $X \times Y$, b_{ij} represents brightness b of the image pixel at location (i, j). $Kernel$ is a kernel of the smoothing filter of size $K \times K$, $Mask$ is an array of size $K \times K$ that contains image area to be convolved with the smoothing filter , $ResLevel$ is the desired level of image smoothing.

```
$K = 5$, $l = 0$, $ResLevel = 2$
while $l < ResLevel$ { Iterate $l$ from 0 until $l$ reaches the desired level of smoothing }
    for $j = 0$ to $Y$ do { Iterate through each pixel of the image column }
        for $i = 0$ to $X$ do { Iterate through each pixel of the image row }
            if ( $j-2 \geq 0$ AND $j+2 < Y$ AND $i-2 \geq 0$ AND $i+2 < X$ ) then { If the pixel to be processed is
            inside of the processing area (i.e. the filter does not exceed the image dimensions) }
                for $p = -2$ to $2$ do { Iterate through each pixel of the mask column }
                    for $q = -2$ to $2$ do { Iterate through each pixel of the mask row }
                        $Mask[(p+2) \cdot K+q+2] = I[(j+p) \cdot X+i+q]$ { Fill in the mask }
                    end for
                end for
                $I[j \cdot X+i] = convolve(Mask, Kernel)$ { Convolve the mask with smoothing kernel }
            else { In all other cases }
                $I[j \cdot X+i] = I[j \cdot X+i]$ { Copy intensity values without a change }
            end if
        end for
    end for
    $l = l + 1$ { Increase the level of smoothing }
end while
```

2.1. FEATURE ROI FORMATION

EDGE DETECTION AND EDGE MAP CONSTRUCTION

The edge detection module simulates the mechanisms of pre-attentive edge detection in the visual cortex of the human brain. As it has been demonstrated (Marr, 1982), the neurons of the primary visual cortex have a remarkable property of orientation selectivity. This property provides the detection of local oriented edge and the definition of edge orientation. According to the concept of columnar organization (Hubel & Wiesel, 1962), the neighbouring neurons in the primary visual cortex have similar orientation selectiveness. Together they form an orientation column or iso-orientation domain. A set of orientation columns with common receptive field forms a module of the cortex called a hypercolumn. A schematic representation of the neuronal hypercolumn organization is shown in Figure 3a. The hypercolumn neurons of different orientation selectivity are represented by a set of convolution kernels which result from differences of two oriented Gaussians with shifted centres:

$$G_{\varphi_k} = {}^{1}\!/_{Z} \cdot (G^{-}_{\varphi_k} - G^{+}_{\varphi_k}) \tag{3}$$

$$G^{-}_{\varphi_k} = {}^{1}\!/_{2\pi\sigma^2} \cdot \exp\left[-\left(\left((p-\sigma\cos\varphi_k)^2 + (q-\sigma\sin\varphi_k)^2 \right) \Big/ 2\sigma^2 \right) \right] \tag{4}$$

$$G^{+}_{\varphi_k} = {}^{1}\!/_{2\pi\sigma^2} \exp\left[-\left(\left((p+\sigma\cos\varphi_k)^2 + (q+\sigma\sin\varphi_k)^2 \right) \Big/ 2\sigma^2 \right) \right] \tag{5}$$

$$Z = \Sigma (G^{-}_{\varphi_k} - G^{+}_{\varphi_k}), \quad G^{-}_{\varphi_k} - G^{+}_{\varphi_k} > 0 \tag{6}$$

where $\sigma = 1.2$ is a root mean square deviation of the Gaussian distribution; φ_k is the angle of the Gaussian rotation, $\varphi_k = k \cdot 22.5°$; $k = 0 \div 15$; p and q denote 7×7 size of the filter $p,q = -3 \div 3$; $i = 0 \div (X-1)$; $j = 0 \div (Y-1)$. It has been demonstrated (Gizatdinova and Surakka, 2006) that ten orientations depicted in Figure 3b (i.e. $k = 2 \div 6$ and $k = 10 \div 14$) are sufficient for representation of prominent facial features in frontal faces.

The maximum response of all ten kernels defines a contrast magnitude of the local edge at its pixel location. The orientation of the local edge is estimated by the orientation of the kernel that gives the maximum response (see also Algorithms 2.1 and 2.2):

$$g_{ij\varphi_k} = \sum_{p,q} b^{(l)}_{i-p,\,j-q} G_{\varphi_k} \tag{7}$$

Algorithm 2.1: Construction of the set of convolution kernels[*]

Require: A set of convolution kernels $Kernels = \{Kernel_k\}$. Each kernel $Kernel_k$ from the set $Kernels$ is of size $K \times K$ and is used to detect one of 16 orientations of the local edge which are defined by $k = 0 \div 15$[**].

```
$\varphi_k = 0$, $K = 7$, $\sigma = 7/6$, $Pi = 3.14159265358979323846$
for $k = 0$ to 16 do { Iterate through all orientations }
    $\varphi_k = (k \cdot 2 \cdot Pi)/16$ { Calculate the angle of Gaussian rotation }
    $sum_k = 0$ { Temporary variable }
    for $p = 0$ to $K$ do { Iterate through each element of the kernel column }
        $y = p - K/2$ { Temporary variable }
        for $q = 0$ to $K$ do { Iterate through each element of the kernel row }
            $x = q - K/2$ { Temporary variable }
            $R1 = (x - \sigma \cdot \cos\varphi_k)^2 + (y - \sigma \cdot \sin\varphi_k)^2$ { Calculate coefficient 1 }
            $R2 = (x + \sigma \cdot \cos\varphi_k)^2 + (y + \sigma \cdot \sin\varphi_k)^2$ { Calculate coefficient 2 }
            $Kernel_k[p \cdot K+q] = 1/(2 \cdot Pi \cdot \sigma^2) \cdot [\exp(-R1/(2 \cdot \sigma^2)) - \exp(-R2/(2 \cdot \sigma^2))]$ { Fill in the kernel matrix }
            if $Kernel_k[p \cdot K+q] > 0$ then $sum_k = sum_k + Kernel_k[p \cdot K+q]$
            end if
        end for
    end for
    for $p = 0$ to $K$ do { Iterate through each element of the kernel column }
        for $q = 0$ to $K$ do { Iterate through each element of the kernel row }
            $Kernel_k[p \cdot K+q] = Kernel_k[p \cdot K+q]/sum_k$ { Normalize elements of the kernel }
        end for
    end for
end for
```

[*] Runs only once.

[**] Ten orientations defined by $k = 2 \div 6$ and $k = 10 \div 14$ can be used.

Algorithm 2.2: Convolution of the image with a set of convolution kernels

Require: A gray scale image $I = \{b_{ij}\}$ of pixel size $X \times Y$, b_{ij} represents brightness b of the image pixel at location (i, j). A set of convolution kernels $Kernels = \{Kernel_k\}$. Each kernel $Kernel_k$ from the set $Kernels$ is of size $K \times K$, $Mask$ is an array of size $K \times K$ that contains image area to be convolved with convolution kernels, $g = \{g_{ij}\}$ is an array of edge magnitudes of size $X \times Y$ and $Orientations = \{Orient_{ij}\}$ is an array of orientations of size $X \times Y$, 16 orientations of the local edge are defined by $k = 0 \div 15$[*].

```
for $j = 0$ to $Y$ do { Iterate through each pixel of the image column }
    for $i = 0$ to $X$ do { Iterate through each pixel of the image row }
        if ( $j - 3 \geq 0$ AND $j + 3 < Y$ AND $i - 3 \geq 0$ AND $i + 3 < X$ ) then { If the pixel to be processed is inside
        of the processing area (i.e. the filter does not exceed the image dimensions) }
            for $p = -3$ to 3 do { Iterate through each pixel of the mask column }
                for $q = -3$ to 3 do { Iterate through each pixel of the mask row }
                    $Mask[(p+3) \cdot K+q+3] = I[(j+p) \cdot X+i+q]$ { Fill in the mask }
                end for
            end for
            $Res_k = 0$, $Orient_k = fake$ { Temporary variables }
            for $k = 0$ to 16 do { Iterate through every orientation }
                $\mathrm{Re}s_k = Convolve(Mask, Kernel_k)$ { Convolve mask with each kernel }
            end for
            $g(j \cdot X + i) = \max(\{\mathrm{Re}s_k\})$ { Select the kernel that gave the maximum response }
            $Orient(j \cdot X + i) = k$   { Select the orientation of the kernel that gave the maximum response }
        **else** { In all other cases }
            $g(j \cdot X + i) = 0$   { No response }
            $Orient(j \cdot X + i) = fake$   { No orientation }
        **end if**
    **end for**
**end for**
```

*Ten orientations defined by $k = 2 \div 6$ and $k = 10 \div 14$ can be used.

After the smoothed low-resolution image is filtered with a set of ten convolution kernels, the extracted local oriented edges are thresholded and grouped into edge regions which represent candidates for feature ROIs. The average contrast of the whole smoothed low-resolution image is used for contrast thresholding. Edge grouping is based on the neighbourhood distances between edge points and limited by a maximum number of edges in the region. Empirical testing has defined optimal thresholds for edge grouping. To get a more detailed description of the located feature ROIs, edge detection and edge grouping are applied to the original high-resolution image within the limits of the located regions. In this case, the threshold for contrast filtering is determined by doubling the average contrast of the high-resolution image.

EDGE HISTOGRAM MATCHING

The existence of facial feature in the image is verified by applying a set of rules which define a specific distribution of local oriented edges inside the located ROIs. The rules define edge histogram that has two dominants corresponding to horizontal orientations (Gizatdinova and Surakka, 2006): 1) the horizontal orientations are represented by the greatest number of extracted edges; 2) a number of edges corresponding to each of horizontal orientations is more than 50 percent greater than a number of edges corresponding to other orientations taken separately; and 3) the histogram cannot have zero number of edges of any orientation. Unlike feature ROIs, noisy edge regions such as shadows, elements of clothing, hair and decoration have an arbitrary distribution of local oriented edges and, therefore, can be filtered out. A relaxation of the rules 2-3 has been introduced (Gizatdinova and Surakka, 2007). If edge histogram fully corresponds to the original rules, the corresponding ROI is labelled as a primary candidate. If edge histogram satisfies to the relaxed criterion, the corresponding ROI is labelled as a secondary candidate. In further analysis, secondary candidates are considered in composing face-like constellations if there are missing features. The steps of the edge histogram matching are summarized in Algorithm 3.

Algorithm 3: Edge histogram matching

Require: A set of edge histograms $H = \{H_i\}$, where $i = 0 \div M$, M is a number of ROIs (regions of connected edges received from the edge map construction stage) and $|H| > 0$. Each histogram H_i from the set H represents a distribution of local oriented edges in a given ROI. For each histogram H_i, the amounts of local edges corresponding to 10 orientation constitute a set $N_i = \{N_{i_k}\}$, where 10 orientations are defined by $k = 2 \div 6$ and $k = 10 \div 14$.

```
for $i = 0$ to $M$ do { Iterate over all histograms from the set $H$ }
    if $N_{i4} = 0$ $OR$ $N_{i12} = 0$ then { If the number of local edges corresponding to at least one of the horizontal orientations equals to zero }
        return* 0 { The ROI that corresponds to a given histogram $H_i$ is discarded as noise }
    else
        $N_{i\max} = \max(\{N_{i_k}\})$, $k = 2,3,5,6,10,11,13,14$ { Find a non-horizontal orientation with the maximum number of local edges }
        if ( $N_{i4} < N_{i\max}$ $OR$ $N_{i12} < N_{i\max}$ ) then { If the number of local edges corresponding to at least one of the horizontal orientations is less than the maximum number of non-horizontal local edges }
            return 0 { The ROI that corresponds to a given histogram $H_i$ is discarded as noise }
        else if ( $N_{i4} \geq 2 \cdot N_{i\max}$ $AND$ $N_{i12} \geq 2 \cdot N_{i\max}$ $AND$ $N_{i_k} \neq 0$ ) then { If the number of edges corresponding to each horizontal orientation is more than 50 percent greater than the maximum number of non-horizontal edges and if all elements from the set $N_i$ are not zeros }
            return 1 { The ROI that corresponds to a given histogram $H_i$ is marked as a primary feature ROI }
        else {In all other cases}
            return 2 { The ROI that corresponds to a given histogram $H_i$ is marked as a secondary feature ROI }
        end if
    end if
end for
```

[*] **return** labels the ROI either as a feature ROI or as a noise and makes a transfer to the next histogram analysis.

2.2. FACE CANDIDATE FORMATION

EDGE PROJECTION

This stage verifies and corrects the formation of feature ROIs. The earlier method has failed at the step of edge map construction due to erroneous connection of edges belonging to neighbouring features into one region. In order to separate the merged feature ROIs, a simple but effective procedure of x/y-edge projection is proposed. If there is a merged feature, edge points are projected to x-axis for eye regions or y-axis for nose and mouth. Projections are obtained from calculating a number of edge points along the corresponding columns or rows of the feature edge map. If the number of projected edge points is smaller than a threshold, edge points are eliminated. After each elimination step, if the region still is not separated, the threshold is increased by 5 edge points. The initial threshold equals to a minimum number of edges in the column or row of the given feature ROI. The elimination iterates until the ROI is separated or until all edge points are eliminated. The procedure of x/y-edge projection is initiated on the step of structural correction, in which a prediction about feature merging is made.

STRUCTURAL CORRECTION

The geometry of a typical up-right face has been widely utilized in order to find a proper spatial arrangement of facial features (Campadelli at al., 2007; Cooray and O'Connor, 2004; Vukadinovic and Pantic, 2005). It has been shown (Ekman et al., 2002) that the interocular distance *d* is a facial measure that is not affected by facial expressions. Because of this property *d* can be used as a relative measure of distances among other features. This way, facial measures *d1*, *d2* and *d3* from Figure 4 can be represented as portions of *d* (Farkas, 1994). However, *d1*, *d2* and *d3* are all affected by facial expressions (Ekman et al., 2002). Therefore, the values for these measures $d1 \in [0.1\text{d}, 0.6\text{d}]$, $d2 \in [0.5\text{d}, 0.7\text{d}]$ and $d3 \in [1.2\text{d}, 1.7\text{d}]$ have been acquired from a representative image set of varying facial expressions. Because *d* depends on the size of the facial region, this measure has been also defined as interval between its minimum and maximum values for each database: $d \in [35,85]$ for Cohn-Kanade, $d \in [60,90]$ for POFA and $d \in [55,80]$ for JAFFE databases. The flowchart of the algorithm for structural correction is depicted in Figure 5. Structural correction classifies the upper face feature ROIs either as eye and eyebrow or as eye region (which includes eye and eyebrow). The lower face feature ROIs are classified as nose and mouth. After the face-like arrangement of the features has been found, the location of the face in the image is also known. Although not implemented in this study, a subdivision between eye and eyebrow in eye region ROIs can be obtained by applying y-projection to the resulting region. Another possibility is to calculate the vertical derivative of the combined ROI. The highest values will correspond to the eyebrow ROI. As it has been reported in (Campadelly et al., 2005), this technique allows splitting the combined ROI into two parts, one corresponding to the eye and another to the eyebrow.

3. EVALUATION OF LOCALIZATION ACCURACY

3.1. REFERENCE ANNOTATION

In order to define the size and location of facial features in the image, a set of characteristic feature points have been selected as follows. Thus, four points which describe eyes and mouth are the most right, left, top and bottom points of these features. The eyebrow is described by its top, bottom-left and bottom-right points. A box that bounds the feature is constructed from these points for eyes, eyebrows, eye regions and mouth. A centre of the feature is then defined as a centre point of the bounding box (Farkas, 1994). Nose is described by the centre point of the nose tip. The top-left, bottom-right and centre points of the bounding box represent a set of attributes that characterize well the size and location of the feature ROI in the image from both informational and evaluation point of view.

The accuracy of localization is commonly evaluated by comparing the results of automatic localization against reference feature locations annotated by a human (Jesorsky et al., 2001; Rodriguez et al., 2006). For this, the selected feature points have been annotated in all databases. Ideally, this should be repeated multiple times by different annotators in order to eliminate errors and subjective decisions from the reference data. Because of tedious character of this work and a large amount of images to be marked, the additional annotation by multiple annotators has been performed on a smaller dataset. 9 participants have annotated 25 images which well reflected variations in facial expressions. The annotation results among the participants revealed the average localization error of 2.14 pixels and standard deviation of 0.5 pixels in feature point annotation. This means that 95% of participants (assuming normal distribution) marked feature points within a circle of maximum $2.14 + 2 \cdot 0.5 = 3$ pixel diameter. This value estimates the inconsistency in human annotation and has been used to correct the singly-annotated ground truth data.

3.2. LOCALIZATION ERROR MEASURE

Figure 6 demonstrates the final result of the feature ROI localization. The size and location of the facial feature in the image are calculated from the top, left, right and bottom coordinates of the corresponding ROI. Mass centre of the located ROI indicates an estimate for the feature centre. A new rectangular error measure R has been designed in order to evaluate the accuracy of localization for eye, eye region and mouth ROIs:

$$\max(d(p_{tl}, \overline{p}_{tl}), d(p_{br}, \overline{p}_{br}) \leq R \qquad (8)$$

where p_{tl} defines the coordinates of the top-left and p_{br} defines the coordinates of the bottom-right boundaries of the reference feature location; $\overline{p}_{tl}$ and $\overline{p}_{br}$ define the coordinates of the automatically located feature positions; $d(p_{tl}, \overline{p}_{tl})$ and $d(p_{br}, \overline{p}_{br})$ are Euclidian pixel distances. If $\overline{p}_{tl}$ (or $\overline{p}_{br}$) is found inside the box that bounds the annotated feature, it is compulsory that it is located in the top-left (or bottom-right) quadrant of the bounding box. $R = 0,1,2,...$ is a real number of pixels that sets a desirable accuracy of localization. For a given R, the result of the feature localization is considered correct, if the feature location satisfies to the criterion from Equation 8. For each feature, the average localization rate is defined as a ratio between the total number of correctly located features (for a given R) and the number of images used in testing.

It can be argued that the accuracy of localization is ultimately defined by the application at hand. The application that will use the output of the localization method may require highly accurate feature ROI localization which leads to smaller values of R. Other applications may work with somewhat relaxed criteria. For this purpose in the next section we present the rates of localization calculated for different R. It is intuitively understandable that localization rates for small values of R will be relatively low, increasing with relaxation of the criterion. The rapidity of the increase of the localization rates can be considered as another performance characteristic of the localization method.

The proposed rectangular measure sets a strict requirement for the localization method to bound feature ROI as close to the borders of the feature as possible. This is different from the criterion of point error measure (Jesorsky et al., 2001; Rodriguez et al., 2006) that only checks if the centre of the feature has been located close enough to the ground truth location:

$$\max(d(p_0, \overline{p}_0), d(p_0, \overline{p}_0) \le R_0 \tag{9}$$

where p_0 and $\overline{p}_0$ define the coordinates of the centre of the reference and the automatically located feature location, respectively; $d(p_0, \overline{p}_0)$ is Euclidian pixel distance; R_0 is a real number that defines a desirable accuracy of localization. In order to make the evaluation criterion be invariant to the size of the image, the error measure is usually normalized by the interocular distance *d* (Jesorsky et al., 2001). In evaluation by the point error measure it has become a standard to report results for $R_0 = 0.25 \cdot d$ that approximates a distance of the half an eye width. Because many studies on facial feature localization have adopted the point evaluation criterion, we also have used it in order to compare our results to those from the literature.

4. RESULTS

The performance of the method has been examined on three public databases of facial expressions: the Cohn-Kanade AU-Coded Facial Expression (Cohn-Kanade) (Kanade et al., 2000), the Pictures of Facial Affect (POFA) (Ekman and Friesen, 1976) and the Japanese Female Facial Expression (JAFFE) (Lyons et al., 1998) databases. For each subject from the AU-coded Cohn-Kanade database, one neutral face and several expressive faces of the highest intensity have been selected. In sum, a total of 97 neutral and 486 expressive images have been selected. From this data the datasets of cropped images with face, hair and sometimes shoulders included (with cropped plane background and image indexes) have been composed. The POFA database consists of 14 neutral and 96 expressive images of 14 Caucasian individuals (57% female). On average, there are 16 images per facial expression. JAFFE database consists of 30 neutral and 176 expressive images of 10 Japanese females. There are about 30 images per facial expression in average. All the images from all databases have been preset to 200 by approximately 300 pixel arrays. No face alignment has been performed. The complexity of expressions has been represented by a variability of deformations in soft tissues (wrinkles and protrusions), variety of mouth appearances including open and tight mouth and self occlusions (semi- and closed eyes, bitted lips, visibility of teeth and tongue).

The method has been able to locate feature ROIs, whose shape and size in the image varied significantly with changes in facial expressions. Figure 6 shows that mouth ROI has been located no matter whether the mouth is open or closed and whether the tongue is visible or not. Eye ROI has been located no matter whether the whites of the eyes are visible or not.

Figure 7 demonstrates the average rates of feature ROI localization when the localization accuracy has been evaluated by the proposed rectangular error measure R. The performance of the method with localization accuracy evaluated by the conventional point error measure R_0 is shown in Figure 8. In the figures, x-axis shows the localization error measure R and y-axis shows the average rates of feature ROI localization calculated for a given R. As described above, the single-annotated reference data includes 3 pixel error of feature point localization. It follows that in the evaluation, a highly accurate localization (that is ideally close to human marking) is considered if the reference and the automatically located feature points are found inside of a circle of 3 pixel diameter ($R = 3$). Presenting the results in Figures 7 and 8, we assume that the ground truth annotation data may contain the localization error of maximum 3 pixels. For example, for POFA expressive dataset the value $R = 12$ defines the average feature localization rates calculated for those feature locations which satisfied to the criteria of Equations 8 and 9 (i.e. feature points have been found within a distance of $R = 12 \pm 3$ pixels from the reference feature points). To facilitate the comparison of our results to those reported in the literature, Figures 7 and 8 show the localization rates for different R and R_0, both calculated as portions of d ($R = 0.2d$ and $R = 0.25d$).

The figures show average localization rates separately for expressive (top row) and neutral (bottom row) datasets. To facilitate the analysis of the results, the results for eye ROIs and eye region ROIs localization have been combined. As it is seen from the curves on the figures, the average localization rates are approximately equal for neutral and expressive sets for a given database. This demonstrates that a presence of facial expressions has not affected the ability of the method to locate the correct area.

The new rectangular error measure evaluates the result of feature ROI localization taking into account the position of the feature in the image and its true size, resulting in more strict accuracy evaluation criterion of feature ROI localization. The comparison of curves from Figures 7 and 8 shows a slight degradation in the average localization rates for the Cohn-Kanade and JAFFE expressive datasets. On the whole, however, the overall performance of the method as evaluated by the point and rectangular error measures is sufficiently high.

The lowest localization rates and the biggest number of localization errors have been observed for the Cohn-Kanade expressive dataset. This fact suggested a need for more detailed analysis to reveal the influence of particular facial behaviours on the feature ROI localization rates. Tables I-IV demonstrate the effect of single and conjoint facial muscle activations on the method performance when the localization accuracy has been evaluated by the proposed rectangular error measure.

5. DISCUSSION

COMPARISSON OF THE RESULTS

The eye positions have been argued to be sufficient criteria to declare successful face detection (Jesorsky *et al*., 2002) and initiate face recognition system (Campadelli at al., 2007). The majority of eye ROI localization results reported in the literature have been evaluated by the point error measure. For this reason, the results from Figure 8, which have the same evaluation criteria, will be used. In Table V we present a comparative analysis of our results against the feature ROI localization rates obtained with support vector machines (Campadelli at al., 2007), analysis of vertical and horizontal intensity histograms (Vukadinovic and Pantic, 2005), intensity-based template matching (Campadelli and Lanzarotti, 2004) and cascade of boosted classifiers (Wilson and Fernandez, 2006). The table shows that the achieved rates of eye ROI localization are comparable to those reported in the literature. The results reported on the Cohn-Kanade database (Vukadinovic and Pantic, 2005) demonstrate 100% success in the eye ROI localization and have been obtained on the neutral facial dataset. Although the rest of the listed methods have been tested on images which may include expressive behaviours in the lower face, the effect of these behaviours is unclear. Our results have been obtained on three databases of complex facial expressions and clearly demonstrate the invariance of the proposed method with regard to a wide variation of facial expressions. Similarly, our results for nose and mouth localization have demonstrated similar or superior performance to the results reported in the literature.

Unfortunately, studies on the feature ROI localization have not reported localization results for different expression categories coded either in terms of facial displays or AUs. For this reason we can compare the results from Tables I-IV only with our own study (Gizatdinova and Surakka, 2006). In that study, the deteriorating effect of expressions of happiness (AUs 6 and 12), anger (AUs 4, 6 and 7) and disgust (AUs 9, 10 and 11) has decreased significantly the feature ROI localization rates. Occurring along or in conjunction, the listed AUs frequently have caused merging of the neighbouring features. In the current version of the method, the applied procedure of x/y-edge projection in many cases has allowed the merged features to be successfully separated. In particular, upper face AUs 4, 6, 7, 43/45, 1+6, 1+7, 4+6, 4+7, 4+43/45, 6+7 and 9 typically have been found difficult to process with the previous version of the method. The improved version of the method has demonstrated a significant improvement in the eye ROI localization for the majority of the listed AUs.

Generally, mouth demonstrates a greater variability in its appearance than the eyes do. For example, in surprise and happy the mouth appearance usually is represented by open mouth (AU 25+26) sometimes with visible teeth and tongue. In anger the mouth could be opened (AU 22+25+26) or closed with tightened lips (AU 23), pressed lips (AU 24) and even lips sucked inside the mouth (AU 28), so that the reddish part of the mouth becomes not visible in the face. In the majority of cases, the proposed method has been able to find the mouth ROI regardless of whether the mouth is open or closed and whether the lips, teeth or tongue are visible or not.

Merging of the nose and mouth features has been largely eliminated, especially in the images with AUs 10, 11, 12, 15, 16, 17, 21, 11+20, 11+25, 12+20, 16+25 and 25+26.

NEW ACCURACY EVALUATION MEASURE

The localization accuracy of the presented method has been evaluated by the conventional point error measure and a new rectangular error measure. The results from Figures 7 and 8 show that the same method can give different results, if different evaluation criteria is applied. The criteria defined in Equation 7 takes into account the location of the feature in the image and its true size and, therefore, puts strict requirement which allows for more accurate localization of feature ROIs. With this respect, high method performance as evaluated by the rectangular error measure has reflected the fact that in most cases the method has precisely located the actual area of the feature that can undergo significant changes in shape and size (e.g. open mouth vs. bitted lips). On the contrary, the majority of the methods have considered the ROI localization to be a success if the centre point of the located ROI has been located within the limits of the feature or did not report about the criteria of what to consider successful feature ROI localization. We believe that the proposed rectangular error measure is an improvement over existing accuracy evaluation metrics of the feature ROI localization.

6. CONCLUSION

The automatic edge-based method of locating regions of prominent facial features from up-right facial images has been presented. The method has been tested on three public databases of complex facial expressions and demonstrated the invariance of feature ROI localization with respect to expressive deformations in the upper and lower face. The localization accuracy has been evaluated by a new rectangular error measure that takes into account the location of the feature in the image and the true feature size. The results have revealed similar or superior performance of the method as compared to state-of-the-art localization methods in the literature and our own earlier results.

The proposed method can be applied directly to the image without any image alignment given that the face takes the biggest part of the image. Alternatively, facial region can be detected first in order to facilitate the feature ROI search. Emphasizing the simplicity of the proposed method, we conclude that it can be used in preliminary localization of facial feature ROIs for their subsequent processing, where coarse feature localization is followed by fine feature point detection. The future plans include further improvement of the robustness of the method for particular facial behaviours, real-time optimization of the method implementation and its application in a real-time system of facial expression analysis.

7. ACKNOWLEDGMENTS

This work has been supported by the Finnish Academy (project numbers 177857 and 1115997), the University of Tampere and the Tampere Graduate School in Information Science and Engineering. The authors thank the creators and actors involved into development of the facial databases for the permission to use and reprint examples of expressive images. The authors also thank the anonymous reviewers for their helpful comments.

REFERENCES

Campadelli, P., Lanzarotti, R., Lipori, G., 2007. Automatic facial feature extraction for face recognition, in: Delac, K., Grgic, M., (Eds.), Face Recognition. I-Tech Education and Publishing, Vienna, pp. 31-58.

Campadelli, P., Lanzarotti, R., Lipori, G., Salvi, E., 2005. Face and Facial Feature Localization, in Proc. Int. Conf. Image Analysis and Processing, pp. 1002-1009.

Cooray, S., O'Connor, N., 2004. Facial feature extraction and principal component analysis for face detection in color images. Lecture Notes in Comp. Science. 3212, 741-749.

Cristinacce, D., Cootes, T., 2006. Feature detection and tracking with constrained local models, in Proc. British Machine Vision Conference, pp. 929-938.

Ekman, P., Friesen, W., 1976. Pictures of Facial Affect. Palo Alto, California, Consulting Psychologists Press.

Ekman P., Friesen W., Hager J., 2002. Facial Action Coding System (FACS). Salt Lake City, UTAH, A Human Face.

Farkas, L., 1994. Anthropometry of the Head and Face, second ed. Raven, New York.

Feris, R., Gemmell, J., Toyama, K., Krüger, V., 2002. Hierarchical Wavelet Networks for Facial Feature Localization, in Proc. Int. Conf. Automatic Face and Gesture Recognition, pp. 118-123.

Gizatdinova, Y., Surakka, V., 2006. Feature-based detection of facial landmarks from neutral and expressive facial images. Trans. Pattern Analysis and Machine Intelligence. 28, 1, 135-139.

Gizatdinova, Y., Surakka, V., 2007. Automatic Detection of Facial Landmarks from AU-Coded Expressive Facial Images, in Proc. Int. Conf. Image Analysis and Processing, pp. 419-424.

Gizatdinova, Y., Surakka, V., 2008. Effect of Facial Expressions on Feature-Based Landmark Localization in Static Grey Scale Images, in Proc. Int. Conf. Computer Vision Theory and Applications, pp. 259-266.

Heisele, B., Serre, T., Poggio, T., 2007. A component-based framework for face detection and identification. Int. J. Comp. Vision. 74, 2, 167-181.

Hjelmas, E., Low, B., 2001. Face detection: A survey. Comp. Vision and Image Understanding. 83, 235–274.

Hurvich, L., 1981. Color vision. Sinauer Associates, Sunderland, MA.

Jesorsky, O., Kirchberg, K.J., Frischholz, R.W., 2001. Robust face detection using the Hausdorff distance, in Proc. Int. Conf. Audio- and Video-Based Person Authentication, pp. 90–95.

Hubel, D.H., Wiesel, T.N., 1962. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. J. Physiology. 160, 1, 106–154.

Kanade, T., Cohn, J., Tian, Y., 2000. Comprehensive database for facial expression analysis, in Proc. Int. Conf. Automatic Face and Gesture Recognition, pp. 46-53.

Lien, J., Kanade, T., Cohn, J., Li, C., 2000. Detection, tracking, and classification of action units in facial expression. J. Robotics and Autonomous Systems. 31, 131-146.

Lyons, M.J., Akamatsu, Sh., Kamachi, M., Gyoba, J., 1998. Coding Facial Expressions with Gabor Wavelets, in Proc. Int. Conf. Automatic Face and Gesture Recognition, pp. 200-205.

Marr, D., 1982. Vision: A computational investigation into the human representation and processing of visual information, in: Freeman, W.H., (Ed.), San Francisco, Freeman Publishers.

Pantic, M., Rothkrantz, J.M., 2000. Automatic analysis of facial expressions: The state of the art. Trans. Pattern Analysis and Machine Intelligence. 22, 12, 1424–1445.

Rodriguez, Y., Cardinaux, F., Bengio, S., Mariéthoz, J., 2006. Measuring the performance of face localization systems. Image and Vision Computing. 24, 8, 882-893.

Song, J., Chi, Zh., Liu, J., 2006. A robust eye detection method using combined binary edge and intensity information. Pattern Recognition. 39, 6, 1110-1125.

Tian, Y.-L., Kanade, T., Cohn, J., 2002. Evaluation of Gabor wavelet-based facial action unit recognition in image sequences of increasing complexity, in Proc. Int. Conf. Automatic Face and Gesture Recognition, pp. 229-234.

Tong, Y., Wang, Y., Zhu, Zh., Ji, Q., 2007. Robust facial feature tracking under varying face pose and facial expression. Pattern Recognition. 40, 11, 3195-3208.

Venkatesh, Y., Kassim, A., Ramana Murthy, O., 2009. A novel approach to classification of facial expressions from 3D-mesh datasets using modified PCA. Pattern Recognition Letters. 30, 12, 1128-1137.

Vukadinovic, D., Pantic, M., 2005. Fully Automatic Facial Feature Point Detection Using Gabor Feature Based Boosted Classifiers, in Proc. Int. Conf. Systems, Man And Cybernetics, pp. 1692–1698.

Wilson P., Fernandez, J., 2006. Facial feature detection using Haar classifiers. J. Computing Sciences in Colleges. 21, 4, 127-133.

Yang, M., Kriegman, D., Ahuaja, N., 2002. Detecting face in images: A survey. Trans. Pattern Analysis and Machine Intelligence. 24, 1, 34-58.

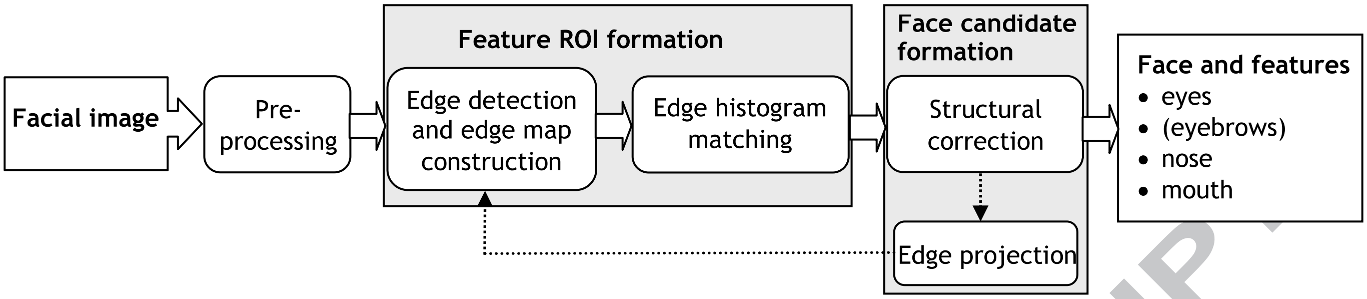
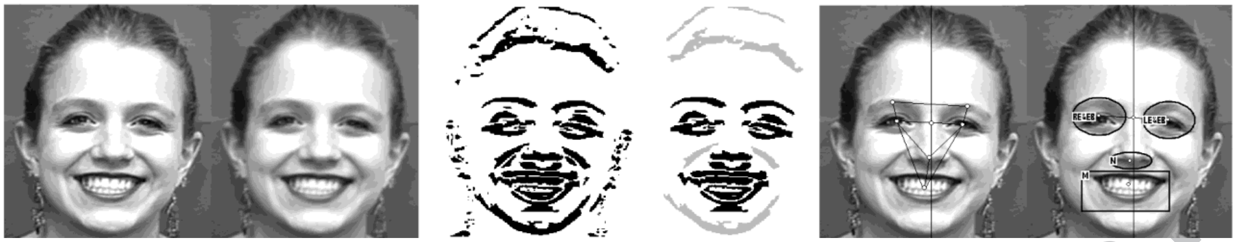
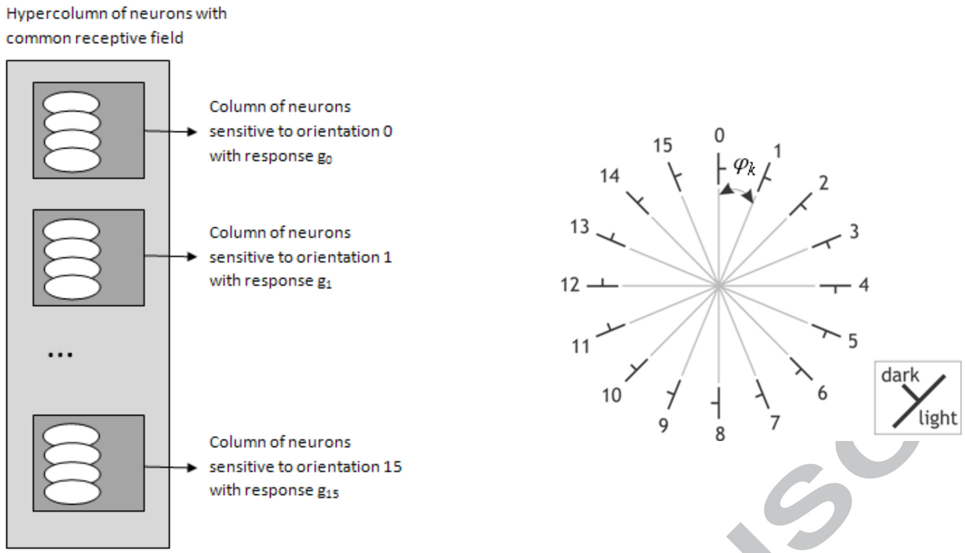

Facial image
Pre-processing
Feature ROI formation
Edge detection and edge map construction
Edge histogram matching
Face candidate formation
Structural correction
Edge projection
Face and features
• eyes
• (eyebrows)
• nose
• mouth

Figure 1.



(a) (b) (c) (d) (e) (f)

Figure 2.


Hypercolumn of neurons with common receptive field
Column of neurons sensitive to orientation 0 with response g_0
Column of neurons sensitive to orientation 1 with response g_1
...
Column of neurons sensitive to orientation 15 with response g_{15}
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
φ_k
dark
light

(a)

(b)

Figure 3.

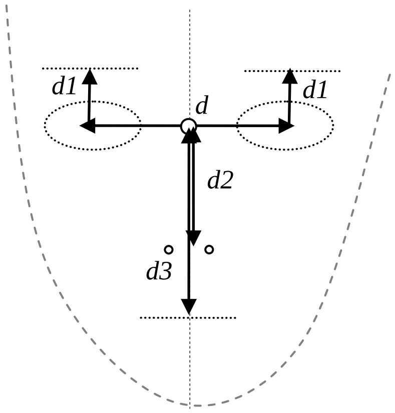


Figure 4.

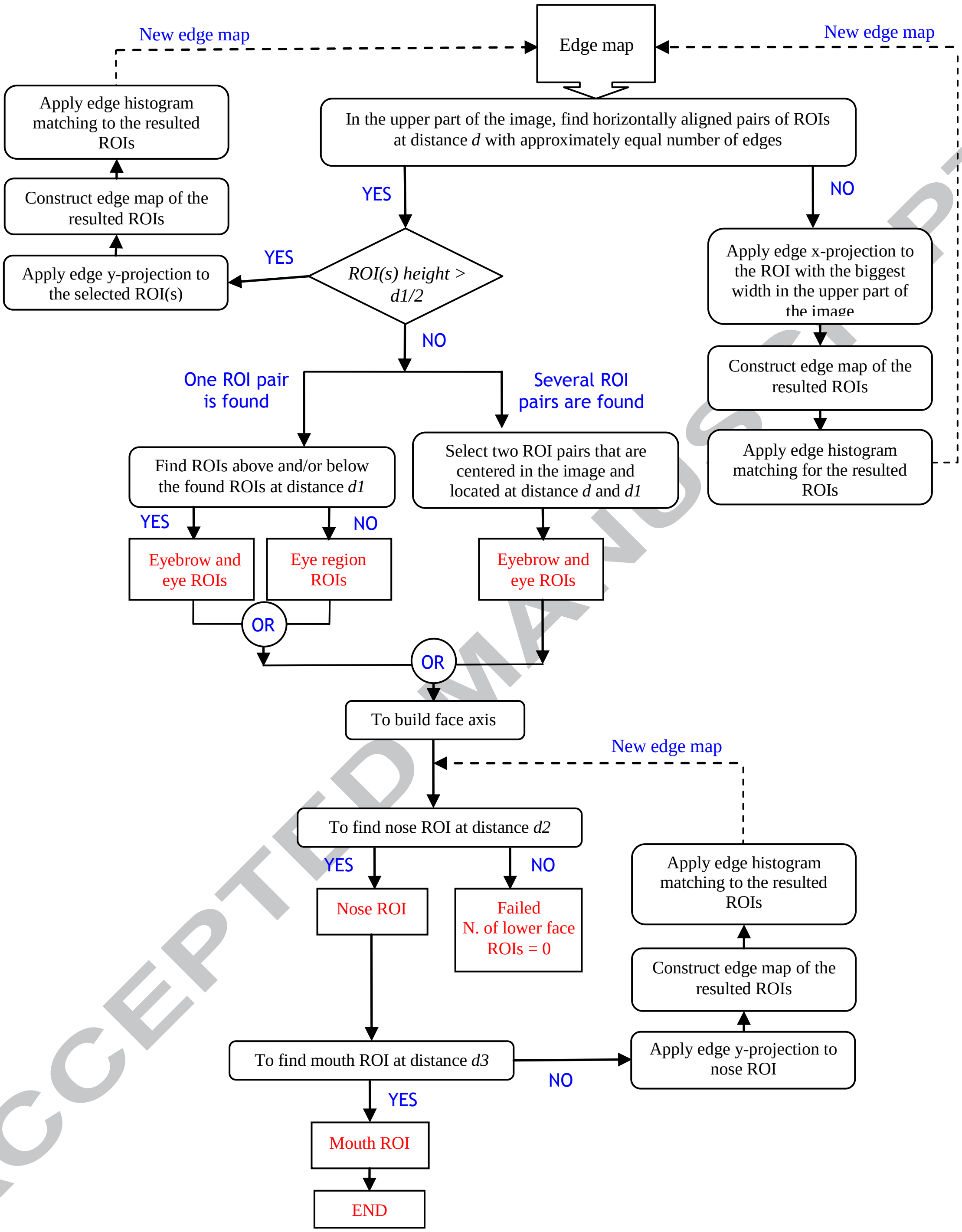
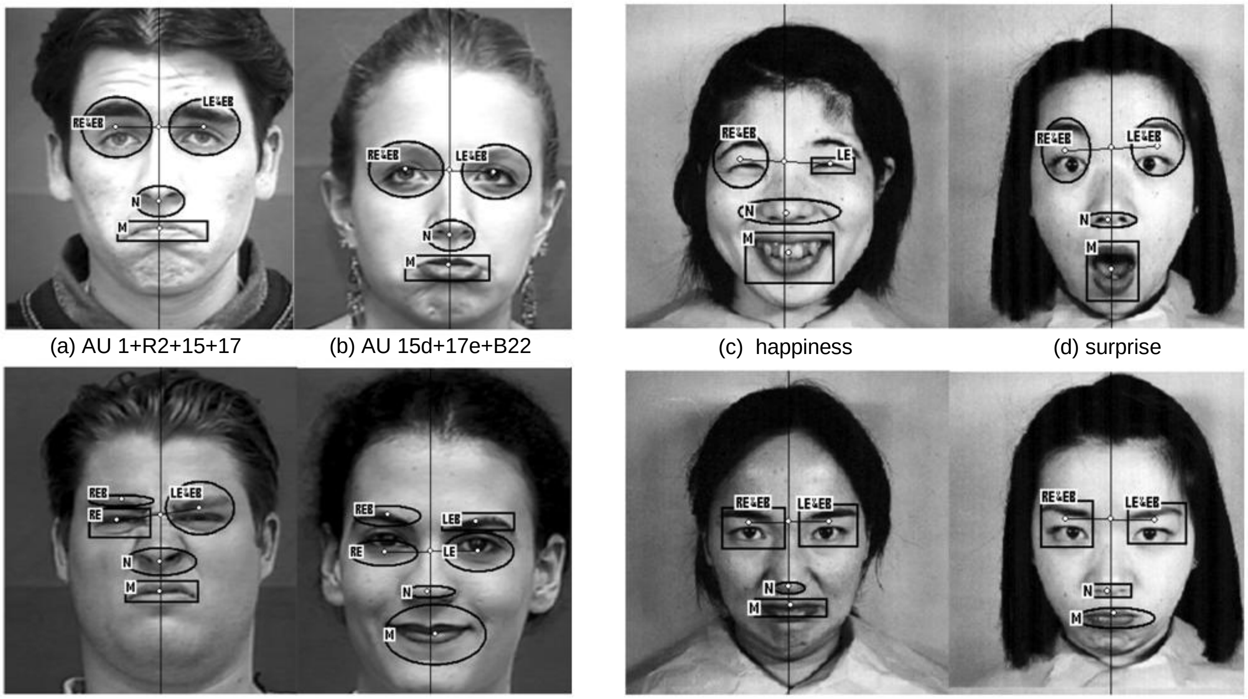

Edge map
New edge map
New edge map
Apply edge histogram matching to the resulted ROIs
Construct edge map of the resulted ROIs
Apply edge y-projection to the selected ROI(s)
In the upper part of the image, find horizontally aligned pairs of ROIs at distance d with approximately equal number of edges
YES
NO
YES
ROI(s) height > d1/2
NO
Apply edge x-projection to the ROI with the biggest width in the upper part of the image
Construct edge map of the resulted ROIs
Apply edge histogram matching for the resulted ROIs
One ROI pair is found
Several ROI pairs are found
Find ROIs above and/or below the found ROIs at distance d1
Select two ROI pairs that are centered in the image and located at distance d and d1
YES
NO
Eyebrow and eye ROIs
Eye region ROIs
Eyebrow and eye ROIs
OR
OR
To build face axis
New edge map
To find nose ROI at distance d2
YES
NO
Nose ROI
Failed N. of lower face ROIs = 0
Apply edge histogram matching to the resulted ROIs
Construct edge map of the resulted ROIs
Apply edge y-projection to nose ROI
To find mouth ROI at distance d3
NO
YES
Mouth ROI
END

Figure 5.



(a) AU 1+R2+15+17 (b) AU 15d+17e+B22 (c) happiness (d) surprise

(e) 4+7+9+17 (f) AU 12c (g) anger (h) anger

Figure 6

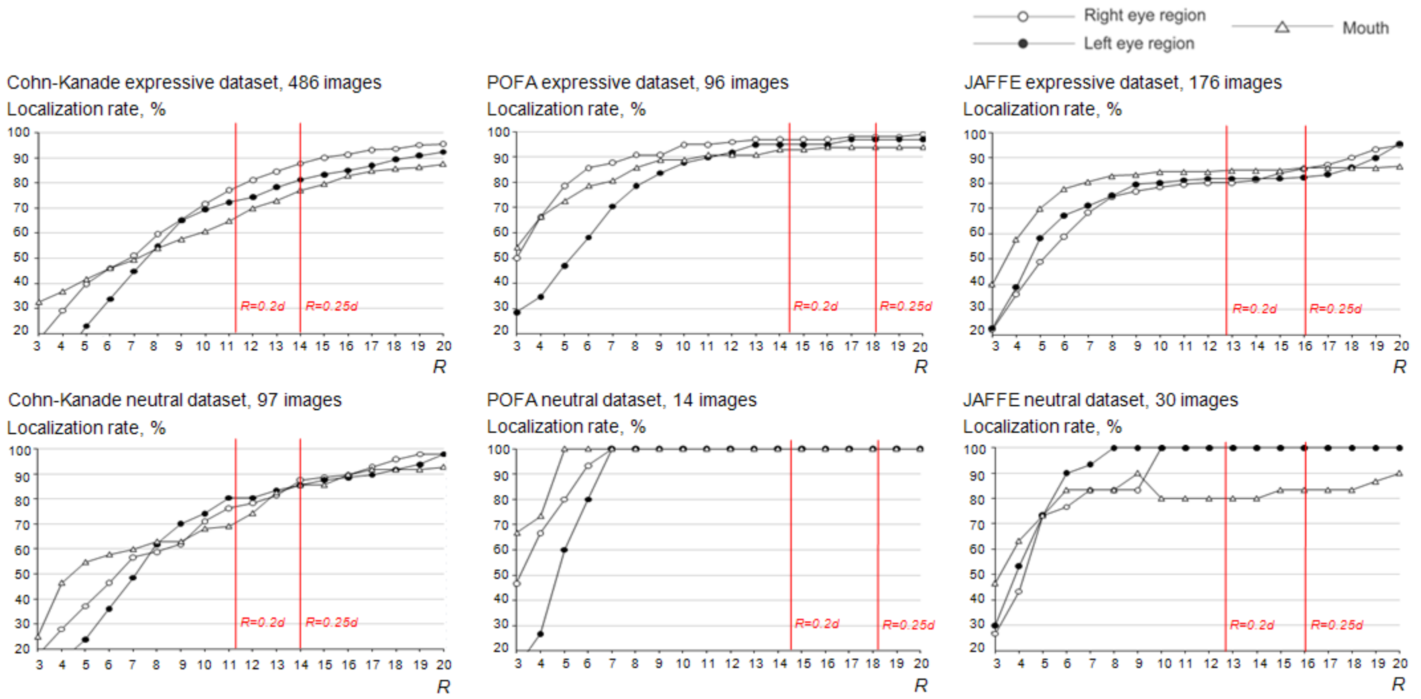

Right eye region
Left eye region
Mouth
Cohn-Kanade expressive dataset, 486 images
POFA expressive dataset, 96 images
JAFFE expressive dataset, 176 images
Cohn-Kanade neutral dataset, 97 images
POFA neutral dataset, 14 images
JAFFE neutral dataset, 30 images
Localization rate, %
100
90
80
70
60
50
40
30
20
3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
R
R=0.2d
R=0.25d

Figure 7.

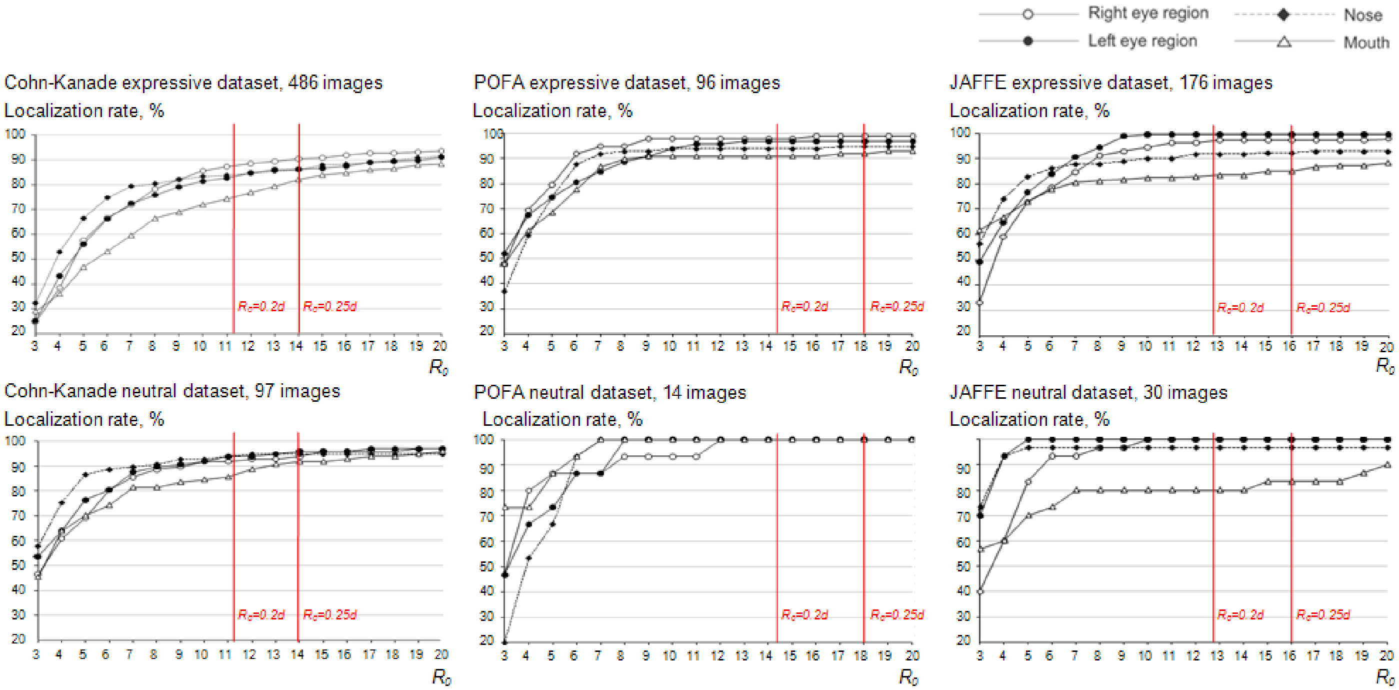

Right eye region
Left eye region
Nose
Mouth
Cohn-Kanade expressive dataset, 486 images
Localization rate, %
POFA expressive dataset, 96 images
Localization rate, %
JAFFE expressive dataset, 176 images
Localization rate, %
Cohn-Kanade neutral dataset, 97 images
Localization rate, %
POFA neutral dataset, 14 images
Localization rate, %
JAFFE neutral dataset, 30 images
Localization rate, %
100
90
80
70
60
50
40
30
20
3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
R_0
R_c=0.2d
R_c=0.25d

Figure 8.

Table I: Localization rates averaged over upper face AUs[a]

Facial features	Upper face AUs						
	1	2	4	5	6	7	43/45
Eye region	89	91	87	94	77	85	96
Nose	86	86	82	87	77	76	85
Mouth	71	64	83	68	78	79	88

[a] Note that AU43 (eye closure) and AU45 (blink) were combined together because they both have the same visual effect on the facial appearance and different durations of these AUs cannot be measured from static images.

Table II: Localization rates averaged over upper face AU combinations

Facial features	Upper face AU combinations											
	1+2	1+4	1+5	1+6	1+7	2+4	2+5	4+5	4+6	4+7	4+43/45	6+7
Eye region	91	82	93	100	84	77	93	88	85	87	91	79
Nose	86	82	85	83	75	82	86	77	77	77	64	68
Mouth	64	84	65	83	75	82	64	85	80	78	91	76

Table III: Localization rates averaged over lower face AUs

Facial features	Lower face AUs													
	9	10	11	12	14	15	16	17	20	23	24	25	26	27
Eye region	85	83	87	77	70	94	90	90	87	95	92	87	92	93
Nose	81	75	84	78	70	96	90	89	79	82	81	84	94	85
Mouth	83	92	84	78	60	91	81	87	79	79	79	79	89	63

Table IV: Localization rates averaged over lower face AU combinations

Facial features	Lower face AU combinations													
	9+17	11+20	11+25	12+16	12+20	12+25	16+20	16+25	17+23	17+24	20+25	23+24	25+26	25+27
Eye region	85	89	89	89	75	73	100	89	93	91	87	94	91	94
Nose	81	86	86	89	50	75	88	89	80	82	79	79	94	84
Mouth	81	86	86	78	50	73	88	78	80	79	79	79	89	66

Table V: Comparative analysis of feature ROI localization methods (localization rates (%) are given for $R_0 = 0.25 \cdot d$)

Study and localization method	Database	Eye region ROI	Nose ROI	Mouth ROI
(Campadelli at al., 2007)				
Support vector machines	BANCA	99.0		
	BioID	95.5		
	FERET	95.6		
	FRGC	97.1		
	XM2VTS	97.8		
(Vukadinovic and Pantic, 2005)	Cohn-Kanade, neutral	100		99
Vertical and horizontal intensity histograms				
(Campadelli and Lanzarotti, 2005)	XM2VTS	98.3		
Intensity-based template	The Univ. of Stirling database	98.5		
matching	UnimiDb	99.2		
(Wilson and Fernandez, 2006)	FERET	93	100	67
cascade of boosted classifiers				
Our study				
	Cohn-Kanade, neutral	95	95	92
	Cohn-Kanade, expressive	88	86	82
	POFA, neutral	100	100	100
	POFA, expressive	98	95	92
	JAFFE, neutral	100	97	83
	JAFFE, expressive	98	92	85