2D/3D Facial Feature Extraction

Hatice Çınar Akakın[1], Albert Ali Salah[2], Lale Akarun[2], Bülent Sankur[1],
[1]Electrical and Electronic Engineering Department, Boğaziçi University,
[2] Computer Engineering Department, Perceptual Intelligence Laboratory, Boğaziçi University,
[hatice.cinar, salah, akarun, bulent.sankur]@boun.edu.tr
Telephone: (90) 212 359 6414, Fax: (90) 212 287 2465, Bebek, İstanbul, Turkey
Corresponding author: Bülent Sankur

ABSTRACT

We propose and compare three different automatic landmarking methods for near-frontal faces. The face information is provided as 480x640 gray-level images in addition to the corresponding 3D scene depth information. All three methods follow a coarse-to-fine suite and use the 3D information in an assist role. The first method employs a combination of principal component analysis (PCA) and independent component analysis (ICA) features to analyze the Gabor feature set. The second method uses a subset of DCT coefficients for template-based matching. These two methods employ SVM classifiers with polynomial kernel functions. The third method uses a mixture of factor analyzers to learn Gabor filter outputs. We contrast the localization performance separately with 2D texture and 3D depth information. Although the 3D depth information per se does not perform as well as texture images in landmark localization, the 3D information has still a beneficial role in eliminating the background and the false alarms.

1. INTRODUCTION

The detection and localization of the face and of its features is instrumental for the successful performance of subsequent tasks in related computer vision applications. Many high-level vision applications such as facial feature tracking, facial modeling and animation, facial expression analysis, and face recognition, require reliable feature extraction. Facial feature points are referred to in the literature as "salient points", "anchor points", or "facial landmarks". The most frequently occurring fiduciary facial features are the four eye corners, the tip of the nose, and the two mouth corners. Additionally, the eyebrows, the bridge of the nose, the tip of the chin, and the nostrils are sometimes used as facial landmarks.

Facial feature detection is a challenging computer vision problem due to high inter-personal changes (gender, race), the intra-personal variability (pose, expression) and acquisition conditions (lighting, scale, facial accessories). The desiderata for facial feature localization list as follows: a) Accurate; b) Precise within a few millimeters; c) Computationally feasible since most systems must run in real time; d) Robust against pose, illumination, expression and scale variations, as well as against occlusions and disturbance from facial accessories; e) Amenable to dynamic tracking.

We distinguish between face detection, where the aim is the coarse localization of the facial bounding box, and face localization, which requires exact localization of the anchor points (as in Hamouz et al.[18]).

Literature surveys indicate that in automatic facial landmarking systems almost always heuristics are involved, which can be particular to a dataset[2,5,7,16,19,28,33]. For example, for 3D faces the nose is found as the maximum protrusion at about the centre of the bounding box, while for the gray-level data the eye sockets correspond to the low valleys in the vertical projection histograms. However, these heuristics may not be always robust. For the above two examples, a streak of hair or a protruding chin in a backward tilted head can mislead face range data to a nose. The valleys of the vertical projection depend is true for non-rotated, non-inclined and non-tilted faces. For this reason, many face recognition approaches assume normalized faces at the outset, and opt for manual localization of the landmarks.

Most approaches use a coarse-to-fine localization scheme to reduce the computational load[2,7,11,14,19,25,29,31]. The initial face detection is mostly performed by using skin color segmentation[5,9,11,21,37]. This may be followed by horizontal and vertical projections of edge images or gray-level intensities[4] and identification of valleys and peaks corresponding to

Image Processing: Algorithms and Systems, Neural Networks, and Machine Learning
edited by E.R. Dougherty, J.T. Astola, K.O. Egiazarian, N.M. Nasrabadi, S.A. Rizvi
Proc. of SPIE-IS&T Electronic Imaging, SPIE Vol. 6064, 60641D, © 2006 SPIE-IS&T · 0277-786X/06/$15

features[33]. A caveat is not to initialize the landmarking process by relying solely upon one detected landmark, e.g. tip of the nose in 3D, as this becomes the soft belly of the algorithm.

The feature localization step following the face detection can be taken up by such techniques as appearance-based, geometric-based and structure-based. Appearance-based approaches aim to model the facial features in a suitable subspace such as principal components analysis (PCA)[1,25], independent components analysis (ICA)[1], discrete cosine transform (DCT)[36], Gaussian derivative filters[2,15] and Gabor wavelets[14,28,29,31]. Once the image is transformed to the subspace representation, different machine learning techniques like boosted cascade detectors[9,11], support vector machines (SVM)[1], and multi-layer perceptrons (MLP)[7,25] are applied. Geometric-based methods use angles, distances and areas between landmarks[28,30,36]. Finally, in the structure-based methods, the ensemble of candidate landmarks is fitted to a model of feature locations and the likelihood is considered. In an seminal paper, Wiskott et al. use an elastic bunch graph to model the distances between the fiducial points[32]. A number of templates (called the bunch) are used to test local feature responses. Other approaches use an optimization function that incorporates terms for local similarity and global structure of the landmark distribution, and penalize transformations differently[11,34]. These models require good initial points (often manually determined) to avoid local minima.

Recently, 3D information has been considered in complementary role for facial feature localization[10]. The 3D information, potentially useful, brings has a few pitfalls. Depending on the sensor type, the acquisition process can be noisy, resulting in holes, spikes, and surface irregularities. The size of the data can be considerably larger as compared to 2D images; in fact 5,000-90,000 samples are not uncommon. A multi-level approach is possible that starts from a coarse initial localization, followed by an iterative closest point (ICP) registration that greatly constrains possible locations for fine landmark localization[20]. In Colbry et al., 2D information in the form of Harris corners is used in conjunction with 3D shape indices to train a feature-driven statistical system that can successfully localize facial features[10]. In Boehnen and Russ, the 3D information supports the 2D system both by filtering out the background, and by supplying true metric information for reliable intra-feature distance calculation[5]. We use 3D information in a similar supportive role in the present work.

In this work, we consider the problem of facial feature localization using joint 2D and 3D information, that is, luminance information coupled with range information. We propose and analyze the performance of three novel facial landmark localization approaches. Our methods are hybrid schemes, where gray-level appearance information plays a predominant role and it is assisted by the 3D information. Certain low-level features of facial images produce good candidates for each landmark point. These multiple locations are tested by a structural subsystem to validate it on the one hand and to refine the accuracy of its location on the other hand. The three fine localization methods employ: i) Gabor features transformed by PCA and ICA; ii) DCT coefficients, analyzed by SVM; iii) Gabor features and depth map analyzed by mixtures of factor analyzers, respectively. The 3D information provides structural verification and background removal in all the methods, while in addition, in the third method it is also employed as a source of local information.

The paper is structured as follows: In Section 2, we introduce three novel facial feature localization algorithms. We present our simulation results in Section 3. Section 4 concludes and indicates future directions.

2. NOVEL FACIAL LANDMARKING ALGORITHMS

The three proposed methods employ a two-stage coarse-to-fine landmarking approach: For a computationally efficient system, we start by searching potential landmark zones on downsampled 2D images. We downsample the face images by a factor of eight (from the size 480x640 down to 60x80). Once the candidate locations are established, we revert back to the higher resolution image and refine the accuracy by using search windows around the coarse landmark locations. Background elimination is implemented by using the depth images, which in turn yields the location of the face coarsely.

Three different techniques are applied to compose discriminative feature vectors of the local facial patches: The first one is the Independent Gabor Features (IGF), where Gabor features are transformed via PCA and ICA in succession, and then classified with binary SVM classifiers (Section 2.1). In the second method, we use DCT coefficients as an alternative, again coupled with SVM classifier for feature localization (Section 2.2). The third method is the Gabor Factor Analysis, where the feature localization is performed with a generative model and complemented with a structural correction scheme (Section 2.3).

2.1. Algorithm I: Facial Features via Independent Gabor Features

The Independent Gabor Features (IGF) has been previously used by Liu and Wechsler for face recognition[23]. These features form a derivative of Gabor feature vectors, computed in different scales and orientations. Gabor kernels have such advantages as being tunable for good spatial localization and high frequency selectivity[23,29]. The image is initially convolved with a Gabor kernel as below:

$$\Psi_j(\vec{x}) = \frac{\vec{k}_j \vec{k}_j^{\,T}}{\sigma^2} e^{\left(-\frac{\vec{k}_j \vec{k}_j^{\,T} \vec{x}\vec{x}^{T}}{2\sigma^2}\right)} \left[e^{(i\vec{k}_j \vec{x})} - e^{(\frac{-\sigma^2}{2})} \right] \tag{1}$$

$$\vec{k}_j = (k_{jx}, k_{jy}) = (k_v \cos\varphi_w, k_v \sin\varphi_w),\ k_v = 2^{-\frac{v+2}{2}}\pi,\ \varphi_w = w\frac{\pi}{8}$$

where $\vec{x} = (x, y)$ is the given pixel location, $j = w + 8v$, and (w,ν) defines the orientation and scale parameters of the Gabor kernels, respectively, and standard deviation of the Gaussian function σ is 2π. The first factor in the Gabor kernel represents the Gaussian envelope and the second factor represents the complex sinusoidal function, known as the carrier. The term, $e^{-\sigma^2/2}$, of the complex sinusoidal compensates for the DC value. The general block diagram of the IGF method is given in Figure 1.

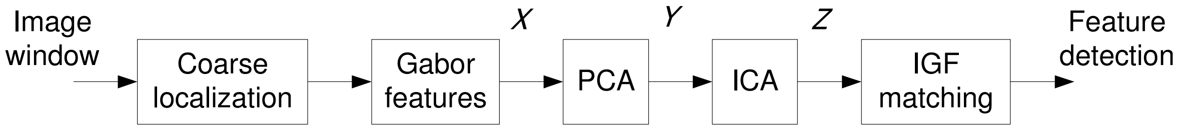

Figure 1. Overview of the IGF method (adapted from[23]). A neighborhood around coarse landmark locations is filtered with Gabor wavelets. The ensuing feature vector is transformed by PCA and ICA in succession before IGF matching. Coarse localization in itself follows a similar data flow, but on downsampled images.

First the dimensionality of these feature vectors is reduced with PCA. Then, they are processed with ICA, which takes the higher-order statistics into account[3]. We designed two Gabor kernel sets. The first set is used for coarse localization on downsampled images, composed of three different scales, i.e., $\nu \in \{0,1,2\}$ and four orientations, $w \in \{0,2,4,6\}$. The second set includes eight transforms, with two different scales, i.e., $\nu \in \{0,1\}$ and four orientations, $w \in \{0,2,4,6\}$. From the Gabor transformed face, we crop patches (i.e. feature generation windows) around the each pixel of the search window. Each resulting Gabor feature is z-normalized by subtracting the component mean and dividing by its standard deviation. Finally, the component vectors over the grid are juxtaposed to form a larger feature vector X. The dimensionality is reduced to dimension $n << N$ via PCA projection as

$$Y = P^t X \tag{2}$$

where $P = [P_1\ P_2\ \ldots\ P_n]$ is the N x n eigenvector matrix corresponding to the n largest eigenvectors of the covariance matrix of X. The ICA method, which expands PCA by considering higher order statistics, was previously employed to derive independent Gabor features that were successful in human face recognition[23]. Other applications of ICA for face recognition can be found[3,12]. The independent Gabor feature vector Z is obtained by multiplying the PCA-transformed features with W, the demixing matrix obtained with the fastICA algorithm:

$$Z = WY \tag{3}$$

The de-mixing matrix W is obtained by maximizing some contrast function of the source data, which are assumed to be statistically independent and non-Gaussian. The derived ICA transformation matrix W is a combination of whitening, rotation, and normalization transformations.

In the coarse level feature localization, 7x7 patches are cropped around each search point and 100-dimensional IGF vectors are obtained by applying the PCA projection on 7x7x12 dimensional Gabor feature vectors. In the fine level, 11x11 patches are cropped around the candidate points and 150-dimensional IGF vectors are obtained via PCA dimension reduction on 11x11x8 dimensional Gabor feature vectors. Computed feature vectors are fed to the SVM classifiers[6].

2.1.1. Support Vector Machine Classification

For the IGF and DCT based methods, we use SVM classifiers. SVMs belong to the class of maximum margin classifiers, such that they find a decision hyperplane for a two-class classification problem by maximizing the margin, which is the distance between the hyperplane and the closest data points of each class in the training set that are called support vectors. This linear classifier is termed the optimal separating hyperplane (OSH). Assuming linearly separable data, a separating hyperplane $wx+b=0$ exists. The set of vectors is optimally separated without misclassification when the margin is maximal. The optimal plane must satisfy the condition $y_i(w.x+b)\geq 1$. The distance between the hyperplane and the points is

$$d(w)=\frac{|w.x+b|}{\|w\|}\geq\frac{1}{\|w\|} \tag{4}$$

therefore the optimal plane is obtained by minimizing $\frac{1}{2}w^T w$ subject to $y_i(w.x+b)\geq 1$. The optimization problem can be solved by the Lagrange function,

$$L(w,b,\alpha)=\frac{1}{2}w^T w-\sum_{i=1}^{N}\alpha_i\left[y_i(w.x+b)-1\right] \tag{5}$$

where α_i are Lagrange multipliers. After solving the optimization problem [6], the OSH has the form:

$$f(x)=\sum_{i=1}^{N}\alpha_i y_i x_i^T x+b \tag{6}$$

In the case of data, which are not linearly separable, we can project the data to into a higher–dimensional space in the hope of finding a linear OSH there. This is done by replacing the inner product $x_i^T x_j$ with a kernel function $K(x_i,x_j)$ that satisfies Mercer conditions, thus allowing fast computations in the low-dimensional space, rather than the new, high-dimensional space:

$$f(x)=\sum_{i=1}^{N}\alpha_i y_i K(x,x_i)+b \tag{7}$$

One of the most frequently used kernel functions is the polynomial kernel $K(\mathbf{x},\mathbf{y})=(1+\mathbf{x}.\mathbf{y})^d$, where *d* is the degree of the polynomial. We have used fifth degree polynomial kernels throughout our experiments. SVM classifiers are trained for both coarse and fine localization stages.

2.2. Algorithm II: DCT-based Facial Feature Extraction

In this approach, instead of using data-driven features for localization, we used lower frequency DCT features[13,24]. DCT coefficients can capture the statistical shape variations and can be a faster alternative for facial landmark detection, compared to local Gabor feature analysis. At the candidate facial feature points, the *C*(*v, u*) matrix containing DCT coefficients is computed as follows:

$$C(v,u)=\alpha(v)\alpha(u)\sum_{y=0}^{K-1}\sum_{x=0}^{K-1}f(y,x)\beta(y,x,v,u)\ \text{for } v,u=0,1,\ldots,K\text{-}1, \tag{8}$$

where $\alpha(v) = \begin{cases} \sqrt{\dfrac{1}{K}} & \text{for } v = 0 \\ \sqrt{\dfrac{2}{K}} & \text{for } v = 1, 2, \ldots, K-1 \end{cases}$ and $\beta(y, x, v, u) = \cos\left[\dfrac{(2y+1)v\pi}{2K}\right]\cos\left[\dfrac{(2x+1)u\pi}{2K}\right]$

The coefficients are ordered according to a zigzag pattern, in agreement with the amount of information stored in them. The first coefficient (DC value) is removed, since it only represents the average intensity value of the block. The remaining (AC) coefficients denote the intensity changes or gray-level shape variations over the image block. To analyse the effect of DCT coefficients both in coarse and fine stages, different number of DCT coefficients are used to form the DCT feature vector. In the coarse localization part we compute 8x8 DCT blocks from the whole face image except the background. For the fine localization, 16x16 DCT blocks are extracted for each point in a window centered at the coarsely estimated location. We use a window of size 19x19, and obtain 361 candidate points for each feature.

2.3. Algorithm III: Gabor Factor Analysis

In this method, we use feature similarity scores locally to determine the landmarks. Gabor features in 8 orientations and a single scale are extracted from a 7x7 window around each landmark in the training set, and each orientation channel is modeled via a generative factor analysis mixture. During the localization phase, each Gabor channel of the test face is converted to a feature similarity score map (conspicuity map). The initial landmark positions are found on the downsampled 2D appearance image, and the fine-tuning is performed on the higher resolution 2D images using also the 3D range image. A structural correction step between coarse and fine stages helps to detect and correct mislocated landmarks. Figure 2 summarizes the coarse landmarking scheme.

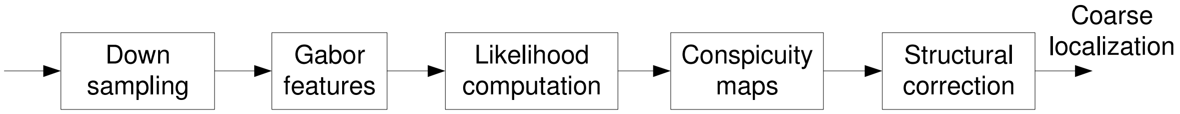

Figure 2. Coarse landmark localization starts with computation of Gabor wavelets on downsampled images. Conspicuity maps for each landmark are computed by summing the likelihoods under mixtures of factor analyzers models. The most conspicuous locations are sent to the structural analysis subsystem for correction.

We downsample the face image by a factor of eight (from the size 480x640 to 60x80) and discard roughly two thirds of these points (from 307,200 pixels, only about 90,000 have depth information as mentioned before, i.e. one third) as they lack depth information. The 60x80=4,800 coarse level search points are reduced to 1,500-2,000 by eliminating pixels with no associated depth value, i.e. the background. We have employed Gabor filters in eight orientations ($w \in \{0,1,2,3,4,5,6,7\}$), a single scale ($\nu \in \{3\}$) and 7x7 window is used for feature generation. Using more scale bands in Gabor analysis or larger feature generation windows did not contribute to the overall accuracy of the system.

The ground truth is obtained by manual landmarking of the training set. The features in each Gabor channel are modeled by an Incremental Mixtures of Factor Analyzers (IMoFA-L)[26]. A mixture of factor analyzers is a mixture of Gaussians, which reduces drastically the number of parameters. This model was used by Yang et al. in a face detection application and was shown to outperform PCA[35]. The IMoFA-L model automatically finds a trade-off between accuracy and complexity by adding components and factors one by one, while monitoring likelihood on a separate validation set. Thus, complex patterns in the data are modeled with more components, and with more factors per component (e.g. mouth corners), whereas simple patterns (e.g. nose) are modeled with smaller number of parameters.

The local analysis proceeds by computing the likelihood of each feature generation window on the test image, creating one conspicuity map per Gabor channel. These channel scores are summed to determine the location with the highest likelihood, hence the exhaustive search at the coarse level produces one candidate per facial feature. The locations of these features are analyzed via a structural correction system. The structural correction system relies on the correct localization threesome landmarks (called a *support set*) chosen out of seven landmarks. There are Comb(7,3) = 35 possible subsets, but 2-3 hypotheses are tested on the average, as a correct support set is usually found among the first few. To test a support set, we apply an affine normalization (translation, scaling and rotation) based on the support set. The expected locations of the rest of the landmarks (the non-support set) are modeled as Gaussian distributions. A support set is validated if the remaining landmarks produce a high structural fitting score. The landmarks, which are detected as outliers, are re-estimated by their expected locations. A landmark l_j is assumed to be an inlier, if its likelihood under the model (denoted with the mean μ_j, and the covariance Σ_j, for j^th landmark) is higher than a fixed threshold:

$$L(l_j, \mu_j, \Sigma_j) > \tau \qquad (9)$$

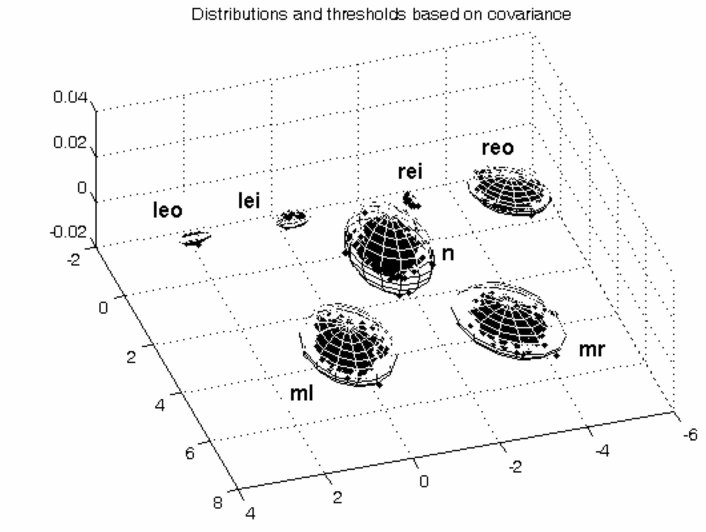

Figure 3. Thresholds for outlier detection are shown as ellipsoids around landmark clusters. The support set consists of the corners of the left eye (leo, lei) and the inner corner of the right eye (rei). As the normalization is based on the landmarks of the support set, they have smaller variations.

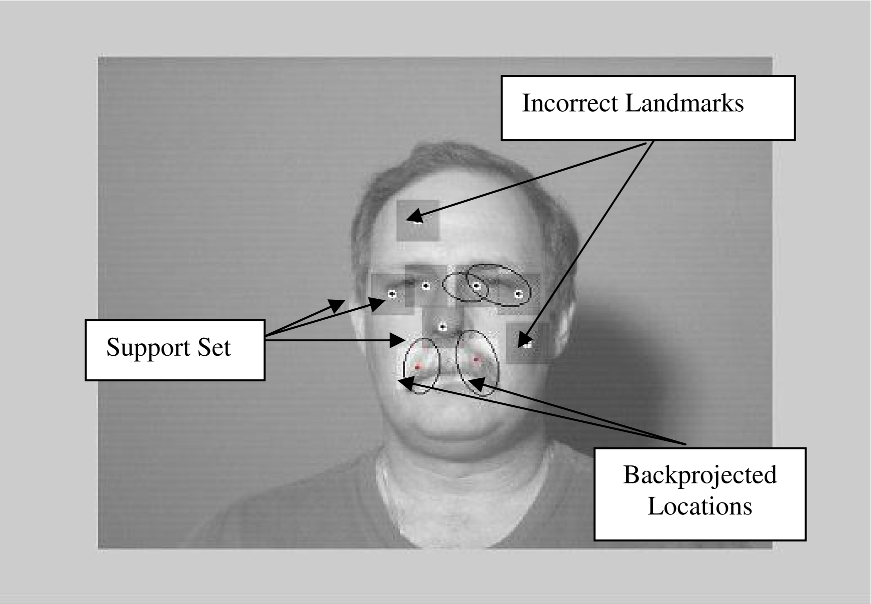

Figure 4. The coarse localization results projected to the original image. The gray boxes are 41x41 neighbourhoods. The ellipsoids indicate the expected locations for each landmark outside the support set. Landmarks in the support set, detected incorrect landmarks and their backprojected coarse locations are shown.

The fine level analysis uses a (19x19) search window around the coarse landmark locations. From the higher resolution depth and Gabor-filtered texture images, 7x7 feature generation windows are cropped to produce 49-dimensional vectors and modeled via IMoFA-L. The conspicuity maps thus obtained are combined to predict the exact location of the facial feature. For the test images, we extract such 49-dimensional feature vectors at each point in the 19x19 search window, and select the location that produces the highest likelihood as the target fiducial point. The texture image by itself outperforms the depth image, but the combination results in a slight increase over the texture-only scenario.

3. EXPERIMENTAL RESULTS

3.1. Database and Testing Methodology

We have employed the University of Notre Dame (UND) database in all our experiments[8]. The data consists of 942 2D images and the corresponding registered 3D point cloud data, both at resolution 480x640. The ground truth is created by manually landmarking seven points, that is, four eye corners, nose tip and mouth corners. The data are randomly split into three disjoint parts, i.e. the training (472 samples), validation (235 samples) and test sets (235 samples). The validation set is used to tune the model parameters of the IMoFA-L. The IGF and DCT-based methods do not need a validation set, hence add these samples to the training set.

To measure the performance of the feature localizers we have used a normalized distance, by dividing localization, error, measured as Euclidean distance in terms of pixels, to the inter-ocular distance. A landmark is considered correctly detected if its deviation from the true landmark position is less than a given threshold, called the *acceptance threshold.*

All proposed algorithms proceed by facial feature localization on downsampled (60x80) face images, followed by a refinement on the corresponding high-resolution (480x640) images. In the refinement stage, the search proceeds with a 19x19 window from around the coarse localization results.

3.2. Performance of Feature Localizers

The performance of the DCT-based, IGF-based and factor analysis-based facial feature localizers are illustrated in Figure 5 and Figure 7 for the seven landmarks. Notice that the performance is given in terms of normalized distance. The horizontal axis indicates the distance threshold beyond which the location is assumed to be incorrect, whereas the vertical axis shows the correct localization percentage.

The performance of the DCT-based algorithm is given in Figure 5, where the effect of the chosen number of DCT coefficients illustrated. In the coarse level, 8x8 DCT blocks are calculated at each facial point. Three tests were performed with 20, 35 and 42 coefficients, respectively, always selected from the upper triangle of the DCT matrix. In the fine stage, 16x16 DCT blocks are calculated around each candidate facial landmark, from which choices of 54, 135 and 177 coefficients were tested. As we increase the number of DCT coefficients, the localization accuracy increases as well, albeit with diminishing returns. This improvement is valid for all landmarks (See Figure 5) except for the nose tip. Increasing the number of DCT coefficients does not improve the performance of the DCT method for the nose tip.

In Table 1 we show the average coarse and fine landmark localization errors for IMoFA-L (2D+3D), DCT (2D and 3D) and IGF (2D) methods. IMoFA-L method with structural correction gives the best results for coarse localization of the facial features. All coarse methods except DCT on depth images give comparable localization results. Although the depth images result in a poorer average localization, the nose tip and inner eye corners are better localized, as the discriminative ravine and peak areas on the depth images correspond to the nose tip and inner eye corners (See Figure 7).

In fine localization part we compare our DCT algorithm with Lades's [21] and Wiskott's[32] Bunch-based methods. Figure 6 illustrates the comparison results. Lades and Wiskott methods are applied to search window which is located around the true landmark positions, our DCT algorithm, which is applied to search window located around the coarse landmark positions outperforms the Lades and Wiskott local feature analysis schemes.

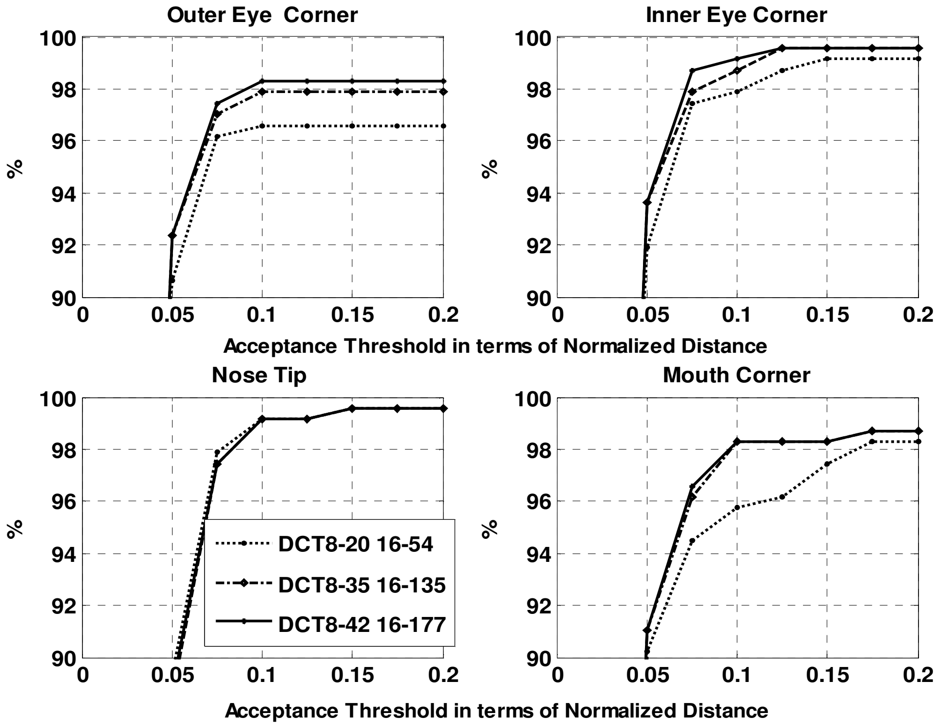

Figure 5. Tuning of the DCT method: results of the fine localization for each landmark type. Legend: DCT8-20 16-54 should be read as: 8x8 coarse level DCT transform and 20 out of 64 coefficients are selected; 16x16 fine level DCT transform and 54 out of 256 coefficients are selected.

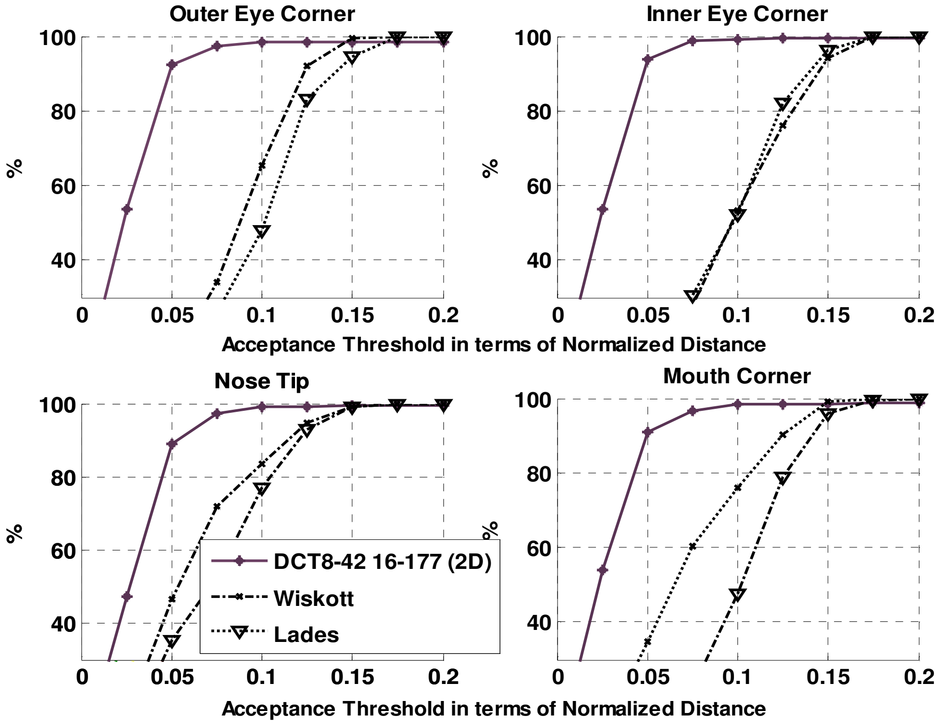

Figure 6 Comparison of fine refinement performances of DCT, Lades and Wiskott

Table 1. Average coarse and fine localization errors in terms of Euclidean pixel distances on the original images.

Method	IMoFA-L (2D+3D)	DCT-8-42, DCT-16-177 + SVM (2D)	IGF 7-100 IGF 11-150 +SVM	DCT-8-42, DCT-16-177 + SVM (3D)
Coarse Localization (60x80 face images)	4.32 pixels	5.04 pixels	6.08 pixels	10.48 pixels
Fine Localization (480x640 face images)	5.08 pixels	2.80 pixels	4.36 pixels	7.81 pixels

In fine localization part, the IGF method was tested on 2D images, with a patch size of 11x11. The DCT method has been observed to outperform the IGF method for all window sizes. The IMoFA-L method has a better accuracy on the lower resolution images, hence provides good initial seeds for landmarks, but its performance is relatively poor for higher resolution images. In other words, its refining stage does not function as well as the DCT or the IGF methods. The main reason for this is the limited availability of training data, as on the higher resolution images, a larger number of samples is necessary for the training of the higher-dimensional generative model.

Finally, we have investigated feature localization performance on the pure 3D depth images, i.e., without appearance images. Both the DCT and IMoFA-L algorithms have shown poorer localization performance as compared to 2D texture images. As pointed out before, a combination of 2D and 3D is more beneficial for the IMoFA-L case, where 3D plays an assist role in validating the landmarks.

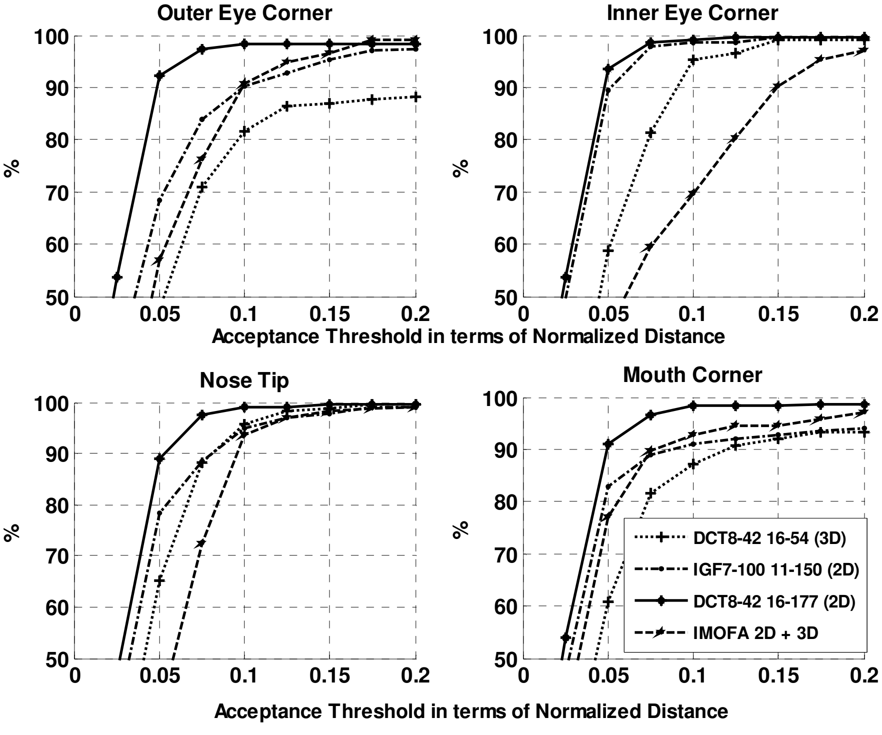

Figure 7. Comparison of fine refinement performances of DCT, IGF and IMoFA-L based methods on 2D and 3D data. Legend: IGF7-100 11-150 should be read as: 7x7 coarse level IGF vector and dimension reduced from 7x7x12 to 100 via PCA; 11x11 fine level IGF vector and dimension reduced from 11x11x8 to 150 via PCA. IMoFA 2D+3D is a combination of the conspicuity maps obtained separately from 2D Gabor features and 3D depth image.

In Figure 7 we can see the fine localization accuracy of the proposed algorithms for facial feature localization. The inner eye corners and mouth corners are relatively tougher to localize. There is a significant difference in the localization performance of different algorithms for different landmark types. However, the DCT method, appropriately tuned, seems to be the winner.

4. CONCLUSIONS

We have proposed three schemes for coarse-to-fine facial landmarking that rely on local 2D and 3D feature information. Gabor wavelet features and DCT coefficients were employed successfully to drive the localization process. An unsupervised mixture of factor technique is contrasted (IMOFA) with the supervised SVM classifiers. (IGF and DCT). We have also proposed a robust structural correction scheme to analyze arbitrary but consistent configurations of landmarks.

We have found that a coarse localization can be efficiently implemented using either Gabor or DCT features. On the other hand, the end-to-end DCT technique seems to be superior in the final refined stage feature locations. The 2D information has proven to be more useful in landmark localization as compared to the 3D information at the same resolution. The 3D information must then be relegated to an assist role in eliminating background clutter or in initializing some of the pronounced landmarks (such as the nose tip).

We have used no specific heuristics for any of the landmarks. In other words the algorithms were all streamlined for all or any of the landmarks. The only requirement is that the landmark should contain sufficient statistical information to drive local feature learning process, and should not be just semantically specified (e.g. middle of cheek). Our local feature analyzers create large basins of attraction around target landmarks. In [27] we have compared our local search methods with with several well-known approaches from the literature[21,32]. In both [21] and [32], Gabor features are extracted from a single point (in contrast to a feature generation window that we have used), with five scales and eight different orientations. [21] uses the magnitudes, resulting in 40-dimensional feature vectors, whereas in [32] both magnitude and phase are used to produce 80-dimensional vectors. The Gabor jet-based methods consult a bunch graph as their template library. We have shown that our methods produced better results, even when phase information was used along with the magnitude in Gabor feature[27].

The possible extensions to this model include illumination compensation as a first step in preprocessing, and the incorporation of other 3D descriptors. Shape can be more informative, but it comes with a computational cost, which must be justified by an appropriate increase in accuracy.

5. REFERENCES

1. ANTONINI (G.), POPOVICI (V.), THIRAN (J.P.), Independent Component Analysis and Support Vector Machine for Face Feature Extraction, *4th Int. Conf. on Audio- and Video-Based Biometric Person Authentication*, 2003.
2. ARCA (S.), CAMPADELLI (P.), LANZAROTTI (R.), An efficient method to detect facial fiducial points for face recognition, in *Proc. 17th Int. Conf. on Pattern Recognition,* 2004.
3. BARTLETT (M.S.), MOVELLAN (J.R.), SEJNOWSKI (T.J.), Face Recognition by Independent Component Analysis, *IEEE Transactions On Neural Networks,* **13**, n°6, 2002.
4. BASKAN (S.), BULUT (M.M.), ATALAY (V.), Projection Based Method for Segmentation of Human Face and its Evaluation, *Pattern Recognition Letters*, **23**, pp. 1623–1629, 2002.
5. BOEHNEN (C.), RUSS (T.), A Fast Multi-Modal Approach to Facial Feature Detection, *Proc. 7th IEEE Workshop on Applications of Computer Vision*, pp. 135-142, 2005.
6. BURGES C.J.C., A tutorial on Support Vector Machines for Pattern Recognition, *Data Mining and Knowledge Discovery,* **2**, pp. 121-167, 1998
7. CAMPADELLI (P.), LANZAROTTI (R.), Localization of Facial Features and Fiducial Points, in *Proc. IASTED Int. Conf. VIIP*, 2002.
8. CHANG (K.I.), BOWYER (K.W.), FLYNN (P.J.), Multi-modal 2D and 3D Biometrics for Face Recognition, in *Proc. IEEE Workshop on Analysis and Modeling of Faces and Gestures,* 2003.
9. CHEN (L.), ZHANG (L.), ZHANG (H.), ABDEL-MOTTALEB (M.), 3D Shape Constraint for Facial Feature Localization Using Probabilistic-like Output, in *Proc. 6th IEEE Int. Conf. on Automatic Face and Gesture Recognition*, 2004.
10. COLBRY (D.), STOCKMAN (G.), JAIN (A.K.), Detection of Anchor Points for 3D Face Verification, in *Proc. IEEE Workshop on Advanced 3D Imaging for Safety and Security, A3DISS,* San Diego, CA., 2005.

11. CRISTINACCE (D.), COOTES (T.), SCOTT (I.), A Multi-Stage Approach to Facial Feature Detection, in *Proc. 15th British Machine Vision Conference*, pp. 277-286, 2004.
12. EKENEL (H.K.), SANKUR (B.), Feature Selection in the Independent Component Subspace for Face Recognition, *Pattern Recognition Letters,* **25**, pp. 1377-1388, 2004.
13. EKENEL (H.K.), STIEFELHAGEN (R.), Local Appearance Based Face Recognition Using Discrete Cosine Transform, *13th European Signal Processing Conf.*, 2005.
14. FERIS (R.S.), GEMMELL (J.), TOYAMA (K.), KRÜGER (V.), Hierarchical Wavelet Networks for Facial Feature Localization, in *Proc. 5th IEEE Int. Conf. on Automatic Face and Gesture Recognition*, 2002.
15. GOURIER (N.), HALL (D.), CROWLEY (J.L.), Facial Features Detection Robust to Pose, Illumination and Identity, *IEEE Transactions on Systems, Man and Cybernetics*, 2004.
16. GU (H.), SU (G.), DU (C.), Feature Points Extraction from Faces, *Image and Vision Computing*, 2003.
17. GUNDUZ (A.), KRIM (H.), Facial Feature Extraction Using Topological Methods, in *Proc. IEEE Int. Conf. on Image Processing*, **1**, 2003.
18. HAMOUZ (M.), KITTLER (J.), KAMARAINEN (J.-K.), PAALANEN (H.), KÄLVIÄINEN (H.), MATAS (J.), Feature-Based Affine-Invariant Localization of Faces, *IEEE Trans. PAMI*, **27**, no 9, 2005.
19. HERPERS (R.), MICHAELIS (M.), LICHTENAUER (K.-H.), SOMMER (G.), Edge and Keypoint Detection in Facial Regions, in *Proc. 2nd Int. Conf. on Automatic Face and Gesture Recognition*, pp. 212-217, 1996.
20. IRFANOĞLU (M.O.), GÖKBERK (B.), AKARUN (L.), 3D Shape-Based Face Recognition Using Automatically Registered Facial Surfaces, in *Proc. Int. Conf. of Pattern Recognition,* **1**, pp. 183-186, 2004.
21. LADES (M.), VORBRUGGEN (J.), BUHMANN (J.) LANGE (J.), VON DER MALSBURG (C.), WURTZ (R.), KONEN (W.), Distortion Invariant Object Recognition in the Dynamic Link Architecture, *IEEE Transactions on Computers*, **42**, 1993.
22. LAI (J.H.), YUEN (P.C.), CHEN (W.S.), LAO (S.), KAWADE (M.), Robust Facial Feature Point Detection under Nonlinear Illuminations, in *Proc. IEEE ICCV Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems*, 2001.
23. LIU (C.), WECHSLER (H.), Independent Component Analysis of Gabor Features for Face Recognition, *IEEE Transactions on Neural Networks*, **14**, pp. 919-928, 2003.
24. PAN (Z.), BOLOURI (H.), High Speed Face Recognition Based on Discrete Cosine Transforms and Neural Networks, Technical Report, University of Hertfordshire, UK, 1999.
25. RYU (Y.S.), OH (S.Y.), Automatic Extraction of Eye and Mouth Fields from a Face Image Using Eigenfeatures and Ensemble Networks, *Applied Intelligence,* **17**, pp.171–185, 2002.
26. SALAH (A.A.), ALPAYDIN (E.), Incremental Mixtures of Factor Analyzers, in *Proc. Int. Conf. on Pattern Recognition*, **1**, pp. 276-279, 2004.
27. SALAH (A.A.), ÇINAR AKAKIN (H.), AKARUN (L.), SANKUR (B.), Exact 2D-3D Facial Landmarking for Registration and Recognition, submitted for publication.
28. SHIH (F.Y.), CHUANG (C.), Automatic Extraction of Head and Face Boundaries and Facial Features, *Information Sciences*, **158,** pp. 117-130, 2004.
29. SMERALDI (F.), BIGUN (J.), Retinal Vision Applied to Facial Features Detection and Face Authentication, *Pattern Recognition Letters*, **23**, pp. 463-475, 2002.
30. SOBOTTKA (K.), PITAS (I.), A Fully Automatic Approach to Facial Feature Detection and Tracking, in BIGUN (J.), CHOLLET (G.), BORGEFORS (G.) (eds.), *Audio- and Video-based Biometric Person Authentication*, *LNCS*, **1206**, pp. 77-84, Springer Verlag, 1997.
31. WANG (Y.), CHUA (C.), HO (Y.), Facial Feature Detection and Face Recognition from 2D and 3D Images , *Pattern Recognition Letters*, **23**, n°10, pp. 1191-1202, 2002.
32. WISKOTT (L.), FELLOUS (J.-M.), KRÜGER (N.), VON DER MALSBURG (C.), Face Recognition by Elastic Bunch Graph Matching, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **19**, n°7, pp. 775-779, 1997.
33. WONG (K.), LAM (K.), SIU (W.), An Efficient Algorithm for Human Face Detection and Facial Feature Extraction under Different Conditions, *Pattern Recognition*, **34,** pp. 1993-2004, 2001.
34. XUE (Z.), LIB (S.Z.), TEOH (E.K.), Bayesian Shape Model for Facial Feature Extraction and Recognition, *Pattern Recognition*, **36**, pp. 2819-2833, 2003.
35. YANG, (M.), AHUJA (N.), KRIEGMAN (D.), Face Detection Using Mixtures of Linear v Subspaces, in *Proc. 4th Int. Conf on Automatic Face and Gesture Recognition*, pp. 70-76, 2000.

36. ZOBEL (M.), GEBHARD (A.), PAULUS (D.), DENZLER (J.), NIEMANN (H.), Robust Facial Feature Localization by Coupled Features, in *Proc. 4th IEEE Int. Conf. on Automatic Face and Gesture Recognition*, 2000.
37. ZHU (X.), FAN (J.), ELMAGARMID (A.K.), Towards Facial Feature Extraction and Verification for Omni-Face Detection in Video-Images, *Image Processing,* **2**, pp. 113-116, 2002.