Volume 2, Issue 5, May 2012 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Robust Facial Land marking for Registration Robust Localization de Points

A. Nageswaran[1], Research Scholar CMJ University, Shilliong,
Dr.Sidappa Naidu[2] , Principal Veltech Multitech Engineering College, Chennai

Abstract: Finding landmark positions on facial images is an important step in face registration and normalization, for both 2D and 3D face recognition. In this paper, we inspect shortcomings of existing approaches in the literature and propose a method for automatic land marking of near-frontal faces. The first level of the two-tier method analyzes faces on a coarse scale and employs Gabor features over eight orientations. The second level improves the accuracy of facial feature locations, as inherited from the first tier, by using Discrete Cosine Transform (DCT) coefficients. In addition, a structural analysis subsystem relates the feature positions in an elastic graph, determines outlier positions and can correct them with back-projections. Tests carried on two separate face databases have indicated that our proposed s scheme is robust and outperforms its competitors.

Keywords: 2D, 3D, Preprocessing, DCT Feature Matching, Structural Analysis

Introduction

Facial feature localization is a critical first step in many subsequent tasks, such as face recognition, pose normalization, expression understanding and face tracking. Although there is not yet a general consensus, the fiduciary or first tier facial features, most often cited in the literature, are the eyes or the eye corners, the tip of the nose and the mouth corners. Similarly, second tier features are the eyebrows, the bridge of the nose, the tip of the chin, etc. The importance of facial features stems from the fact that most face recognition algorithms in 2D and/or 3D rely on accurate feature localization. This step is critical not only directly for recognition techniques based on features themselves, but indirectly for the global appearance-based techniques that necessitate prior image normalization. For example, in 2D face recognition, popular recognition techniques, such as eigenfaces or Fisher faces, are very sensitive to registration and scaling errors. For 3D face recognition, the widely used iterative closest point (ICP) registration technique requires scale-normalized faces and a fairly accurate initialization. In any case, both modalities require accurate and robust automatic land marking.

Schemes Based on 2D Information

The various approaches in the literature for facial feature localization can be classified as appearance-based, geometric-based and structure-based. The majority of approaches uses a preprocessing stage for initial coarse localization, using horizontal and vertical gray-level or edge field projections, followed by some histogram valley detection filter. A second commonality between methods is that most use a coarse-to-fine localization to reduce the computational load. Some algorithms employ a skin colour-based scene segmentation to detect the face first and further colour segmentation for lip detection.

Appearance-based approaches aim to find basis vectors to represent the face and its facial features. Examples of transformations used are principal components analysis (PCA), Gabor wavelets independent components analysis (ICA) discrete cosine transforms (DCT)and Gaussian derivative filters. These transform features capture and model facial features under statistical variability when selected and processed with machine learning techniques like boosted cascade detectors, support vector machines (SVM) and multi-layer perceptrons (MLP).

Schemes Based on 3D Information

3D information is not commonly used in finding facial fiducial points, since 3D face imaging and handling of the resulting data volume are still not mainstream techniques. Furthermore outlier noise makes reliable processing difficult. However, when the scale of the faces are known, ICP can be used to register the face to a 3D template, thereby greatly constraining the possible locations for each facial landmark

The method with 40-dimensional 2D Gabor jets is extended to a 324-dimensional 3D jet method (36-dimensional point signatures from a 3x3 neighbourhood of each landmark). Colbry et al. employ surface curvature-based shape indices under geometrical constraints to locate features on frontal 3D faces .Their method has been generalized to the multi-pose case with the aid of 2D information, such as the output of Harris corner detector on the gray-level information and related geometrical constraints. Conde et al. use SVM classifiers trained on spin images for a purely 3D approach. As their proposed method requires great computational resources, they constrain the search for the landmarks by using a priori knowledge about the face. In the 3D information plays a secondary or support role, in filtering out the background, and to compute intra-feature distances in geometry-based heuristics. In 3D information is used to assist 2D in filtering out the background, and a comparison between 2D and 3D methods under relatively controlled illumination conditions indicates superiority of the 2D approaches. A summary of methods based on 3D information, totally or partially...

Algorithm-1: Gabor Factor Analysis and Structural Information

The first algorithm takes a generative modeling approach for landmark localization at a coarse stage. By modeling the distributions of local features with a state-of-the-art unsupervised model we achieve robust localization. Searching for each landmark independently protects the system against misleading local similarity values caused by missing or occluded landmarks. Figure 1 summarizes the landmarking scheme where a coarse localization is performed using the Gabor factor analysis algorithm, followed by a structural correction.

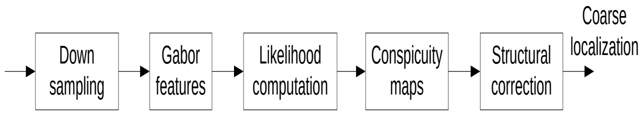

Figure 1. Coarse landmark localization starts with computation of Gabor wavelets

Gabor wavelets are often used in the literature to extract local discriminating features. The image is initially convolved with a Gabor kernel as below:

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

$$\vec{k}_j = (k_{jx}, k_{jy}) = (k_v\cos\varphi_w, k_v\sin\varphi_w),\ k_v = 2^{-\frac{v+2}{2}}\pi,\ \varphi_w = w\frac{\pi}{8} \quad (1)$$

where $\vec{x} = (x, y)$ is the given pixel, $j = w + 8v$, and (w,v) 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, $e^{-\sigma^2/2}$ in the square brackets compensates for the DC value.

Preprocessing

To reduce the computational burden, we down sample the high resolution face images. For the UND database, we have used a down sampling factor of eight (from the size 480x640 down to 60x80), which is a good compromise between precision and computational parsimony. An additional benefit of down sampling is the dramatic reduction in the number of candidate points to be tested, from approximately 300,000 to 1,500-2,000 with the addition of 3D masking. We use the 7x7 neighborhood as a feature generation window, producing 49-dimensional feature vectors for each point at each orientation in the coarse scale (See Figure 3 for the size of the feature generation window).

Algorithm-2: DCT-based Facial Feature Extraction

In our second approach we have used various frequency bands DCT features DCT coefficients can also capture the statistical shape variations and can form competitor features to local Gabor features in both computation effort and accuracy of localization.

Preprocessing

The coarse level process described in is used as the first step; in other words, the DCT scheme inherits the outcome of the Gabor-based coarse-level localizer. An area of size 19x19 around the coarse landmark is probed with DCT windows. In other words, each of the 361 points around the coarse landmark is inspected using a 15x15 feature patch. Each such transplanted patch undergoes DCT transformation and selected bands of the coefficients (subset of 225 coefficients) are extracted as fine localizing features. More specifically, given an image block from the search window $f(x, y)$, where $y, x = \{0, 1, ..., K-1\}$, we decompose it in terms of orthogonal 2-D DCT basis functions. The result is a matrix $C(v, u)$ containing DCT coefficients:

$$C(v,u) = \alpha(v)\alpha(u)\sum_{y=0}^{K-1}\sum_{x=0}^{K-1} f(y,x)\beta(y,x,v,u) \text{ for } v, u = 0, 1, ..., K-1, \quad (2)$$

The DCT coefficient values can be regarded as the relative presence of 2D spatial patterns contained in the visited locality. Once the coefficients of the DCT output matrix are computed, they are re-organized in a zigzag-scanning pattern. This selection favors the high-energy, low-frequency coefficients, 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. . The first 119 coefficients out of 224 (not 225 since DC value is removed) are chosen and z-normalized to form the feature vector. These chosen coefficients belong to the left upper diagonal of the DCT block.

DCT Feature Matching

The procedure for refining the localization is based on identifying the best matching stored feature best matching to the actual data. The information in the training is summarized succinctly by using a clustering approach. The training data consists of the selected DCT coefficients from the 15x15 patches of 707 faces placed on the dead center of features. DCT block size is chosen in order to represent only the desired facial feature in that block. In the training phase, we have used k-means clustering on the z-normalized DCT coefficients to obtain eight codebook vectors (k=8), one separate codebook for each of the 7 facial landmarks. Average cluster population is 88.5 and standard deviation of cluster population is about 21.

Experimental Results

A good feature detector should have high accuracy and should converge to the true feature starting from any point within a large basin of attraction. If the correct localization probability flattens out after an initial monotonic growth with the increasing size of the search window, this indicates the robustness of the scheme.

We compare our method with two other methods for facial feature localization: Lades et al. and Wiskott et al.**Error! Reference source not found.**. Lades et al. have employed a similarity measure that is based on the comparison of Gabor wavelet jets, with five scales and eight different orientations, resulting in 40-dimensional feature vectors overall. Their proposed jet similarity measure is

$$S(J,J')=\frac{\sum_j a_j a_j'}{\sqrt{\sum_j a_j^2 \sum_j a_j'^2}}$$

where a_j denotes the magnitude of the Gabor wavelet for a particular orientation and frequency. Wiskott et al. extend this in **Error! Reference source not found.** by considering the phase information as well, but they especially compensate for sharp changes induced by phase rotation:

$$S_\phi(J,J')=\frac{\sum_j a_j a_j' \cos(\phi_j-\phi_j'-\vec{d}\,\vec{k})}{\sqrt{\sum_j a_j^2 \sum_j a_j'^2}} \qquad (4)$$

Where ϕ_j is the phase, and $\vec{d}\,\vec{k}$ denotes the displacement proportional with the phase rotation rate. Both methods store the features extracted from the training set as bunch models. During testing, features extracted from the test image are compared with all templates in the bunch, and the minimum distance to any template is used for candidate selection. In our comparisons, the bunch graph was learned from the training set, and the similarity measures were evaluated exhaustively within a maximum neighborhood around the true landmark.

The Gabor Factor Analysis Approach

For the Gabor factor approach as explained we down sampled the images of the UND dataset down to size 60x80, and then, obtained mixture distributions of Gabor features. The test images were similarly processed. The likelihood of pixels to correspond to a facial landmark is computed using Eq.3 of This likelihood is called the feature conspicuity map, and is obtained by summing the eight likelihood maps, one for each orientation. The per-orientation conspicuity maps for the left eye outer corner are illustrated in Figure , while the seven inset images in Figure 3(b) show the total (summed) conspicuity maps for each of the seven landmarks. For any one of the landmarks, the peak of the summed conspicuity map is taken as the identified facial landmark. All the identified landmarks are pictured in figure 2.

Structural Analysis

The structural analysis subsystem can be viewed as a hypothesis testing labeling step. Given the seven landmark candidates, the role of the structural analysis is to validate these candidates based upon configurationally information. The permutations of landmark triplets are tested in descending order given by the product of reliabilities of landmarks in the corresponding support sets. Our simulations have shown that three is the optimum number of landmarks for support set size.

Figure 2. Likelihood based conspicuity maps obtained from IMoFA-L outputs for the outer corner of the left eye. Lighter locations indicate higher values.

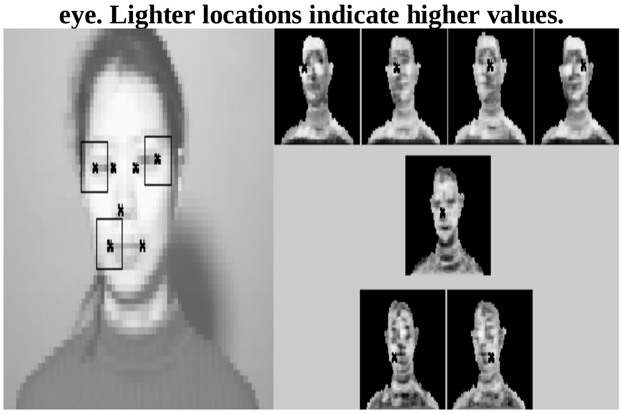

Figure 3. a) Left image: The correctly located landmarks on the downsampled intensity image; 7x7 neighbourhoods

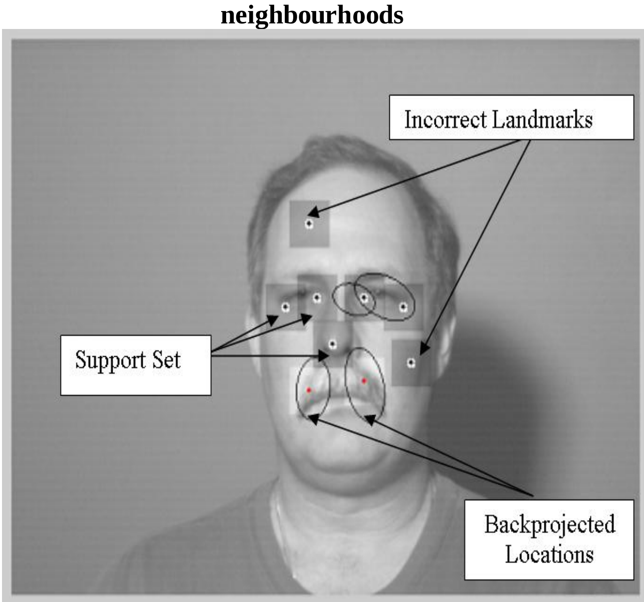

Figure 4. The coarse localization results projected to the original image.

Experiments in Challenging Environments

To test the applicability of our method in a more challenging environment, we have performed additional tests on a subset of the English part of the BANCA dataset This dataset contains significant illumination, pose, and background and face clutter variations, such as eyeglasses and hair, and presents major challenges for feature localization. From the BANCA dataset, three representative sessions with different environmental conditions were selected. 50 subjects from each session contributed one image for training, one image for validation and one image for the test set.

For the coarse level, all parameter settings were retained. Images are down sampled by eight (from to 576x720 to 72x90) and eight Gabor channels are used. Search area is the coarsely located face. Figure 5shows coarse localization examples on the down sampled faces for three sessions that increase in difficulty. It is observed that when eyeglasses occlude the eye corners, facial landmarks cannot be recovered. The structural correction re-estimates 350 landmark locations in 150 test samples.

Figure 5. Samples from sessions 1, 5, and 10 of the English part of the BANCA dataset.

Conclusions and Future Work

In this paper, we have introduced and tested a two-tier method of facial land marking. The coarse level features are extracted by Gabor wavelets, and then each facial feature is learned by a separate model. The configurationally information of the landmarks is also integrated into the localization method. The robustness and satisfactory performance of the coarse land marking are due in part to the elimination of outlier positions with configurationally information, and partly due to observing the image with 56x56 windows (7x7 after down sampling), which are sizes commensurate with the facial features.

The structural subsystem is a used as a post-processing block to detect and correct localization failures. We should also remark that the proposed method is not necessarily specific to faces, but is applicable to any other objects and defined feature sets. One byproduct of the structural analysis is that it can in principle locate facial features that are not rich in texture, like cheeks, chin, or features under total occlusion of a facial accessory.

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*, pp. 111-118, Guildford, UK, 2003.

[2] ARCA (S.), CAMPADELLI (P.), LANZAROTTI (R.), A face recognition system based on automatically determined facial fiducial points, *Pattern Recognition*, **39(3),** pp.432-443**,** 2006.

[3] The BANCA dataset - English part; http://banca.ee.surrey.ac.uk/.

[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, *7th IEEE Workshop on Applications of Computer Vision*, pp. 135-142, Breckenridge, USA, 2005.

[6] BOLME (D.S.), Elastic Bunch Graph Matching, unpublished MS thesis, Colorado State University, 2003.

[7] BRUNELLI (R.), POGGIO (T.), Face Recognition: Features versus Templates, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **15**, pp. 1042-1052, 1993.

[8] BURL (C.), PERONA (P.), Recognition of Planar Object Classes, in *Computer Vision and Pattern Recognition Conference*, pp. 223 - 230, San Francisco, USA, 1996.

[9] CAMPADELLI (P.), LANZAROTTI (R.), Localization of Facial Features and Fiducial Points, in *Proc. IASTED Int. Conf. Visualization, Imaging & Image Processing*, Malaga, Spain, 2002.

[10] CHANG (K.I.), BOWYER (K.W.), FLYNN (P.J.), Multi-modal 2D and 3D Biometrics for Face Recognition, in *IEEE Workshop on Analysis and Modeling of Faces and Gestures*, pp.187-194, Nice, France, 2003.

[11] CHEN (L.), ZHANG (L.), ZHANG (H.), ABDEL-MOTTALEB (M.), 3D Shape Constraint for Facial Feature Localization Using Probabilistic-like Output, in *6th IEEE Int. Conf. on Automatic Face and Gesture Recognition*, pp.302-307, Seoul, Korea, 2004.