A Combined Feature-Texture Similarity Measure for Face Alignment Under Varying Pose

Lixin Fan and Kah Kay Sung
School of Computing
National University of Singapore
Singapore, 117543
{fanlixin, sungkk}@comp.nus.edu.sg

Abstract

We formulate face alignment as a model-based parameter estimation problem in this paper. First, we work within a framework that combines two separate subspace models to represent frontal face patterns and pose change independently. The combined unified non-linear model represents varying pose faces with a complex manifold. Then, we use a feature based similarity measure(FBSM) to evaluate image differences in terms of pose, and match unknown pose faces with the model image using a combined feature-texture similarity measure(FTSM). Noticeable properties of the combined FTSM include (1) its sensitivity to spatial differences between feature points in two images, which is crucial to aligning two initially faraway poses; (2) easy determination of hill-climb directions in parameter space, without computing gradients of error functions. Experimental results demonstrate that, in the absence of significant clutter, a face alignment algorithm using the combined FTSM, can reliably align varying pose faces under different lighting conditions, even when initial poses are far off.

1 Introduction

Face images are affected by many factors including face appearance, lighting condition and pose. To measure how similar a given image resembles a generic face, one must take into account image variations caused by all these factors *separately*. For instance, the well-known eigenface method captures face appearance variations within a linear feature subspace, and uses the distance between the given image from the feature subspace (DFFS) as a difference measure between input pattern and the face class. The method was proven efficient for frontal (or other fixed pose) faces [1][2][3]. In order to represent frontal as well as non-frontal faces variations, Beymer, Jones, Vetter and Poggio [4][5][6], Craw[7], and Cootes and Taylor[8] proposed a *non-linear* model, which combines two parametric linear subspace models (one for face appearances, the other for varying pose) by *image warping*, to represent the distribution of varying pose faces. With the proposed model, one can synthesize a given pose face as a *warped* linear combination of frontal face images, where the warping displacement field is constrained by a subspace model learnt from face images under varying pose. Alignment involves minimizing an appropriate *distance measure* between the original image and the synthesized image.

Existing techniques in [4][5][6][7][8] use an iterative procedure to estimate both texture and pose parameters (see Section 2 for definitions). During each iteration, the synthesized image is compared with the original image using a *pixel-wise intensity difference measure*. We note that using intensity difference measure to quantify structure differences and estimate pose parameters is inappropriate, because *intensity* difference and *structure* difference are, although sometimes related, essentially two different physical concepts. For instance, when faces change in pose, one may only find minor changes in intensity at most pixels, but large amount of spatial movement of some feature points (e.g. eyes, nose, and mouth). Therefore, we propose to use a feature based similarity measure(FBSM) to take into account the *spatial* differences between feature points of two poses, augmenting the intensity difference measure used in previous methods. In this paper, we show that the combined feature-texture similarity measure(FTSM) is sensitive to pose differences between two face images, and a face alignment algorithm using the combined FTSM, can reliably align varying pose faces under different lighting conditions, even when initial poses are far off [1]. Another noticeable property of the FTSM is that, rather than using a stochastic gradient descent algorithm to estimate pose parameters (like in [4]), the feature point

1063-6919/00 $10.00 © 2000 IEEE

correspondence map can be used to directly determine the best hill-climb directions in pose parameter space without computing gradients of error functions (see Section 3.1).

We also note that mutual information based face alignment methods [9][10] have been shown to handle inaccurate initial estimates. Nevertheless, in these approaches, one usually, although not necessarily, use 3D face models which can be hard to acquire.

2 Varying Pose Face Model

We adopt a varying pose face model that combines two separate subspace models representing frontal face and pose change *independently*. Pentland *et al.* and others approximate a frontal face texture with a linear combination of eigenfaces [1][2]:[2]

$$I \approx \widehat{I} = \Phi_T \cdot \alpha \tag{1}$$

where Φ_T is a eigenface matrix obtained by applying PCA on training face images and discarding small eigen vectors. The transformation coefficient $\alpha = \Phi'_T \cdot I$ determines texture variations caused by face appearance as well as lighting change in reconstructed faces. We will refer to α as *texture* parameters below. Distance From Feature Space(DFFS) between I and $\widehat{I}$ is then given by :

$$\epsilon^2(I) = \|I\|^2 - \left\|\widehat{I}\right\|^2 = \|I\|^2 - \|\alpha\|^2 \tag{2}$$

On the other hand, to model spatial movement of image pixels, we define below a *feature based image warping* process, which first finds dense pixelwise correspondences by interpolating a computed sparse *feature points correspondence map* (FPCM), then shifts image pixels accordingly:

$$I_0 \rightarrow I : I = I_0 \circ C \tag{3}$$

where I_0 is the image to be warped, C is a feature points correspondence map and $\circ$ denotes a warping process (see Figure 1). We refer to [11] for detailed procedures.

To learn image variations caused by pose change, a linear subspace of dense correspondence maps is

[1]In this paper, we assume that face images are appropriately centered and normalized in size beforehand, and we only need to align faces under rotations.

[2]We use notations that (1) all m-by-n images are lexicographically ordered to form image vectors $I \in \Re^{N=mn}$; (2) images are subtracted by the average of training images; (3) "′" represents transpose of matrix.

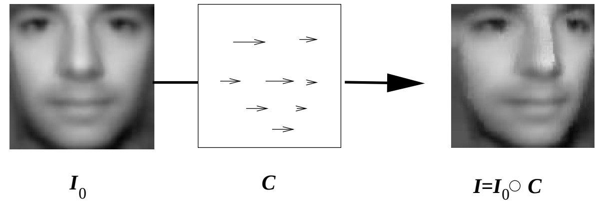

Figure 1: Feature based image warping. Original image I_0, feature points correspondence map C, and corresponding warped image I are shown from left to right.

Figure 2: Left to right: Reference image I_0, and warped images with first three eigenvectors, given pose parameter=1000 ($I_0 \circ 1000E_i$, $i = 1, 2, 3$).

constructed by applying PCA on example interpolated FPCMs, which were obtained by manual registration of face images under varying pose[3]. Figure 2 illustrates an example of warped images with first three eigenvectors of the learnt subspace model. Note that these warped images roughly correspond to three types of pose changing (i.e. in depth left-right rotation, in depth up-down rotation, and in plane rotation). With this model, arbitrary pose variations from the frontal pose can be approximated as a linear combination of these basic warping operations [4][5][6][7][8]. In mathematical terms, the associated FPCM (C) represented by a $\Re^{2N}$ vector, is approximated by[4]:

$$C \approx \widehat{C} = \Phi_P \cdot \gamma \tag{4}$$

where N is the number of feature points, Φ_P is eigenvector matrix obtained by applying PCA on training FPCMs and discarding small eigenvectors. The transformation coefficient vector $\gamma = \Phi'_P \cdot C$ is referred as *pose* parameters. Similarly, one may want to compute the distance between C and $\widehat{C}$. In Section 3.1, we present a feature-based similarity measure used to quantify the differences between FPCMs.

[3]In the learning phase, FPCMs are found by manual registration for all example images. In pose alignment phase, FPCMs are found by a binary edge map matching method, see Section 3.1.

[4]Again, FPCMs are subtracted by the average of training FPCMs.

Combining linear subspaces of face texture and pose change, one can model a new face image as:

$$I \approx I^M = (\Phi^T \alpha) \circ (\Phi^P \gamma) \tag{5}$$

where Φ^T, Φ^P and α, γ are defined as above. In Section 3.1, we introduce a feature-texture similarity measure to find the distance between I and I^M.

3 Combined Feature-Texture Similarity Measure and Pose Alignment

The proposed model captures both texture and pose variations for face images (Equation 5). In order to measure the difference between a given face image (I) and its model approximation (I^M), we need a similarity measure taking into account both texture and pose differences. Unfortunately, traditional *intensity based similarity measures* (such as Sum of Squared Difference, normalized correlations etc.) of misaligned faces do not reflect pose differences well, since they are poor at reflecting the spatial differences between distinctive feature points. On the other hand, *feature based similarity measures* (e.g. Hausdorff measure [13]) capture geometric differences between distinctive feature points, but completely ignore image intensity differences. In the rest of this section, we first introduce a feature based similarity measure(FBSM) to quantify structural(pose) variation in face images, and derive a feature-texture similarity measure combining both an intensity based measure and a FBSM. Finally, we outline a face alignment algorithm using the combined similarity measure.

3.1 Using Image Features for Pose Alignment

A. Feature Point Correspondence Maps

To find the feature point correspondence map between one image I_1 and another image I_2, one need to match similar feature points between the two images. Unlike classical optical flow algorithms that obtain correspondence based on *image flow brightness constraints* [12], an ideal feature matching method should be *lighting invariant* to capture only structure differences between the two images. In general, any matching method having this property will suit our purpose. In this paper, we employ a simple *binary edge map matching* method, which consists of following steps:

1. Compute the binary edge maps E_1 and E_2 fo the two given images;

2. For each edge point f_i in E_1, find the spatially closest edge point f_i' in the other edge map E_2, and obtain the *displacement vector* $\overrightarrow{d_i}$ from two matched edge points:

$$\overrightarrow{d_i} = (x_i' - x_i, y_i' - y_i) \tag{6}$$

where (x_i, y_i) and (x_i', y_i') are coordinates of f_i and f_i' respectively;

3. Construct the feature point correspondence map (FPCM) from displacement vectors of all edge points in image E_1:

$$C_{E_1 \to E_2} = \{\overrightarrow{d_i}(f_i \to f_i') | f_i \in E_1\} \tag{7}$$

Note that the FPCM is *directed*, which means $C_{E_1 \to E_2}$ and $C_{E_2 \to E_1}$ are not necessarily, and very unlikely, to be identically reversed for a given pair of edge maps E_1 and E_2. This implies that the proposed correspondence map based similarity measure should also be directed.

B. A Feature Based Similarity Measure

In the face alignment phase, a FPCM serves two purposes. First, one can evaluate how close two images resemble each other in terms of structural(pose) based on associated FPCMs. In particular, for two given images I_1 and I_2, and the associated FPCMs $C_{E_1 \to E_2}$ and $C_{E_2 \to E_1}$, a *feature based similarity measure*(FBSM) can be defined as:

$$S_{feat}(I_1, I_2) = \max(S^d(C_{E_1 \to E_2}), S^d(C_{E_2 \to E_1})) \tag{8}$$

where $S^d(C_{E_a \to E_b})$ is the root mean squared length of displacement vectors in the directed FPCM $C_{E_a \to E_b}$:

$$S^d(C_{E_a \to E_b}) = \frac{1}{N} \sum_{i \in E_a} \left\| \overrightarrow{d_i} \right\|^2 \tag{9}$$

This similarity measure is in fact a modification of the classical Hausdorff measure, which takes maximum of displacement vectors' lengths [13][14]. Note that this similarity measure is proportional to the average *spatial* distances, rather than intensity differences between two sets of feature points. We argue that this property is crucial to pose alignment, since an optimization algorithm using this similarity (or dissimilarity) measure as an error function will attempt to merge two sets of feature points and eventually align two face poses.

C. Computing Best Hill-climb Direction

During face alignment, FPCMs also serve a second purpose to help estimate the best direction of hill-climb in pose parameter space. One can obtain a dense correspondence map(C) from a sparse FPCM by interpolation, and project (C) into the learnt varying poses subspace:

$$\Delta\gamma = \Phi'_P \cdot C \quad (10)$$

This FPCM based pose re-estimation method has several advantages over other pose re-estimation methods commonly found in the literature. Compared to stochastic gradient descent techniques using intensity based errors ([4]), it is less susceptible to local minima when initial poses are far off, provided one can obtain reasonable FPCMs.

3.2 Combined Feature-Texture Similarity Measure

A. Pose Aligned Intensity Difference

To estimate the face texture parameter α, we minimize the intensity difference between pose aligned faces:

$$S_{text}(I_1, I_2) = SSD(I_1, (I_2 \circ C)) \quad (11)$$

where *SSD()* is the sum of square distance, and C is the correspondence map associating two images I_1 and I_2.

B. A Combined Feature-texture Similarity Measure

Considering both feature(pose) and texture similarities (Equations 8 and 11), we can define a combined similarity measure:

$$S(I_1, I_2) = K_s \cdot S_{feat} + K_t \cdot S_{text} \quad (12)$$

where K_s and K_t are two user-specified weighting parameters. If no priori knowledge is accessible, one can simply set them with equal values(e.g. 1.0) without emphasizing either aspect.

It is worth mentioning that existing methods in [4][5][6][7][8] measure the intensity differences between structural normalized faces, which is in fact a degenerate case of the proposed combined similarity measure ($K_s = 0$,$K_t = 1$). We note that these methods match frontal (or near frontal) faces well, but may experience difficulties with initially faraway poses, because intensity differences no longer consistently reflect structural(pose) difference under this situation. In contrast, the combined similarity measure explicitly accounts for spatial difference between feature points, and can reliably align two poses, even when they are initially far off (see Figure 6).

3.3 Face Alignment

With the proposed parametric varying pose face model, one is able to formulate pose alignment as a parameter estimation problem, with objective to minimize the FTSM between the original image (I) and the model image (I^M). The optimal pose parameter is estimated using the *maximum likelihood* principle:

$$(\alpha^*, \gamma^*) = \arg\min S(I, I^M(\alpha, \gamma)) \quad (13)$$

We outline a face alignment algorithm, which given an input face image, alternates between estimating texture and pose parameters in an iterative manner:

1. Set α_0^* and γ_0^* such that I^M corresponds to a prototype frontal face image (see Figure 3);
2. With fixed α_{i-1}^*, estimate $\gamma_i^*(i = 1, 2, 3, ...)$ with objective to minimize $S(I, I^M(\alpha_{i-1}^*, \gamma_i^*))$ using FPCM based hill-climb method (Section 3.1);
3. With fixed γ_i^*, estimate α_i^* with objective to minimize $S(I, I^M(\alpha_i^*, \gamma_i^*))$ using Levenberg-Marquardt method [15];
4. **if** ($S >$threshold) $\{i = i + 1;$ and go to step 2; $\}$ **else** stop looping;
5. Output (α_i^*, γ_i^*), $S(I, I^*(\alpha_i^*, \gamma_i^*))$ and $I^M(\alpha_i^*, \gamma_i^*)$;

One can compute the optimal FPCM based on estimated pose parameter γ^* using Equation 4, then apply it to prototype feature points to obtain the warped feature points which are aligned with given face poses. Figure 4 shows an example of pose alignment. More experimental results and discussions will be presented in the next section.

4 Face Alignment Results

We used the proposed model and the similarity measure to align prototype face features with unknown pose face images. First, we construct a frontal face subspace model, and a varying pose model, using a subset of the MIT Beymer face database as training data. The training set includes 60 different persons' face images (100x100 pixels), under 5 fixed poses (300 face images in total). All frontal training faces are shape-normalized to align feature points. Training FPCMs are obtained by manual registration of non-frontal face images with reference to a prototype frontal face. The Beymer face database also contains a separate set of testing face images under various poses, which fall in range of $[-60^o, 60^o]$ left-right rotations, and $[-10^o, 10^o]$ down-up rotations.

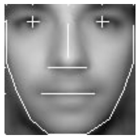

Figure 3: Feature points and face prototype used in this paper. There are 16 line segments and 24 end points of line segments.

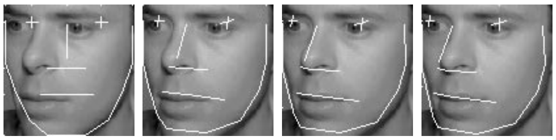

Figure 4: Unknown pose face alignment with feature points marked (Left to Right: Iteration 0,1,5 and 9).

Figure 4 shows intermediate pose alignment outputs during execution of above algorithm, and Table 1 summarizes the corresponding estimated pose parameters and error measures. Note that the estimated pose parameters do not only enable one to align face features under varying pose, but also provide useful information about face images. For instance, comparing the final pose parameter in Table 1 with eigen vectors in Figure 2, one can find out that the first component($\gamma_{[0]} = -1599.95$) and the second component($\gamma_{[1]} = -155.83$) correspond to a left and slightly downward in-depth rotation respectively, which is a fairly accurate semantic description of face pose in Figure 4.

The face alignment algorithm was tested with 50 Beymer database testing faces, which belong to 10 (6 male and 4 female) different people under 5 unknown poses.In addition, we also tested the algorithm with 20 new face images taken under different lighting conditions. Table 2 summarizes the success rate of pose alignment under various conditions, and Figure 7 shows selected alignment results. More than 70% faces were well aligned, 27% fairly well aligned and one face misaligned (Figure 5), under various conditions including different poses, face appearances, gender and lighting conditions. Because of its sensitivity to spatial differences between feature points, the proposed method can reliably align faces even when the initial pose estimate is far off. For example in Figure 6, we deliberately set the initial pose to a far-away left rotation, the algorithm could still converge to the ulimate correctly in few iterations.

Figure 5: Face misaligned.

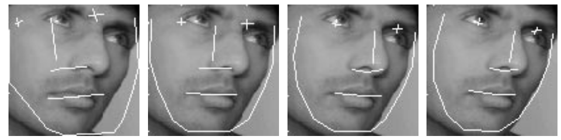

Figure 6: Pose alignment when initial pose is far off (Left to right: Iteration 0, 1, 5 and 9).

Figure 7: Face alignment results (row 1 to row 4: Beymer test faces; row 5: faces taken under different lighting conditions).

iterations	$\gamma_{[0]}$	$\gamma_{[1]}$	$\gamma_{[2]}$	Error
0	0.00	0.00	0.00	39.38
1	-1007.21	-23.65	65.31	34.10
5	-1325.43	-100.66	150.17	29.36
9	-1599.95	-155.83	198.18	27.05

Table 1: Pose parameters (first three components) and error measures ($K_s = 1.0$, $K_t = 1.0$).

#Faces	Good	Fair	Misaligned
70	50 (71.4 %)	19 (27.1 %)	1 (1.4 %)

Table 2: Face alignment results - we define a good pose alignment as having at most 2 feature points misaligned (i.e. 10 pixels away from correct positions), and a fair pose alignment having at most 5 feature points misaligned, otherwise a pose misalignment. Face size is 100 x 100 pixels.

5 Conclusions and Future Work

This paper shows how one can compute and use sparse correspondence maps of distinctive image features, to better deal with structural (pose) mismatches within an existing model-based image alignment framework. The sparse feature point correspondence maps (FPCMs) contribute to the alignment process in two ways:

1. They explicitly encode, and are therefore used to quantify spatial displacements between distinctive feature points in a pair of mis-aligned images. When combined with a standard pose-adjusted intensity based similarity measure, the outcome is an error measure that better reflects pose differences between a pair of mis-aligned images.

2. They are used for re-estimating pose parameters, in a way similar to how dense optical flow fields have been used for the same purpose. The pose re-estimation step does not involve stochastic gradient descent, and is hence efficient. When image flow brightness constraints do not work well for computing dense optical flow fields (e.g. when initial pose and/or texture estimates are far off), our FPCMs can lead to more reliable results.

We note that our proposed extensions work well only when feature-based correspondences can be reasonably established. While this has been largely true for our face images, and Huttenlocher et. al. [13] have even reported reliable feature-based alignment results under more challenging conditions, we believe that more discriminative feature matching methods will be needed to deal with significant image clutter. This is one direction of our current research.

Acknowledgments

We thank Dr. David Beymer and the MIT Artificial Intelligence Laboratory for making available the Beymer face database. Ms. Zeng Ying helped us manually label the training data. Lixin Fan is supported by NUS research scholarship. This research is supported by a NUS ARF Research Grant: RP3960623.

References

[1] M. A. Turk, and A. P. Pentland, "Face Recognition Using Eigenfaces", *Journal of Cognitive Neuroscience*, Vol. 3, No. 1, 1991.

[2] B. Moghaddam and A. Pentland, "Probabilistic Visual Learning for Object Detection", *M.I.T., Media Lab. Perceptual Computing Section Technical Report*, No. 326, June, 1995.

[3] B. A. Pentland, B. Moghaddam, and Thad Starner "View-Based and Modular Eigenspaces", *M.I.T., Media Lab. Perceptual Computing Section Technical Report*, No. 245.

[4] Michael J. Jones, and Tomaso Poggio, "Model-Based Matching by Linear Combinations of Prototypes", *M.I.T. A.I. Memo*, No. 1583, Nov. 1996.

[5] Thomas Vetter, Michael J. Jones, and Tomaso Poggio, "A Bootstrapping Algorithm for Learning Linear Models of Object Classes", *M.I.T. A.I. Memo*, No. 1600, Feb. 1997.

[6] David Beymer, "Vectorizing Face Images by Interleaving Shape and Texture Computations", *M.I.T., A.I. Memo*, No. 1537, Sept. 1995.

[7] I. Craw, N. Costen, T. Kato and S. Akamatsu, "How should we represent faces for automatic recognition?", *IEEE Trans. on PAMI,* vol. 21, pp. 725-736, Aug. 1999.

[8] T.F. Cootes and C.J. Taylor, "Active Appearance Models", *Proc. ECCV*, Vol. 2, pp. 484-498, 1998.

[9] P. Viola and W. Wells, "Alignment by Maximization of Mutual Information", *Proc. ICCV*, Cambridge, MA, 1995.

[10] A. L. Ratan, W.E.L.Grimson, and W.M. Wells III, "Object detection and localization by dynamic template warping", *Proc. CVPR*, pp. 634-640, 1998.

[11] Thaddeus Beier and Shawn Neely. "Feature-based image metamorphosis". In *SIGGRAPH'92 Proceedings*, pages 35-42, Chicago, IL, 1992.

[12] B.K.P. Horn and G. Schunck. "Determining Optical Flow", *Artificial Intelligence*, 17:185-203, 1981.

[13] D. P. Huttenlocher, G.A.Klanderman, and W.J.Rucklidge, "Comparing Images Using the Hausdorff Distance", *IEEE Trans. on PAMI*, vol. 15, pp. 850-863, Sept. 1993.

[14] M. P. Dubuisson and A. K. Jain, "A modified Hausdorff Distance for Object Matching", in *Proc. 12th Int. Conf. Pattern Recognition,* pp. 566-568, Oct. 1994.

[15] W. Press, *et al. "Numerical Recipes in C"*, Second Edition, Cambridge University Press, 1992.