FACE ALIGNMENT USING INTRINSIC INFORMATION

Yuchi Huang[†,‡], *Stephen Lin*[‡], *Hanqing Lu*[†], *Heung-Yeung Shum*[‡]

†Institute of Automation, Chinese Academy of Sciences
‡Microsoft Research Asia

ABSTRACT

Previous 2-D face alignment algorithms are generally quite sensitive to illumination variation and poor initialization. To account for these two obstacles, two forms of relatively lighting invariant descriptors—*Intrinsic Gray-level Information* and *Intrinsic Edge Information*—are adopted in our algorithm to direct shape search. The former is recovered from local intensity normalization and useful at localizing face contours accurately despite its dependency on initialization. The latter is extracted from normalized local regions by Canny edge filtering and is robust at coarse alignment in spite of poor initialization. The different merits of these two forms of intrinsic information motivate us to employ them at different stages of our face alignment process. Extensive experimentations show that this proposed approach allows our system to handle not only illumination variation, but also poor initialization.

1. INTRODUCTION

Many approaches have been proposed to align faces from cluttered images. Active Shape Models (ASM) [1] and Active Appearance Models (AAM) [2], proposed by Cootes et al. are two popular algorithms for face alignment. However, there are two problems with previous face alignment algorithms. First, the search strategy in these methods tends to collapse when there exists significant shading and shadowing which can effectively mask subtle features and introduce misleading features as well. These illumination effects can substantially reduce the robustness and accuracy of face alignment. Second, even for images under relatively uniform lighting, convergence to correct shape points cannot be guaranteed in instances of poor initialization. Typically we have to provide a good initialization manually to prevent the face localization process from converging to local minima.

To reduce the misleading effects of shading and shadows, we propose a patch normalization technique to recover *Intrinsic Gray-level Information* from local regions. This kind of gray-level information gives accurate and robust alignment results for instances with good initializations, but complete dependence on it throughout the entire search process can lead the alignment process astray because of its sensitivity to initialization. To give a good initialization automatically, we propose to use information on prominent edges, called *Intrinsic Edge Information* in place of Intrinsic Gray-level Information in the initial search phase. Extracted from normalized local regions by Canny edge filtering, prominent edge information can effectively lead the search to locate some subset of the features robustly, and decrease the likelihood of poor convergence by providing a rough alignment. This kind of hierarchical search strategy uses different information at different stages, and allows our algorithm to handle not only illumination variation, but also poor initialization. With this proposed approach, our face alignment system yields accurate, stable and efficient performance under a wide range of illumination conditions, as evidenced in extensive experimentation.

2. PATCH FILTERING AND EDGE EXTRACTION

In this section, we describe the two relatively illumination invariant feature descriptors used in our alignment technique. The first one for finer alignment is based on local intensity normalization, and the second one for coarse alignment involves a Canny filter to extract prominent edge information. Since shading and shadows often diminish the appearance of features, they decrease the likelihood of correct convergence. To reduce the diminishing effects of shading and shadows, patch filtering is proposed for local intensity normalization, which makes a feature more distinguishable from its surrounding area, as demonstrated in Fig 1(b). Our formulation of the patch filter begins with the Lambertian lighting model, which describes a gray-level image $I(x, y)$ as

$$I(x, y) = \rho(x, y)\mathbf{n}^\top(x, y)\mathbf{s} \tag{1}$$

or more generally, the Lambertian model with shadows can be represented as

$$I = \min(\rho\mathbf{n}^\top \sum_l \mathbf{s}_l, 0) = \min(\rho\mathbf{n}^\top \mathbf{S}, 0) \tag{2}$$

where $\rho(x, y)$ is the reflectance (albedo) associated with point (x,y) in the image, $\mathbf{n}(x, y)$ denotes the surface normal

0-7803-8554-3/04/$20.00 ©2004 IEEE.

(a) (b) (c)

Fig. 1. (a) the input image, (b)its patch-filtered image, and (c) its intrinsic edge image.

of the face at (x,y), and **S** is the light source direction and intensity, which can be represented as a linear combination of multiple point light sources. This equation can be seen as a product of a reflectance component (ρ) and an illumination component ($\mathbf{n}^T\mathbf{S}$) as observed by Barrow and Tennenbaum [3].

From Retinex theory [4], the illumination image component can be approximated as the low frequency component of I, determined by convolution of the image with a low-pass Gaussian filter, which we denote as F_1. Dividing image intensities by this illumination component then yields an illumination-invariant descriptor:

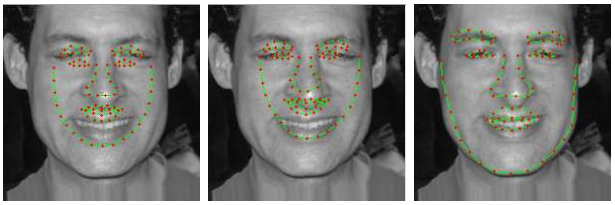
$$R = \frac{I}{I * F_1} \tag{3}$$

This descriptor normalizes a local patch with respect to illumination intensity, under the assumption that the illumination intensity is fairly even over the local patch. This division by the illumination component, however, can emphasize noise in the patch. To reduce this side effect, we filter out the high-frequency components in the numerator by convolving it with a low-pass filter F_2 with a larger passband than F_1:

$$R_1 = \frac{I * F_2}{I * F_1} \tag{4}$$

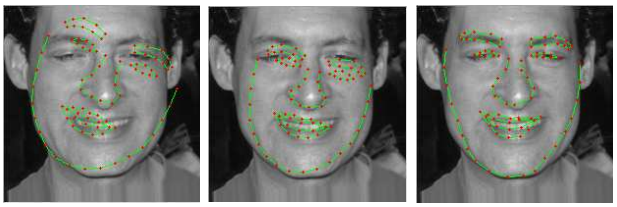
where the division is pixel-wise. The overall result of patch filtering is shown in Fig. 1(b). Although image noise is still somewhat amplified, the features become much more apparent than before patch filtering.

Because gradient-based edge detection methods are sensitive to edge magnitude and smoothness, which can be significantly affected by illumination conditions, it is not suitable to use them to extract Intrinsic Edge Information directly from the original images. In our algorithm, we employ patch filtering locally at first, and then use the modified Canny Filter suggested by Fleck [5] to extract prominent edges (shown in Fig. 1(c)). Because this method identifies most prominent feature points in an image, it is relatively robust to convergence problems caused by noise and local variations.



(a) (b) (c)

Fig. 2. (a) Initialization #1: scale variation. (b) Results using patch-filtered gray-level feature. (c)Results using intrinsic edge feature.



(a) (b) (c)

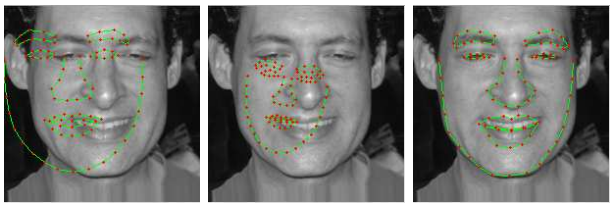
Fig. 3. (a) Initialization #2: rotation variation. (b) Results using patch-filtered gray-level feature. (c)Results using intrinsic edge feature.

3. HIERARCHICAL SEARCH MODEL

To distinguish the relative merits of the two relatively illumination invariant features, their performance should be measured with respect to initialization sensitivity and alignment accuracy. We use as feature models the principal components of the filtered edge values and also the principal components of the filtered gray-level values in local windows centered on each feature point. The principal components analysis is computed from 200 images of size 200x200 under various non-extreme illumination conditions. Three different poor initializations illustrated in Figs. 2-4 are used to test the sensitivity of these two features to initialization. Fig. 5 gives the statistical results on a different set of 200 images. It is clear that the edge-based method is more robust to poor initialization. Although much image information is disregarded in edge images, it is very useful at coarsely localizing face features in spite of poor initialization. From the experimental results presented in Sec. 4, it is also apparent that the patch filtering method provides higher accuracy for final alignment.

The different merits of these two methods motivates us to employ Intrinsic Edge Information and Intrinsic Gray level Information at different stages of the alignment process. The hierarchical implementation of our alignment method is summarized in the following steps. As in many ASM implementations, the number of resolutions we use is 4.

1. For each training and test image, a Gaussian image



(a) (b) (c)

Fig. 4. (a) Initialization #3: displacement variation. (b) Results using patch-filtered gray-level feature. (c) Results using intrinsic edge feature.

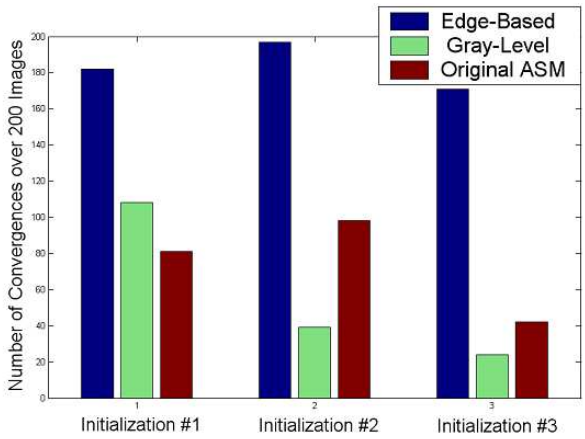


Fig. 5. Comparison of different search features. The y-axis represents the number of images out of 200 on which shape points converge to relatively correct positions, as opposed to images on which shape points converge to totally wrong positions.

pyramid is built. The base image is denoted as level 0, and the roughest image is taken as level L. Similar to the original ASM method, a statistical shape model is built from the training images using PCA. For the highest level of the pyramid, the PCA models of the edge features are built from the training images. For the other levels,the PCA models of patch features are computed.

2. An initialization for each test image is determined. For images under non-extreme illumination, the initial shape can be given by a face detection algorithm. For images under extreme illumination, the initial shape is provided manually or could be provided by color-based detection methods.

3. In the search phase,

   (a) Set $l = L$.

   (b) If $l = L$, use the edge feature. Otherwise, use the patch feature.

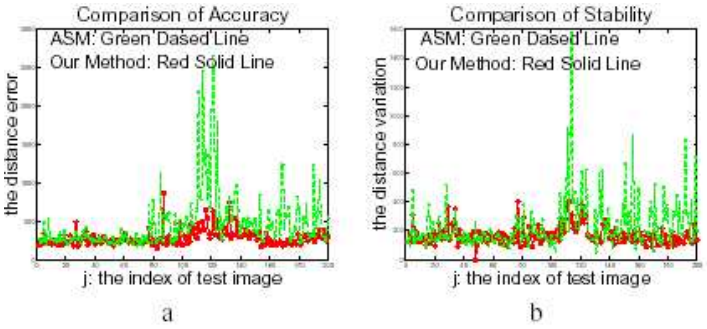


Fig. 6. Statistical comparison with ASM. (a) Accuracy; (b) Stability.

(c) Search the positions of all points until 90% of the points converge, and then project the shape into the PCA shape subspace.

(d) If $l > 0$, then decrement l by one and return to (b).

4. EXPERIMENTS

To test the performance of our alignment system, we do substantial experimentation on two groups of images, under general illuminations without significant facial shadows and under extreme illuminations, which consist of a single point light source at a large angle from the viewing direction. Because our algorithm requires some additional time to compute the edge and patch features in the image, it is slightly slower than the original ASM search scheme, but it nevertheless takes only about 0.4 to 0.7 seconds to align a face in a 200x200 image on a P-4 1.4G computer with 256M memory. We manually labelled 400 pictures under general illumination, each of size 200X200. Of these images, 200 were used for training and the remaining 200 for testing. Even though the faces are fairly well illuminated, some of these images present problems to ASM. In this section, we first compare the accuracy and stability of our algorithm to the original ASM method on the 200 test images. To measure accuracy, the distance between the searched feature positions (x_{k1}, y_{k1}) and the manually annotated feature points (x_{k2}, y_{k2}) is taken as the estimated alignment error:

$$D = \sum_k \sqrt{(x_{k1} - x_{k2})^2 + (y_{k1} - y_{k2})^2} \qquad (5)$$

As a rough measure of stability, we input the manually aligned shape to the alignment algorithms as the initialization, and then observe the variation between this initialization and the resulting shapes after search. Fig. 6(b) exemplifies the greater stability of our method in comparison to the original ASM method.

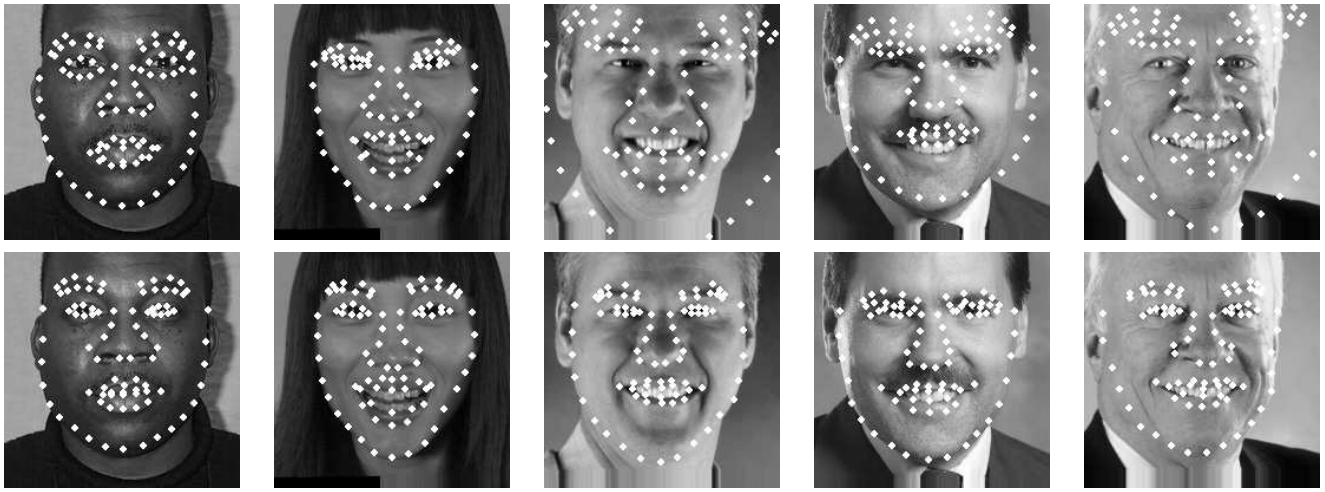


Fig. 7. Comparison under good illumination. Top row: ASM. Bottom row: our method.

Fig. 8. Alignment results for different people under extreme illumination.

Additional comparisons under general illumination are given in Fig. 7 for examples with exaggerated expressions, facial hair, or shading variation on the faces. We selected images from the CMU PIE database [6] and YALE FACE DATABASE [7] to test our method on extreme illuminations. Since ASM collapses entirely on extreme images, a statistical comparison between ASM and our algorithm is not meaningful. Experiments on these images show that our system can give reasonable results under various shadings and shadows, as exemplified in Fig. 8. For some images with significant shadowing, although our algorithm may not accurately locate some of the feature points, it rarely collapses to a totally wrong result.

5. REFERENCES

[1] T.F. Cootes, C. J. Taylor, D. H. Cooper and J. Graham. "Active shape models - their training and application", *CVIU*, 61(1):389, January 1995.

[2] T.F. Cootes, et al. "Active appearance model", *Proc. 5th ECCV*, Freiburg, Germany, 1998.

[3] H.G. Barrow and J. Tenenbaum. "Recovering intrinsic scene characteristics from images", *Computer Vision Systems*, pp. 3-26. Academic Press, 1978.

[4] D. J. Jobson, et al. "Properties and Performance of a Center/Surround Retinex", *IEEE Trans. on Image Proc.*, Vol. 6, No. 3, 1997, pp. 451-462

[5] Margaret M. Fleck. "Some Defects in Finite-Difference Edge Finders", *IEEE Trans. on PAMI*, No. 3, Vol. 14. March 1992. pp 337-345

[6] T. Sim, S. Baker et al. "The CMU Pose, Illumination, and Expression Data-base", *Proc. IEEE Int. Conf. on Auto. Face and Gesture Recog.*, May, 2002

[7] Athinodoros S. Georghiades and Peter N. Belhumeur. "From Few to many: Illumination cone models for face recognition under variable lighting and pose", *IEEE Trans. PAMI*, Vol. 23, No. 6, pp 643-660, 2001