Active Wavelet Networks for Face Alignment

Changbo Hu, Rogerio Feris, Matthew Turk
Dept. Computer Science, University of California, Santa Barbara

Abstract

The active appearance model (AAM) algorithm has proved to be a successful method for face alignment and synthesis. By elegantly combining both shape and texture models, AAM allows fast and robust deformable image matching. However, the method is sensitive to partial occlusions and illumination changes. In such cases, the PCA-based texture model causes the reconstruction error to be globally spread over the image. In this paper, we propose a new method for face alignment called active wavelet networks (AWN), which replaces the AAM texture model by a wavelet network representation. Since we consider spatially localized wavelets for modeling texture, our method shows more robustness against partial occlusions and some illumination changes.

1 Introduction

Many computer vision tasks such as face recognition and facial expression analysis require the accurate alignment between a given face and a canonical face. Extensive research has been conducted on this topic, especially using model-based approaches [5, 3].

Among model-based methods, the active appearance model (AAM) algorithm [5] has achieved good results in face alignment. The method makes use of statistical models of both shape and texture, allowing fast and robust deformable image matching. Several variations of AAM have also been proposed to improve the original algorithm, namely view-based AAM [6], Direct Appearance Models [9], a compositional approach [2] and 3D AAM [1].

Despite the success of AAM and its variations, problems still remain to be solved. AAM is sensitive to illumination changes, especially if the lighting in the test image significantly differs from the lighting encoded in the training set. Moreover, under the presence of partial occlusion, the PCA-based texture model of AAM causes the reconstruction error to be globally spread over the image, thus impairing alignment.

This paper proposes a new method, called active wavelet networks (AWN), in which a Gabor wavelet network representation (GWN) [11] is used to model the texture variation in the training set. The GWN approach represents a face image through a linear

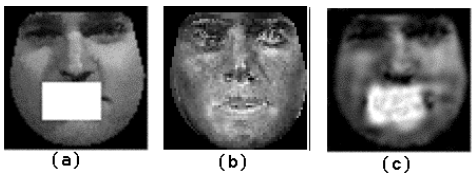

Figure 1: (a) Partial occluded image. (b) PCA reconstruction. Note that the error is spread over the image. (c) Wavelet reconstruction.

combination of 2D Gabor functions whose parameters (position, scale and orientation) and weights are optimally determined to preserve the maximum image information for a chosen number of wavelets.

Because of the localization property of wavelets, when partial occlusion or highlight illumination problems arise, the matching is more robust than with AAM. Figure 1 illustrates a comparison between PCA and a Gabor wavelet reconstruction for a partial occluded face image. Note that the error is globally spread over the image in PCA, whereas it remains local in the wavelet representation.

Our method also offers some advantages regarding efficiency, since some computations are limited to the spatial support of the filters, rather than the whole image.

The remainder of this paper is organized as follows. Related work is discussed in Section 2. In Section 3, we present the AWN approach for face alignment. Experiments are presented in Section 4 and conclusions are given in Section 5.

2 Related Work

Recently, Gabor wavelet networks [11] has been proposed as an effective approach for object representation, with successful applications in face tracking [7] and pose estimation [10]. The idea is to represent an object as a linear combination of 2D Gabor wavelets, with parameters and weights optimally determined from the continuous space. Face alignment may be performed by affinely transforming the wavelet representation to match a new image. However, since a GWN is optimized for a particular image, alignment becomes a problem when different individuals are considered.

This problem was tackled by anonymous author [8] using a set of exemplars, but the method is computationally expensive. The solution we adopt in this paper is to optimize a single GWN over a set of face images. This requires all the face images to have the same shape. More details about this technique will be described in section 3.2.

Another wavelet-based approach is the bunch-graph method [13]. In this approach, Gabor jets are used as feature vectors and an elastic graph matching algorithm is adopted for face alignment. Our method is faster than the bunch graph method because we use a

sparser representation based on GWN and a more efficient search technique based on the active appearance algorithm.

The work most closely related to this paper is the active appearance model method [5]. AAM uses PCA to encode both shape and texture variation, as well as the correlations between them. By assuming a linear relationship between appearance variation and texture variation and between texture variation and pose variation, AAM learns the linear regression models from training data. The model search is driven by the residual of the search image and model reconstruction.

The Shape-AAM method [4] is a variation of the standard AAM algorithm, generally applied when shape modes are fewer than texture modes. Instead of manipulating the combined appearance parameters, ShapeAAM uses image residuals to drive shape and pose parameters, and then compute texture parameters directly from the image, given the current shape. In this paper, our texture model is the GWN representation from the shape-free image set and the search method is similar to Shape-AAM. Our method enables the search to be robust to partial occlusion and some illumination changes, while providing more efficiency.

It is worth mentioning that AAM is better suited for image synthesis than our method, since a small number of eigenfaces can generate photo-realistic images, whereas a high number of wavelets would be required for this purpose.

3 Active Wavelet Networks

In this section, we introduce active wavelet networks for face alignment. Our method starts with a training set, in which each image is labelled with landmark points on the subject's face. Thus, each sample has a labelled shape and an image texture.

Consider the training set of shape and texture to be $\Omega = \{(\mathbf{x}_i, \mathbf{g}_i^x)\}, i = 1...N$,where N is the number of training images, $\mathbf{x}_i = \{(x_j^i, y_j^i)\}, j = 1...M$, is a shape specified by a set of M points, and $\mathbf{g}_i^x$ is the texture enclosed by the shape $\mathbf{x}_i$. We model the shape variation by PCA, and the texture is represented by a GWN model. We will describe the shape model and the GWN texture representation in the following subsections.

3.1 Statistical Shape Model

Given the training set, all shapes are aligned to a common coordinate frame and then the shape variation can be modelled by PCA in a lower dimensional shape space. So, a normalized shape $\mathbf{x}$ can be approximated as:

$$\mathbf{x} = \bar{\mathbf{x}} + \mathbf{Pb} \tag{1}$$

Figure 2: (a) Labelled training image. (b) Shape-free texture.

where $\bar{\mathbf{x}}$ is the mean shape, $\mathbf{P}$ is a set of orthogonal modes of variation and $\mathbf{b}$ is a set of shape parameters. Using the shape landmarks as control points, we can warp the training images to the mean shape. Figure 2 illustrates a labelled image and its texture warped into the mean shape. The set of shape-free textures $G = \{\mathbf{g}_i^{\bar{x}}\}, i = 1...N$ is used to learn the GWN representation, as described next.

3.2 Wavelet Network Model

We have used a wavelet network to model the face texture as an alternative to Principal Component Analysis in standard AAM. As already mentioned, the use of spatially localized wavelets allows more robustness with respect to partial occlusions and illumination changes.

The constituents of a wavelet network are single wavelets and their associated coefficients. We adopted the odd-Gabor function as the mother wavelet. It is well known that Gabor filters are recognized as good feature detectors and provide the best trade-off between spatial and frequency resolution [12]. Considering the 2D image case, each single odd Gabor wavelet can be expressed as follows:

$$\psi_{\mathbf{n}}(\mathbf{x}) \quad = \quad \exp\left[-\frac{1}{2}(\mathbf{S}(\mathbf{x}-\mu))^T(\mathbf{S}(\mathbf{x}-\mu))\right] \times \sin\left[(\mathbf{S}(\mathbf{x}-\mu))^T \begin{pmatrix} 1 \\ 0 \end{pmatrix}\right] \qquad (2)$$

where $\mathbf{x}$ represents image coordinates and $\mathbf{n} = (s^x, s^y, \theta, \mu^x, \mu^y)$ are parameters which compose the terms $\mathbf{S} = \begin{pmatrix} s^x\cos\theta & -s^y\sin\theta \\ s^x\sin\theta & s^y\cos\theta \end{pmatrix}$, and $\mu = \begin{pmatrix} \mu^x \\ \mu^y \end{pmatrix}$, that allow scaling, orientation, and translation. A Gabor wavelet network for a given image consists in a set of n wavelets $\{\psi_{\mathbf{n}_k}\}$ and a set of associated weights $\{w_k\}$, specifically chosen so that the GWN reconstruction:

$$\hat{I}(\mathbf{x}) = \sum_{k=1}^{n} w_k \psi_{\mathbf{n}_k}(\mathbf{x}) \qquad (3)$$

best approximates the target image. We modified the original formulation of GWNs to allow the optimization of a single GWN in a set of shape-free images, obtained through warping, as described in previous section.

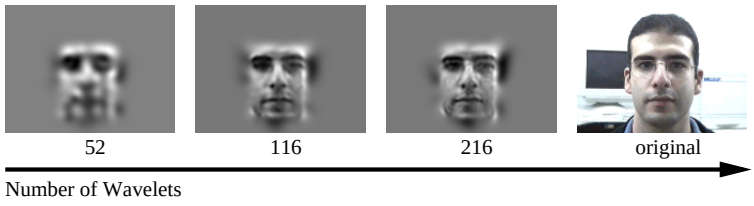

Figure 3: The image shows a facial reconstructions with variable accuracy, considering (from left to right) 52, 116 and 216 wavelets.

3.2.1 Calculation of Wavelet Parameters

Assuming that we have a set of shape-free face images of different people, $\{\mathbf{g}_i^{\bar{x}}\}, 1 \leq i \leq N$, that are truncated to the region that the face occupies, we can calculate the GWN representation parameters as follows:

1. Randomly drop n wavelets of assorted position, scale, and orientation, within the bounds of the normalized face images.
2. Perform gradient descent (e.g., via Levenberg-Marquardt optimization) over the set of wavelet parameters to minimize the total sum of differences between the training images and their wavelet reconstructions:

$$\arg\min_{\mathbf{n}_k, w_{ik}} \left\| \sum_{i=1}^{N} \mathbf{g}_i^{\bar{x}} - \left(\sum_{k=1}^{n} w_{ik} \psi_{\mathbf{n}_k}(\mathbf{x}) \right) \right\|^2 . \tag{4}$$

One advantage of the GWN approach is that one can trade-off computational effort with representational accuracy, by increasing or decreasing the number n of wavelets (see Figure 3).

3.2.2 Calculation of Texture Parameters

In the standard Shape-AAM method, the texture parameters for a given image are computed by projecting the image into an eigenspace learned from the training set. In our method, the texture parameters $\{t_k\}, k = 1...n$ correspond to wavelet coefficients, obtained by orthogonally projecting the image into the learned wavelet subspace.

However, Gabor wavelet functions are not orthogonal, thus implying that, for a given family Ψ of Gabor wavelets, it is not possible to calculate t_k by a simple inner product of the Gabor wavelet $\psi_{\mathbf{n}_k}$ with the image.

In fact, a family of dual wavelets $\tilde{\Psi} = \{\tilde{\psi}_{\mathbf{n}_1} \ldots \tilde{\psi}_{\mathbf{n}_n}\}$ has to be considered. The wavelet $\tilde{\psi}_{\mathbf{n}_j}$ is the dual wavelet of the wavelet $\psi_{\mathbf{n}_i}$ iff $\langle \psi_{\mathbf{n}_i}, \tilde{\psi}_{\mathbf{n}_j} \rangle = \delta_{i,j}$.

Given a normalized face image $\mathbf{g}$ and a set of optimized wavelets $\Psi = \{\psi_{\mathbf{n}_1}, \ldots, \psi_{\mathbf{n}_N}\}$, the texture parameters are given by:

$$t_k = \langle \mathbf{g}, \tilde{\psi}_{\mathbf{n}_k} \rangle. \tag{5}$$

It can be shown that $\tilde{\psi}_{\mathbf{n}_k} = \sum_l \left(A^{-1}\right)_{k,l} \psi_{\mathbf{n}_l}$, where A is the wavelet interference matrix, with $A_{k,l} = \langle \psi_{\mathbf{n}_k}, \psi_{\mathbf{n}_l} \rangle$.

3.3 AWN Search

Given a new face image, and a rough estimation of face pose, the search process aims to determine shape and pose parameters that best fit the model into the new image. The AWN search algorithm is a variation of the Shape-AAM method, where the main difference is the calculation of texture parameters and image reconstruction, which are based on the GWN model.

Let $\mathbf{g}^{\bar{x}}$ be the normalized image enclosed by a shape $\mathbf{x}$ and $\hat{\mathbf{g}}^{\bar{x}}$ its GWN reconstruction. The residual between both images is:

$$\delta \mathbf{g} = \mathbf{g}^{\bar{x}} - \hat{\mathbf{g}}^{\bar{x}} \tag{6}$$

The residual δg is used to drive the shape parameters $\mathbf{b}$ and the affine pose parameters $\mathbf{p}$, assuming a linear relationship:

$$\delta \mathbf{b} = \mathbf{B}\delta \mathbf{g}, \quad \delta \mathbf{p} = \mathbf{P}\delta \mathbf{g} \tag{7}$$

where the two regression matrices $\mathbf{B}$ and $\mathbf{P}$ are computed offline, by perturbing the face model parameters on training data. Our search algorithm can be described as follows.

Given a new face image, a starting shape $\mathbf{x}$ and pose $\mathbf{p}$,

1. Sample the image enclosed by the current shape and normalize it to obtain $\mathbf{g}^{\bar{x}}$
2. Use GWN to compute texture parameters using eq. 5 and reconstruct the texture $\hat{\mathbf{g}}^{\bar{x}} = \sum_{k=1}^{n} t_k \psi_{\mathbf{n}_k}$.
3. Compute the residual using eq. 6
4. Predict the the shape and the pose parameters using eq. 7.
5. If the change of $\delta \mathbf{g}$ is small enough or the maximum number of iterations was reached, stop; else go to step 1.

A successful search results in a AWN model that is well aligned with the input face image.

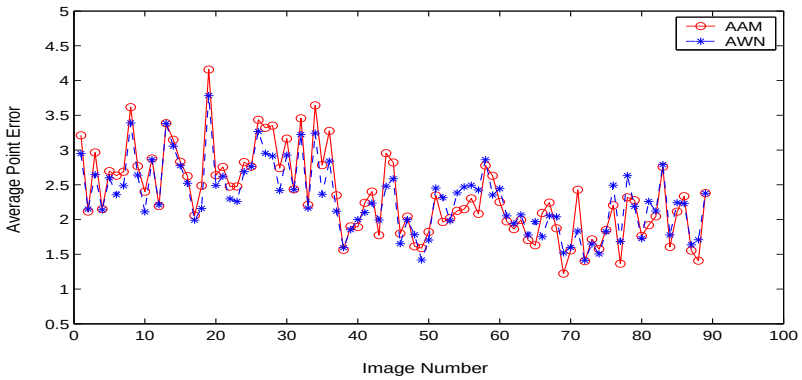

Figure 4: Comparison of precision of AAM and AWN in images captured under normal conditions. The two methods achieve similar performance.

4 Experimental Results

In this section, we present our experimental results, comparing our method with AAM. We have learned the AWN and AAM models in a training set containing 40 face images of different individuals. We derived a statistical shape model using a 15-dimensional eigenspace, capturing 98% of variation in the training set. The texture model for AAM was built using 75 modes, also capturing 98% of the total texture variation.

Considering a small test database of 90 images of different individuals under normal conditions, the performance of AWN and AAM is similar, as shown in Figure 4. The graph shows the average error per shape point for both methods for each image in the test set. We tested several different wavelet subspaces for the AWN texture model, varying the number and initial parameters of wavelets. Nine wavelets were used for comparison in Figure 4. We also verified that the range of initial location for good convergence is about 8 pixels in both methods. This range could in fact be enlarged by a multi-scale approach. In this experiment, the initial model location was randomly chosen with maximum range of 8 pixels from the ground-truth.

We evaluated the performance of AAM and AWN against occlusions in simulated images, using white patches of different sizes at random locations. Figure 5 shows examples of such simulated images. The correct convergence rate in 50 partial occluded images, for both AAM and AWN methods, is shown as a function of the occluding patch size in Figure 6. The occluded parts are easily detected in our wavelet representation and the algorithm compensates to perform robust alignment. In this experiment, we used 60 small-support wavelets in the wavelet network model. We considered the convergence to be correct when the average error per shape point was less than 3.5 pixels. Figure 7 shows an example of alignment on a face with sunglasses, where AAM fails due to the occluded

Figure 5: Examples of random occluding white patches of sizes 5x5,10x10 up to 50x50.

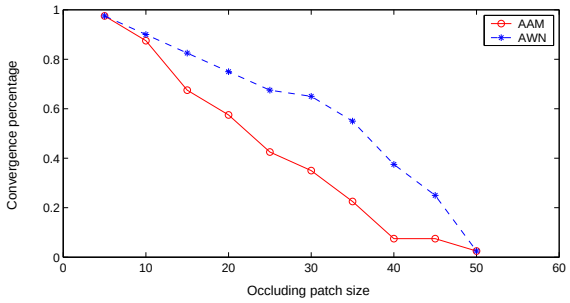

Figure 6: The graph shows the correct convergence rate in 50 partial occluded images, for both AAM and AWN methods, considering each size of the occluding patch.

part, while our method works well.

We also did experiments with images under different illumination conditions. AWN showed to be more robust due to the fact that local highlight changes have a correspondent local effect in the representation. Figure 8 shows an example of matching under varying illumination where our method performs better than AAM. In future work we intend to use an explicit illumination model and also quantitatively evaluate the robustness of AWN against illumination changes.

The AWN approach offers the advantage in efficiency of using just the support of Gabor filters, rather than the whole image, for faster subspace projection and reconstruction. So far, this improvement has been minor because the time for image reconstruction in AAM is not significant compared to the time for other computations, such as warping the image into the mean shape. Our method takes 10ms per iteration in a 1.6 GHz Pentium IV, using images of size 128x192.

5 Conclusions

We have presented a new method for automatic face alignment called active wavelet networks (AWN). Our method takes the advantages of active appearance models for de-

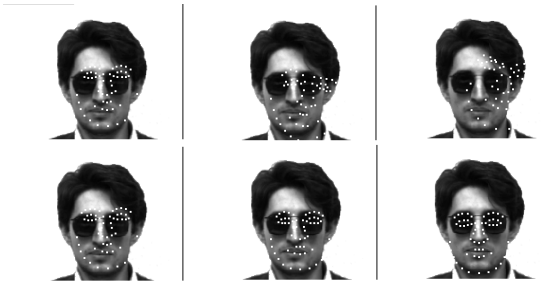

Figure 7: Example of sensitivity of AAM (upper) and AWN (lower) to partial occlusion. Images from left to right are initialization, iteration 2 and final matches.

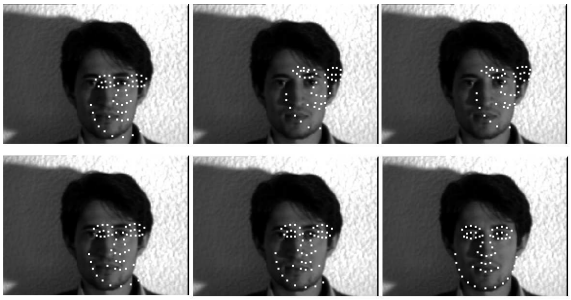

Figure 8: Example of sensitivity of AAM (upper) and AWN (lower) to illumination changes. Images from left to right are initialization, iteration 2 and final matches.

formable image matching, while relying in a texture model based on Gabor wavelet networks. Our experimental results show that AWN is more robust to illumination changes and partial occlusion than AAM.

As future work, we plan to extend our approach to view-based face alignment and recognition. We also intend to use an explicit model for handling illumination changes.

Acknowledgments

We are thankful to Volker Krueger for valuable discussions and part of the source code.

References

[1] J. Ahlberg. Using the active appearance algorithm for face and facial feature tracking. In *ICCV'01 Workshop on Recognition, Analysis and Tracking of Faces and Gestures in Real-Time Systems*, Vancouver, BC, Canada, 2001.

[2] S. Baker and I. Matthews. Equivalency and efficiency of image alignment algorithms. In *Computer Vision and Pattern Recognition*, pages 1090–1097, 2001.

[3] T. Cootes, D. Cooper, C. Taylor, and J. Graham. Active shape models – their training and application. *Computer Vision and Image Understanding*, 61(1):38–59, 1995.

[4] T. Cootes, G. Edwards, and C. Taylor. A comparative evaluation of active appearance model algorithms. In *British Machine Vision Conference*, pages 680–689, Southampton, UK, 1998.

[5] T. Cootes, G. Edwards, and C. Taylor. Active appearance models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(6):681–685, 2001.

[6] T. Cootes, G. Wheeler, K. Walker, and C. Taylor. View-based active appearance models. *Image and Vision Computing*, 20:657–664, 2002.

[7] R. Feris, J. Gemmell, K. Toyama, and V. Krueger. Hierarchical wavelet networks for facial feature localization. In *ICCV'01 Workshop on Recognition, Analysis and Tracking of Faces and Gestures in Real-Time Systems*, Vancouver, BC, Canada, 2001.

[8] R. Feris, V. Krueger, and R. Cesar Jr. Efficient real-time face tracking in wavelet subspace. In *ICCV'01 Workshop on Recognition, Analysis and Tracking of Faces and Gestures in Real-Time Systems*, Vancouver, BC, Canada, 2001.

[9] X. Hou, S. Li, H. Zhang, and Q. Cheng. Direct appearance models. In *Computer Vision and Pattern Recognition*, pages 828–833, 2001.

[10] V. Krueger, S. Bruns, and G. Sommer. Efficient head pose estimation with gabor wavelet networks. In *British Machine Vision Conference*, University of Bristol, 2000.

[11] V. Krueger and G. Sommer. Gabor wavelet networks for object representation. *Journal of the Optical Society of America*, 2002.

[12] B.S. Manjunath and R. Chellappa. A unified approach to boundary perception: edges, textures, and illusory contours. *IEEE Transactions on Neural Networks*, 4(1):96–107, 1993.

[13] L. Wiskott, J. Fellous, N. Krueger, and C. Malsburg. Face recognition by elastic bunch graph matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19:775–779, 1997.