Component-based Face Recognition with 3D Morphable Models

by

Jennifer Huang

Submitted to the Department of Electrical Engineering and Computer Science
in Partial Fulfillment of the Requirements for the Degrees of
Bachelor of Science in Electrical Engineering and Computer Science
and Master of Engineering in Electrical Engineering and Computer Science
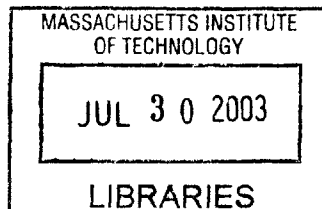at the Massachusetts Institute of Technology
December 17, 2002

Author _____ ___ ____ _____

Department of Electrical Engineering and Computer Science
December 17, 2002

Certified by _____

Tomaso Poggio
Thesis Supervisor

Accepted by _____

Arthur C. Smith
Chairman, Department Committee on Graduate Theses

Component-based Face Recognition with 3D Morphable Models
by
Jennifer Huang

Submitted to the Department of Electrical Engineering and Computer Science

December 17, 2002

In Partial Fulfillment of the Requirements for the Degrees of
Bachelor of Science in Electrical Engineering and Computer Science
and Master of Engineering in Electrical Engineering and Computer Science
at the Massachusetts Institute of Technology

ABSTRACT

This thesis presents a novel approach to pose and illumination invariant face recognition that combines two recent advances in the computer vision field: component-based recognition and 3D morphable models. In a first step a 3D morphable model is used to generate 3D face models from only two input images for each person in the training database. A vast number of synthetic training images under varying pose and illumination conditions are rendered using the 3D models. These synthetic faces are used to to train a component-based face recognition system. The resulting component-based face recognition system achieved 90% accuracy and significantly outperformed a comparable global face recognition system. These results show the potential of the component-based approach toward a fully pose and illumination invariant system.

Thesis Supervisor: Tomaso Poggio
Title: Uncas And Helen Whitaker Professor

Contents

List of Figures

Chapter 1

Introduction

1.1 Background

Within computer vision, face recognition has become increasingly relevant in today's society. The recent interest in face recognition can be attributed to the increase of commercial interest and the development of feasible technologies to support the development of face recognition. Major areas of commercial interest include biometrics, law enforcement and surveillance, smart cards, and access control [25]. Unlike other forms of identification such as fingerprint analysis and iris scans, face recognition is user-friendly and non-intrusive. Possible scenarios of face recognition include: identification at front door for home security, recognition at ATM or in conjunction with a smart card for authentication, video surveillance for security.

As real-world applications for face recognition systems continue to increase, the need for an accurate, easily trainable recognition system becomes more pressing. Early work on face recognition involved methods such as principal component analysis

[10], elastic bunch graph matching [24], and optical flow based techniques [1]. Current systems have advanced to be fairly accurate in recognition under constrained scenarios, but extrinsic imaging parameters such as pose, illumination, and facial expression still cause much difficulty in correct recognition. The problem of face recognition can be divided into two major areas: detection of the face region and identification of the detected region. Attempting recognition on an inaccurate detected face region is hopeless.

Face recognition techniques can be classified as holistic or component based. Holistic or global approaches treat the entire face as one unit and are characterized by the lack of a priori decomposition of the image into semantically meaningful facial parts. The component approach detects the facial parts first and then performs recognition. The justification behind the component-based approach is that the face has a strong configurational appearance that can be exploited and the combination of multiple components will reduce the inaccuracy of an individual component detector. However, the complexity of the system increases with the multiple detection tasks which need to be performed. Component-based methods are mainly characterized by the use of a priori knowledge to represent the face in smaller parts and its constraints on the spatial configuration of parts [16].

Recently, component-based approaches have shown promising results in various object detection and recognition tasks such as face detection [17, 9], person detection [11], and face recognition [4, 23, 12, 8]. The system in [9] was a SVM based recognition system which decomposed the face into a set of components that were interconnected by a flexible geometrical model. The premise behind the system was that changes in

the head pose mainly lead to changes in the position of the facial components which could be accounted for by the flexibility of the geometrical model. In other words, the components themselves changed little compared to the change in position.

The system consisted of a component-based face detection unit and a face recognition unit. In the face detection unit, fourteen component classifiers and a geometrical classifier were used to detect the face and extract the components from an image. The component-based face recognition unit used the extracted components to perform recognition. Their experiments showed that the component-based system consistently outperformed holistic recognition systems in which classification was based on the whole face pattern.

1.2   Motivation and Goals

A major drawback of the system was the need of a large number of training images taken from different viewpoints and under different lighting conditions. In many real world applications, this data is not available. This drawback is eliminated with the incorporation of 3D morphable models in this thesis. The addition of a morphable model allows the computation of a 3D face model using an analysis by synthesis method [3] with only two face images. The ability to generate an arbitrary number of synthetic images under varying pose and illumination conditions through the use 3D morphable models eliminates the need for a large number of real training images.

Previous research has demonstrated both the advantages of component-based systems over global systems and the uses of 3D morphable models. The combination

of the two ideas promises even greater potential for a face recognition system. The implication of this combination of component-based recognition with 3D head models is that any person can be added to the database with only two real images. In addition, the system can theoretically be extended to be completely pose and illumination invariant by generating synthetic images with a full range of pose and illumination conditions.

Given the limited time and processing power constraints, the goal of this thesis was to train a face recognition system which dealt with a subset of poses and illumination conditions. The system has six people in the database and handles rotations from frontal to approximately half profile.

1.3   Outline of Thesis

The outline of the thesis is as follows: Besides component-based recognition which was already discussed in Chapter 1.1, the main ideas behind this thesis are Support Vector Machine (SVM) classification and 3D morphable models. Chapter 2 gives a technical background of the thesis, explaining both SVM classification and 3D morphable models. Since the face recognition system can be broken up into face detection and face recognition, they will also be discussed in separate chapters. Chapter 3 describes both the component-based and the global approach to face recognition. Similarly, Chapter 4 outlines component-based and global face recognition and presents experimental results. Finally, conclusions and suggestions for future work are presented in Chapter 5.

Chapter 2

Technical Background

2.1 Support Vector Machine Classification

2.1.1 Machine Classification

Classifiers, including Support Vector Machines, are generally trained through supervised learning. In supervised learning, a machine chooses the best function relating the inputs and outputs. This function is judged by its ability to generalize on new inputs which were not in the training data. In this section, regularization theory is used to derive SVMs [20], following the approach in [19].

2.1.2 Regularization Framework

Regularization gives solutions in the form of:

$$f(\mathbf{x}) = \sum_{i=1}^{\ell} \alpha_i K(\mathbf{x}, \mathbf{x}_i) \qquad (2.1)$$

where the $\mathbf{x}_i, i = 1, \ldots, \ell$ are the input, K is the kernel, and α_i a set of parameters dependent on the input. The solution f is found by minimizing functionals of the type:

$$\Phi[f] = \frac{1}{\ell} \sum_{i=1}^{\ell} V(y_i, f(\mathbf{x}_i)) + \lambda \|f\|_K^2, \qquad (2.2)$$

where V is a *loss function*, $\|f\|_K^2$ is a smoothness, or *regularizing*, term, which is the norm in the Reproducing Kernel Hilbert Space (RKHS) defined by the kernel K, and λ a positive parameter controlling the relative weight between the data and the regularizing term. The loss function is a measure of goodness of the predicted output $f(\mathbf{x}_i)$ with respect to the given output y_i. The choice of the loss function determines different learning techniques, each leading to a different learning algorithm for computation of coefficients α_i in (2.1).

SVM classification uses the following loss function V:

$$V(y, f(\mathbf{x})) = (1 - y f(\mathbf{x}))_+, \qquad (2.3)$$

where $(t)_+ = t$ if $t > 0$ and zero otherwise.

By solving the Quadratic Programming (QP) problem with linear constraints, the coefficients α_i in (1) can be found. The the loss function (2.3) remarkably yields *sparse* solutions. A sparse solution refers to the fact that only a small fraction of the coefficients α_i in the expansion (2.1) are nonzero. These non-zero α_i have corresponding data points which are termed *support vectors*. Support vectors are unique in that they are sufficient to determine the solution to a given training set. SVM's are also

11

interesting in that the separating surface has the maximum distance to the support

vectors, which are the closest points in the training data. This graphical solution can

be seen in Figure 2.1.
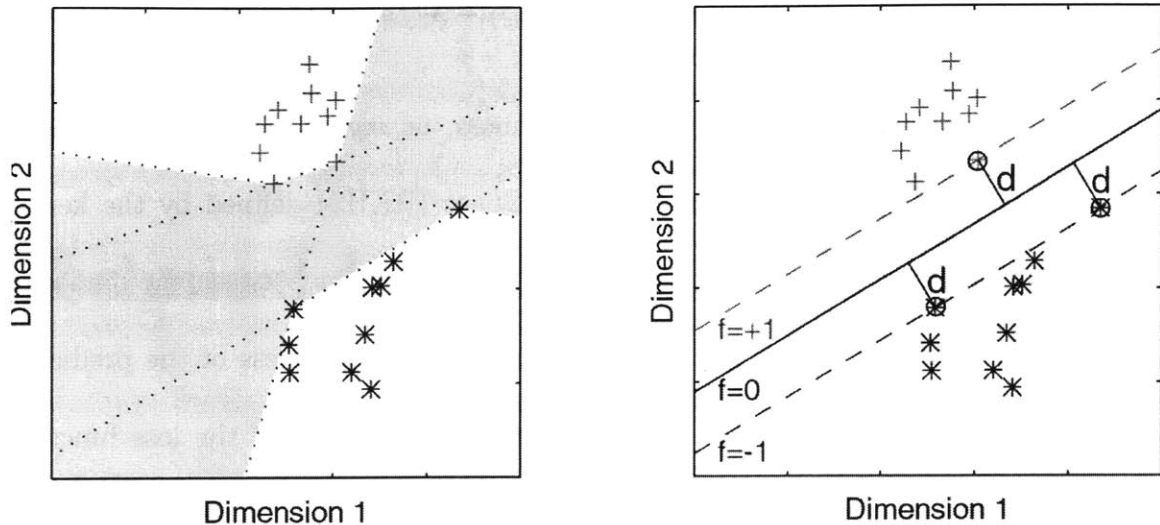
Figure 2.1: A simple linear SVM separation between two classes. On the left, the
figure shows all possible separations of the data without error. On the right, the figure
shows the maximum margin separation. Note the support vectors lie at distances d
from the hyperplane.

The benefit of the SVMs is that the function $f(\mathbf{x})$ is independent of the dimen-

sionality of the feature space. Both the quadratic minimization and the recognition

depend only on the dot products of the unknown vector with the sample vectors.

More detailed explanations of Support Vector Machines can be found in [21] and [5].

2.1.3 Multi-class classification

A SVM can only separate two classes. A number of strategies to solve L-class problems

with binary SVM classifiers exist (see e.g. [15]). Two common solutions to the multi-

class classification problem are the one-vs-all and the pairwise approach. Figure 2.2

diagrams the difference between the two methods.

i) A one-vs-all approach implies that a SVM classifer is trained for each class L in the database, which separates that class from the rest[6, 18].

ii) A pairwise approach implies that $L(L-1)/2$ machines are trained, each separating a pair of classes. These pairwise classifiers are arranged in trees, where each tree node represents an SVM. The classification can be determined by traversing the outputs of the tree. Some implementations of the pairwise approach include [14], [7], and [13].

In this thesis, the one-vs-all approach is chosen for its simplicity and training efficiency. Only L SVMs have to be trained compared to $L(L-1)/2$ SVMs in the pairwise approach. The two strategies are fairly comparable in run-time complexity and performance.

A) Pairwise bottom-up

A or B or C or D

A or B | C or D

A   B   C   D

Training: $L(L-1)/2$
Run-time : $L-1$

B) 1-vs.-All

A / B,C,D

B / A,C,D

C / A,B,D
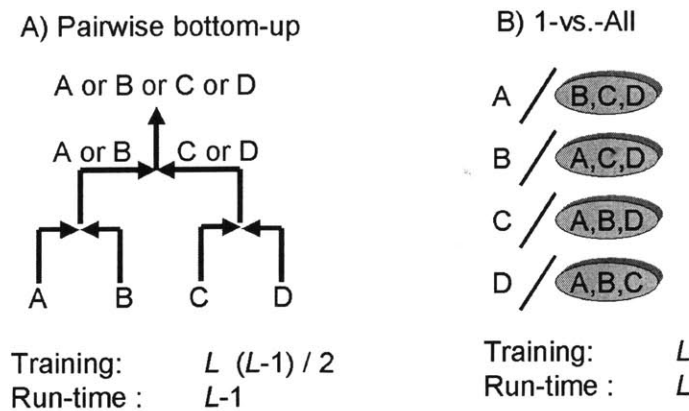
D / A,B,C

Training: L
Run-time : L

Figure 2.2: Diagram of pairwise and one-vs-all approach of multi-class classification with binary classifiers

2.2　3D Morphable Models

2.2.1　Description

Morphing between 3D objects is a well-known computer graphics technique. 3D morphable face models apply the general concept into the vector space representation of face models. The main idea behind the morphable face model approach is that given a sufficiently large database of 3D face models any arbitrary face can be generated by morphing between the ones in the database.

The 3D morphable model used in this thesis was developed by Volker Blanz and Thomas Vetter [3], who extended the 2D approach in [22]. The generation of 3D head models was done in collaboration with their lab. Their database of 3D models was built by recording the faces of 200 subjects with a 3D laser scanner. Then 3D correspondences between the head models were established in a semi-automatic way using techniques derived from optical flow computation. Using these correspondences, a new 3D face model can be generated by morphing the existing models in the database.

To create a 3D face model from a set of 2D face images, an analysis by synthesis loop is used to find the morphing parameters such that the rendered images of the 3D model are as close as possible to the input images. These parameters include shape and texture coefficients, illumination, orientation, and face position. The optimization algorithm starts with manual alignment of the average face (of the 200 head models) with the face in the image. Iteratively, the algorithm attempts to minimize the error between the synthetic reconstruction at that point with the input image with respect

to the the sum of square errors over all color channels and all pixels [3].

2.2.2 Training Images Generated from Face Models

Morphable models allow for the simplistic generation of 3D face models which are used in the training of the face recognition system. First, high quality frontal and half-profile pictures are taken of each subject under ambient lighting conditions. These images are then used used as input to the analysis by synthesis loop which yields a face model. This face model can be used to graphically render synthetic face images under varying pose and illumination conditions. Examples of the pairs of input images and corresponding synthetic images created by rendering the 3D face models are shown in Figure 2.3.
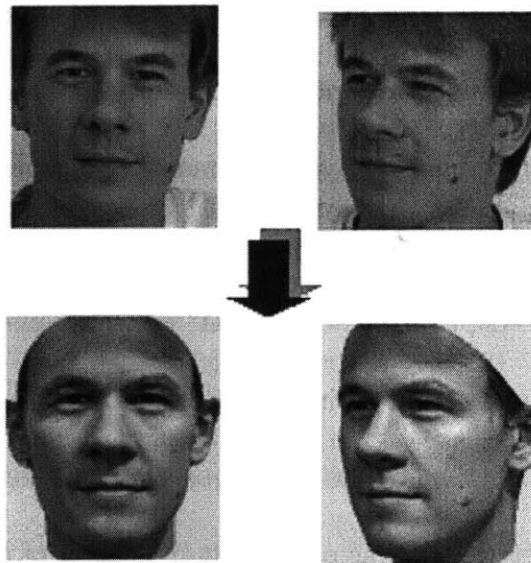

Figure 2.3: Generation of the 3D model. The top images are the real images used to generate a 3D model. The bottom images are synthetic images generated from the model. Notice the similarity between the original and synthetic images.

Using the 3D models, synthetic images such as the ones in Figure 2.4 can easily
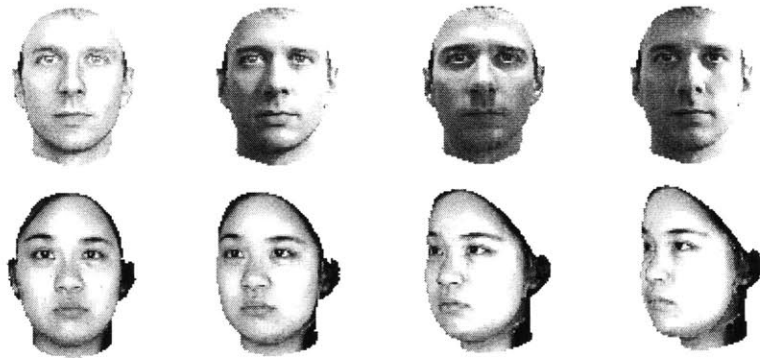
Figure 2.4: Synthetic training images. Synthetic face images generated from the 3D head models under different illuminations (top row) and different poses (bottom row).

be created by rendering the models. The 3D morphable model also provides the full 3D correspondence information which allows for automatic extraction of facial components and facial regions. This greatly simplifies the problem of obtaining the large quantity of training images required to train a robust face recognition system.

The 3D face models were used in the creation of the positive training set. The negative training set was built from randomly extracting 13655 patterns of size 58×58 from a database of non-face images. Synthetic faces were generated at a resolution of 58×58 for the 6 subjects under varying pose and illumination conditions. Specifically, the faces were rotated in depth from $-34°$ to $34°$ in $2°$ increments. Two illumination models were used to simulate real lighting conditions. One model consisted of ambient light alone, while the other model was composed of directed light in addition to ambient light. The directed light was pointed at the center of the face and positioned between $-90°$ and $+90°$ in azimuth and $0°$ and $75°$ in elevation. The angular position of directed light was incremented by $15°$ in both directions. This training set was used to train both the component-based and global detection units.

For face recognition, the training set used for detection was modified. Upon further

examination, the images used for detection were unrealistic in their reflectance and lighting schemes. Some adjustments were made to the lighting conditions and skin reflectance parameters in order to introduce more realistic training images which would better represent the real world images. This modified training set of 11526 images was used for both the component-based and global recognition units.

Chapter 3

Face Detection

As previously mentioned, the task of face detection is crucial to a successful recognition. Two face detection units, component-based and global, are described in detail in this Chapter. These outputs of these two units will be used as inputs to the face recognition units described in the next Chapter.

3.1 Component-based Face Detection

The component-based detector performs two tasks: the detection of the face in a given input image and the extraction of the facial components which are later needed to recognize the face.

The architecture of a two-level component-based face detection system was taken from [8] and shown schematically in Figure 3.1. The first level consists of fourteen independent component support vector machine classifiers. The shape of the components was learned by an algorithm described in [9] to achieve optimal detection

	Eyebrows	Eyes	Nose Bridge	Nose	Nostrils	Cheeks	Mouth	Lip	Mouth Corners
Width	19	17	18	15	22	21	31	13	18
Height	15	17	16	20	12	20	15	16	11

Table 3.1: Size of the fourteen components of the component-based detector.

results. Figure 3.2 shows examples of the fourteen components and their sizes are listed in Table 1. The components included the left and right eyebrows, left and right eyes, bridge of the nose, nose, left and right nostrils, left and right cheeks, mouth, lip, and left and right mouth corners.

On the second level, the geometrical classifier takes the first level outputs of the component classifiers and performs the face detection. The first level outputs consisted of the maximum continuous outputs of the component classifiers within rectangular search regions around the expected positions of the components in a 58 by 58 window. The rectangular search regions were determined from statistical information about the location of the components in the training images. In addition, the geometrical classifier also used the maxima locations (x, y positions relative to the the upper left corner of the 58 by 58 window).

On a novel image, a 58 times 58 resolution window slides across the image and first detects the components and then feeds the information into the second level classifier. The maximum result over the entire image is taken to be the location of the face. Once the face is detected, the components are extracted for later use in the face recognition unit.

The training set, as previously discussed in Chapter 2.2.2, was used in training the component-based face detector. To train the first level classifiers, fourteen components were automatically extracted from every synthetic face image based on the
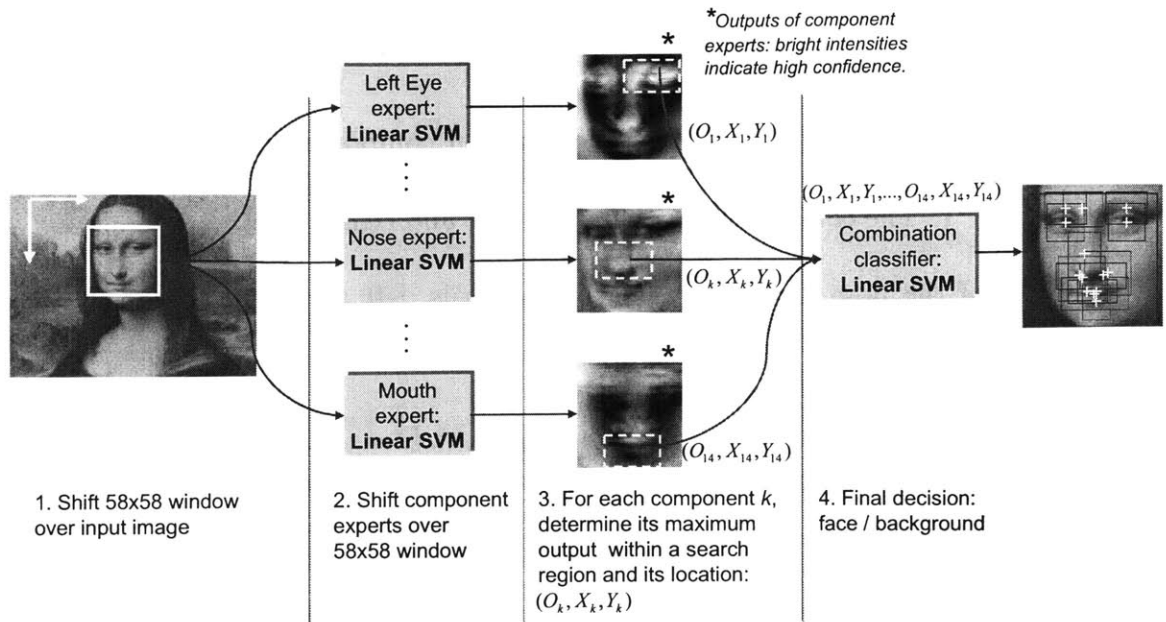
19

Figure 3.1: System overview of the component-based face detector using three components.
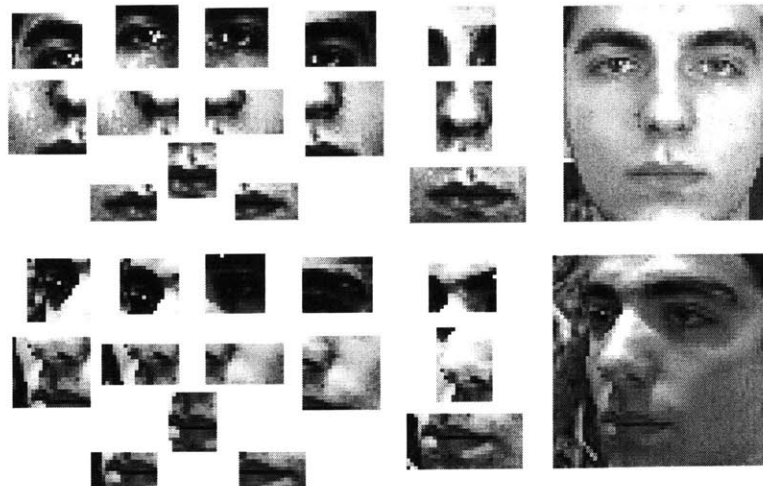

Figure 3.2: Examples of the fourteen components extracted from a frontal view and half profile view of a face.

correspondence information given by the morphable model. Similarly, negative components, corresponding to the size of their respective positive component images, were extracted from the negative training set. Each component classifier was trained on the previously described set of extracted facial components and on a set of randomly selected non-face patterns. These components are shown in Figure 3.2. Next, the second level face detector was trained with the output of the first level classifiers on face images and non-face images.

Various experiments were performed in the attempt to produce the best face detection unit. On a kernel level, component and geometrical classifiers were trained with both linear and polynomial kernels. The slight increase in accuracy obtained by using polynomial classifiers was determined to be not worth the tradeoff in computing time. Another problem was the slight shifting of detection when the position information was included in the training of the face detector. This problem was mitigated by training the face detector without position information. Different types of histogram equalization and feature types (gray value, gradients, Harr wavelets) were also attempted. The resulting face detection used gray pixel value features and was produced solely from the face models in the database.

Both the face detection unit previously described and the face detection unit used in the [9] were used as inputs to the face recognition units. The face detection unit from the previous system performed more accurately (approximately 10%). Since the concentration of this thesis was on face recognition and not on face detection, the face detector from the previous system was used as the final face detection unit. The previous system developed their face detection unit more extensively, namely

with more synthetic head models, more difficult training images, and bootstrapping techniques.

3.2   Global Face Detection

Global face detection differs from component-based face detection in that the whole face is treated as one unit. The same training set, as described in Chapter 2.2.2, was used. Instead of extracting individual components, a region of the face was extracted for the training of the face detector. Using the correspondence information given by the morphable model, the extracted region was based on the position of the upper corner of the eyebrows and the middle of the chin. Histogram equalization is then performed on each extracted face image. Figure 3.3 shows a sample training images and the histogram equalized results. Finally the gray pixel values of each image are stored in a single feature vector. Similarly, the gray pixel values of the histogram equalized negative training images are stored into a single feature vector.

Figure 3.3: Global face detection training images. The original images are on the left, and the histogram equalized images are on the right.

A linear SVM is trained on the positive and negative feature vectors to yield the global detector. On a novel image, a 58×58 resolution window slides across the image. The maximum result of the global classifier is taken to be the location of the face. The 58×58 region is extracted for later use in the global face recognition unit.

Chapter 4

Face Recognition

4.1   Component-based Face Recognition

Face recognition can be attempted once the face is detected in the image. The face recognition units described in this Chapter use the outputs of the face detection units from the previous Chapter. Two types of face recognition, component-based and global, are described and tested in this Chapter.

4.1.1   Architecture

The component-based face recognition system consists of six people classifiers, one for each person in the database. Each classifier was trained in a one vs. all approach. In other words, a SVM was trained from each subject in the database to separate her/him from all the other subjects.

To determine the identity of a person at runtime, the component-based face detection and component-based recognition units are used. Figure 4.1 shows the pro-

gression from a novel image to recognition. First, the component-based face detector detects the face portion of the image and extracts the components. These components are used by the component-based face recognizer to determine the identity of the person. The outputs of the different SVM face classifiers are compared. The identity associated with the face classifier with the highest normalized output is taken to be the identity of the face.
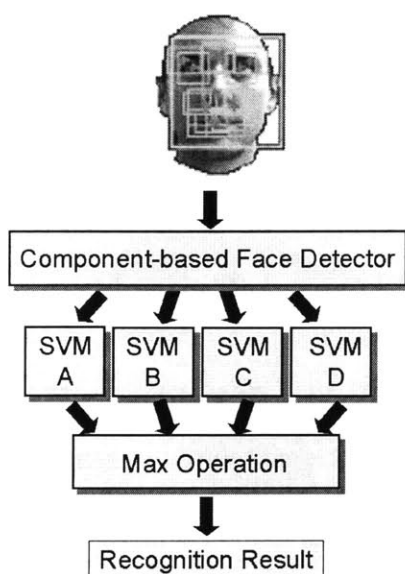

Figure 4.1: Component-based face recognition overview. The face and components are extracted by the face detection unit. The highest output of the recognition classifiers is used as the recognition result.

4.1.2 Training

From the fourteen components extracted by the face detector, only nine components were used for face recognition. Four components were eliminated because they strongly overlapped with other components or contained few gray value structure (e.g. cheeks). The lip component was also eliminated because of its inaccuracy. Fig-

24

ure 4.2 shows the composite of the nine extracted components used for face recognition for some example images. An extra histogram-equalized face region component was added to improve recognition. [1] Some examples of this face component are shown in Figure 4.3.
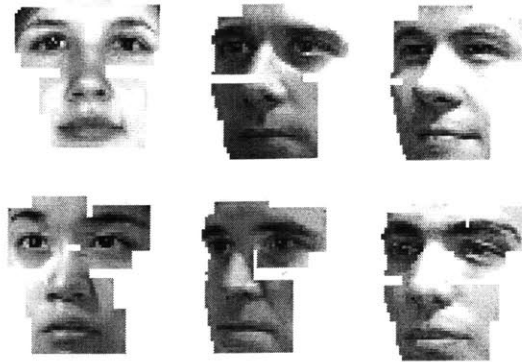

Figure 4.2: Composite of the nine components retained for face recognition.

Figure 4.3: Extra histogram equalized face component added for face recognition

The component-based face detection unit was applied to each synthetic face image in the modified training set, described in Chapter 2.2.2, to detect the components and thereby the facial region. Histogram equalization was then preformed on the bounding box around the components. The gray pixel values of each component were then taken from the histogram equalized image and combined into a single feature vector. Feature vectors were constructed for each person, and corresponding classifiers

---

[1] The location of the extra component was computed by taking the bounding box around the other nine detected components and then subtracting from the larger edge to form a square component. This square was then normalized to 40×40, histogram-equalized, and then stored.

were trained.

4.2 Global Face Recognition

The global face recognition system also consists of six people classifiers, one for each person in the database. Each classifier was also trained in a one vs. all approach. To determine the identity of a person at runtime, the global face detection and global recognition units are used. Figure 4.4 shows the progression from a novel image to recognition. First, the global face detector detects and extracts the face region. This region is then used by the global face recognizer to determine the identity of the person. The outputs of the global SVM face classifiers are compared. The identity associated with the face classifier with the highest normalized output is taken to be the identity of the face.

The same modified training images which were used for component-based recognition were also used for global face recognition. The global face detection unit was applied to each synthetic face image in the training set to detect the facial region. Similar to the global face detection unit, these extracted face regions were histogram-equalized and stored in feature vectors. Feature vectors were constructed for each person, and corresponding classifiers were also trained for each person.
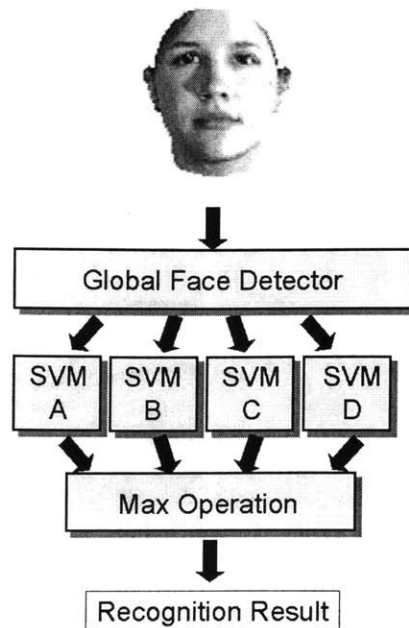
Figure 4.4: Global face recognition overview. The face is extracted by the face detection unit and fed to the global face recognition unit. The highest output of the recognition classifiers is used as a recognition result.

4.3 Results

4.3.1 Test Set

A test set was created by taking images of the six people in the database. The subjects were asked to rotate their faces in depth and the lighting conditions were changed by moving a light source around the subject. The test set consisted of 200 images of each person, under various pose and illumination conditions. Figure 4.5 contains examples of the images in the test set. Although subjects were asked to rotate only in depth, some faces were slightly rotated in the image plane was well.
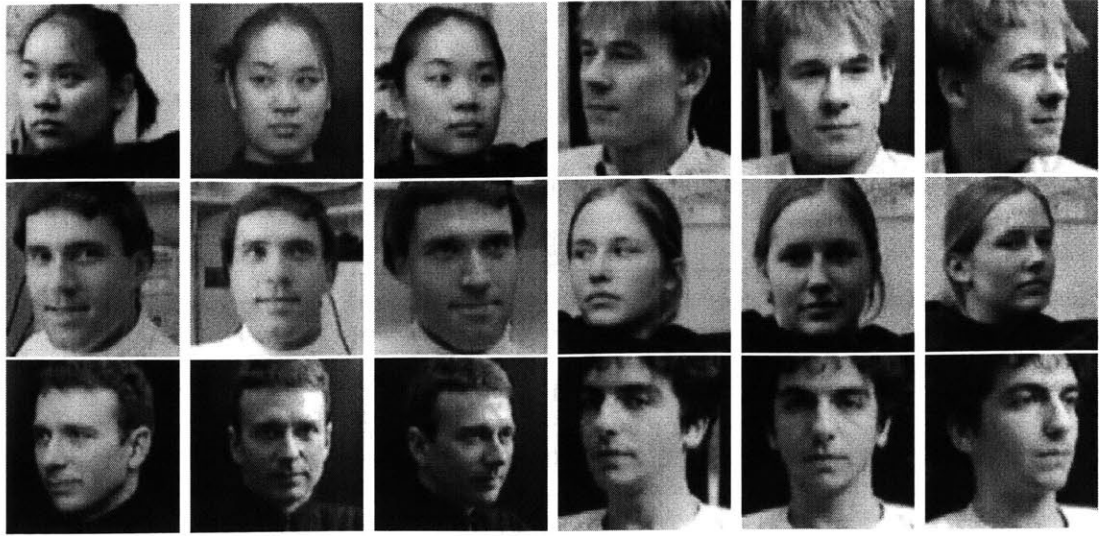
Figure 4.5: Examples of the real test set. Note the variety of poses and illumination conditions.

4.3.2 Experiments

Issues with Face Detection

As previously mentioned, a correct detection is extremely crucial to the subsequent face recognition. No exact numbers can be given on false face detections. The exact effect of a small shift is also not clearly defined. By sight, the number of clearly incorrect detections for global detection, examples of which are shown in Figure 4.7, is 1.8%. The number of clearly incorrect detections for component-based detection, examples of which are shown in Figure 4.6, is 3.8%.
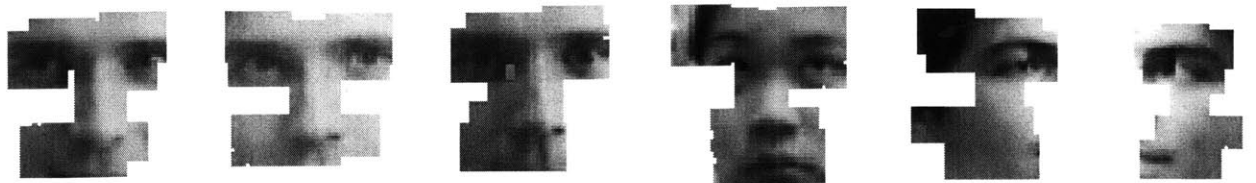

Figure 4.6: Incorrect component-based face detections.

Figure 4.7: Incorrect global face detections

Linear vs. Polynomial Kernels

For both component-based and global face recognition, linear and 2nd degree polynomial SVMs were trained. For the whole face system the accuracy of the polynomial classifier exceeds that of its linear counterpart while for the component-based system the polynomial and linear classifiers are roughly equal. In the component-based case, the polynomial kernels only performed 1.7% more accurately for maximum recognition; in the global case, the polynomial kernels performed 10% more accurately for maximum recognition. The resulting ROC curves can be seen in Figure 4.8.

Component-based vs. Global Classification

Trained on the same data, the component-based recognition system significantly outperformed the global recognition system. The resulting ROC curves for the 2nd degree polynomial versions of global and component-based recognition on the test set can be seen in Figure 4.9. Component-based recognition was approximately 50% more accurate than the global recognition.

This large discrepancy in results can be attributed to two main factors: difference in robustness and preprocessing between the two models. Figure 4.10 compares the performance of a simple global and component template on sample rotations in pose. The component template can adjust to the slight variations in pose, while the fixed global template is less robust against changes in the depth and in the image plane.

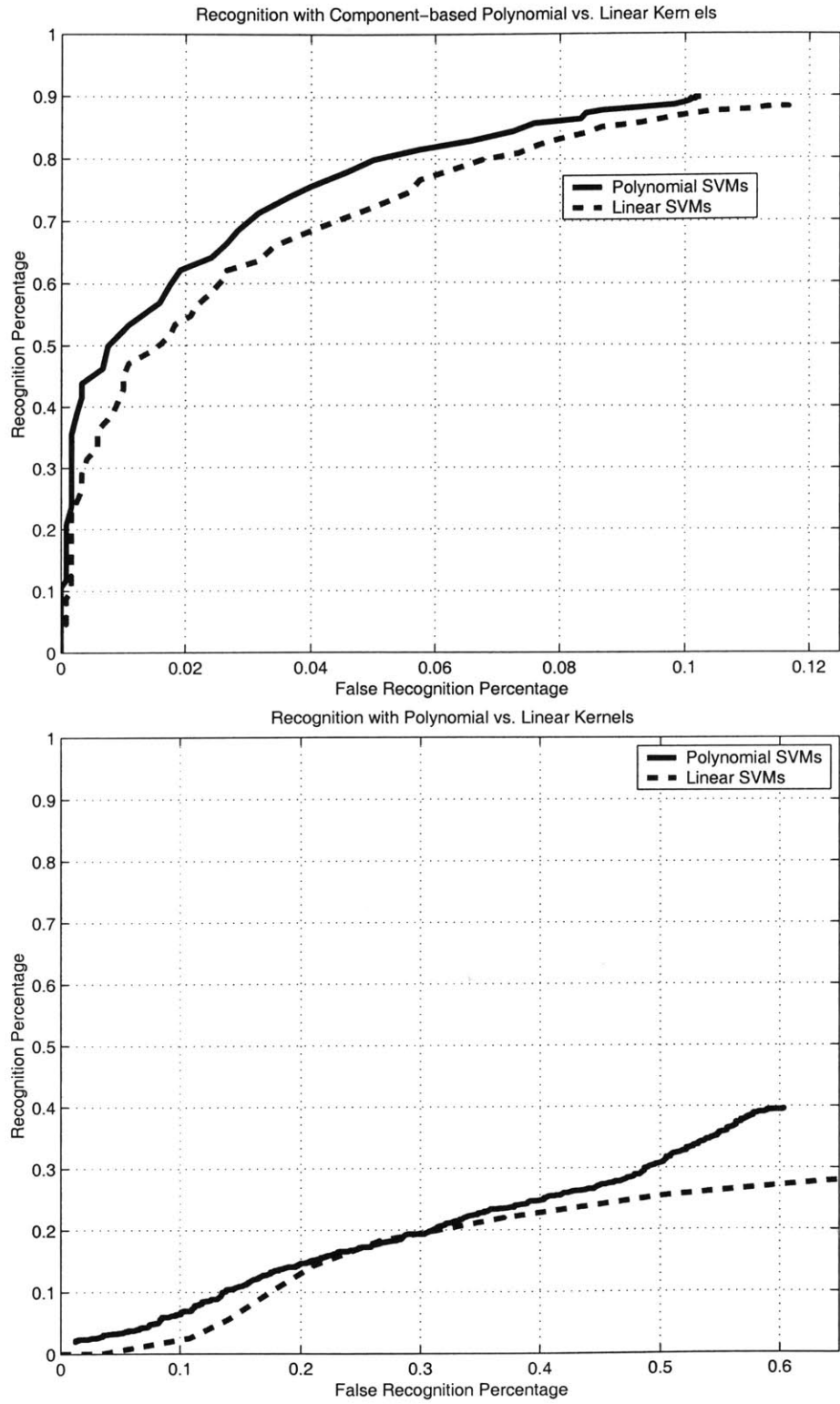Figure 4.8: ROC curves for linear and polynomial kernels. The top diagram shows the performance of linear and polynomial component-based recognition and the bottom diagram shows the performance of linear and polynomial global recognition.
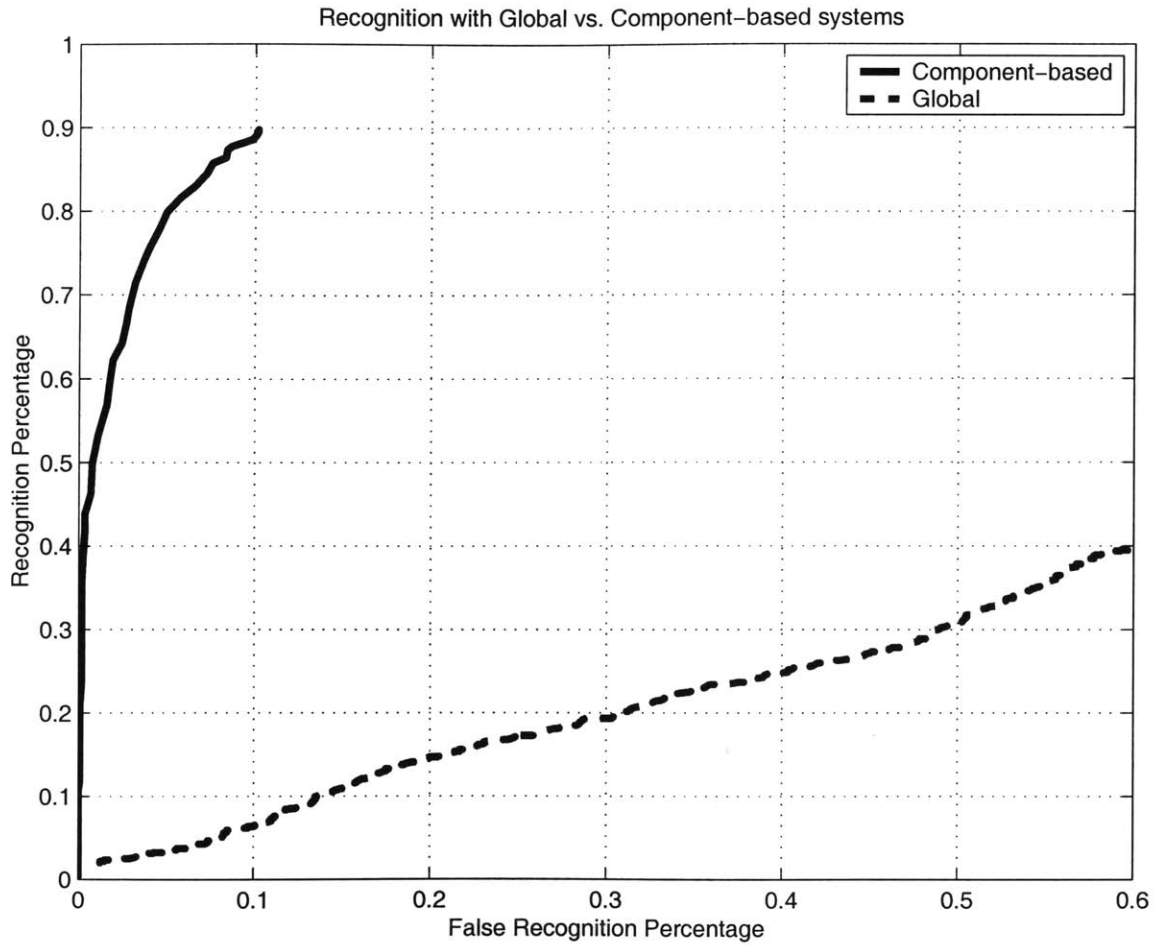
Figure 4.9: ROC curves for component-based and global classifiers. This diagram compares the performance of global and component-based polynomial recognition systems trained on the same data.

This can be explained by the flexibility in the geometry of the components in the component-based approach.

In addition the the inherent sensitivity to rotation, the real test set also had non-uniform backgrounds since the images were taken under real conditions. Component-based recognition had the benefit of performing histogram equalization on just the bounding box around the components. This eliminated any effect of a non-uniform background. However, global recognition performed histogram equalization on the extracted face region which occasionally contained significantly distracting backgrounds. This background problem most likely contributed to the further decrease in performance. Thus, the combination of the greater robustness of component-based face recognition and superior histogram processing capabilities most likely resulted in the discrepancy in performance.

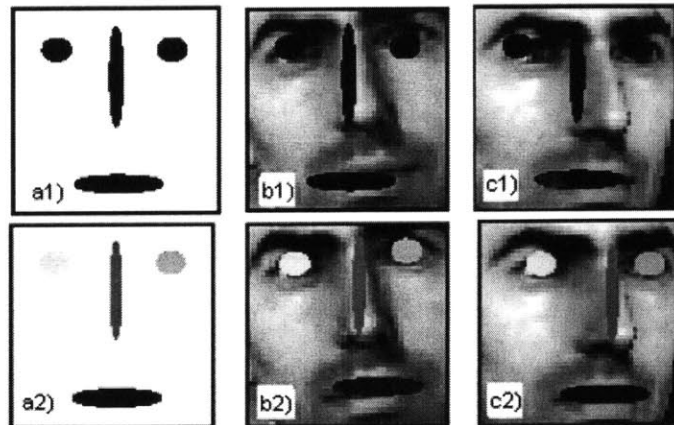Figure 4.10: Sensitivity of global classification to changes of pose. a1) and a2) are the global and component-based schematic component templates, respectively. b1) and b2) show the template performances on rotations in the image plane, while c1) and c2) show the performances on rotations in depth. The component-based template has a flexibility geometry which compensates more robustly for changes in pose.

Chapter 5

Conclusions and Suggestions for

Future Work

This thesis involved a new development face recognition with the incorporation of 3D morphable models and component-based face recognition. This combination allowed the training of a face recognition system which required only two face images of each person. From these two images, 3D face models were computed and then used to render a large number of synthetic images under varying poses and lighting conditions. These synthetic images were then used to train a component-based face detection and recognition system. A global face detection and recognition system was also trained for comparison.

Results on real test images show that the component-based recognition system clearly outperforms a comparable whole face recognition system. Component-based techniques yielded a recognition rate of 90% for faces rotated to approximately to $\pm 45°$ in depth under varying illumination conditions. In comparison, global techniques only

performed at a recognition rate of 40% on the same test set. These results point to the overall robustness of component-based face recognition in comparison with global recognition.

The success of this thesis also proves the feasibility of using 3D morphable models in training face recognition systems. Future work includes the expansion of the system to include a wider range of rotations and illumination conditions. Extension of of pose and illumination invariance would involve training on synthetic images over a larger range of views and conditions. However, when expanding near the profile view, other complications may arise. For example, the components needed for profile recognition might be different than those used for mostly frontal recognition. In that case, components must be relearned.

Another area of improvement is the accuracy in face detection, which was not explored in depth in this thesis. In [2], face detection accuracy was improved by using a more sophisticated geometrical model for the positions of the components along with more carefully selected negative training data.

Finally, the number of faces currently in the database is not large and could be increased. However, increasing the number of people causes additional issues. First, the speed of the system will decrease significantly with additional users. The number of classifiers in a one vs. all approach grows linearly with the number of classes (people). Another issue is separability between people in the database. Second degree polynomial SVMs are able to separate six people, but will not necessarily be able to separate a hundred. New kernels (e.g. higher degree polynomials and gaussians) might have to be considered as the system grows. The importance of choosing the

right components/features to separate the people also becomes increasingly important with additional people.

Bibliography

[1] D. Beymer. Face recognition under varying pose. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 756–761, 1994.

[2] S. M. Bileschi and B. Heisele. Advances in component-based face detection. In *Proceedings of Pattern Recognition with Support Vector Machines, First International Workshop, SVM 2002*, pages 135–143, Niagara Falls, 2002.

[3] V. Blanz and T. Vetter. A morphable model for synthesis of 3D faces. In *Computer Graphics Proceedings SIGGRAPH*, pages 187–194, Los Angeles, 1999.

[4] R. Brunelli and T. Poggio. Face recognition: Features versus templates. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(10):1042–1052, 1993.

[5] Recognition Christopher Burges. A tutorial on support vector machines for pattern.

[6] C. Cortes and V. Vapnik. Support vector networks. *Machine Learning*, 20:1–25, 1995.

[7] G. Guodong, S. Li, and C. Kapluk. Face recognition by support vector machines. In *Proc. IEEE International Conference on Automatic Face and Gesture Recognition*, pages 196–201, 2000.

[8] B. Heisele, P. Ho, and T. Poggio. Face recognition with support vector machines: global versus component-based approach. In *Proc. 8th International Conference on Computer Vision*, volume 2, pages 688–694, Vancouver, 2001.

[9] B. Heisele, T. Serre, M. Pontil, and T. Poggio. Component-based face detection. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, pages 657–662, Hawaii, 2001.

[10] B. Moghaddam, W. Wahid, and A. Pentland. Beyond eigenfaces: probabilistic matching for face recognition. In *Proc. IEEE International Conference on Automatic Face and Gesture Recognition*, pages 30–35, 1998.

[11] A. Mohan, C. Papageorgiou, and T. Poggio. Example-based object detection in images by components. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, volume 23, pages 349–361, April 2001.

[12] A.V. Nefian and M.H. Hayes. An embedded HMM-based approach for face detection and recognition. In *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 6, pages 3553–3556, 1999.

[13] J. Platt, N. Cristianini, and J. Shawe-Taylor. Large margin dags for multiclass classification. *Advances in Neural Information Processing Systems*, 2000.

[14] M. Pontil and A. Verri. Support vector machines for 3-d object recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 637–646, 1998.

[15] R. Rifkin. *Everything old is new again: a fresh look at historical approaches in machine learning*. PhD thesis, M.I.T., 2002.

[16] S. McKenna S. Gong and A. Psarrou. *Dynamic Vision: From Images to Face Recognition*. Imperial College Press, Ireland, 2000.

[17] H. Schneiderman and T. Kanade. A statistical method for 3D object detection applied to faces and cars. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 746–751, 2000.

[18] B. Schölkopf, C. Burges, and V. Vapnik. Extracting support data for a given task. In U.M. Fayyad and R. Uthurusamy, editors, *Proceedings of the First International Conference on Knowledge Discovery and Data Mining*, Menlo Park, CA, 1995. AAAI Press.

[19] M. Pontil T. Evgeniou and T. Poggio. Regularization networks and support vector machines. In *Advances in Computational Mathematics*, volume 13, pages 1–50, 2000.

[20] A.N. Tikhonov and V.Y. Arsenin. *Solutions of Ill-posed Problems*. W.H. Winston, 1977.

[21] V. Vapnik. *Statistical learning theory*. John Wiley and Sons, New York, 1998.