Experiments in Detecting Facial Features

Gloria Chow Xiaobo Li
Department of Computing Science
University of Alberta
Edmonton, Alberta
CANADA T6G 2H1

Abstract

This paper reports experiments in detecting facial features (eyes, mouth, and chin-lines) from a front-view ID-type picture. First, at low resolution, a context module defines a face template in terms of intensity valley regions, using morphological filtering and 8-connected blob coloring. The objective is to generate a list of hypothesized face locations ranked by face likelihood. The detailed analysis is left for the high resolution eye and mouth modules, to confirm as well as refine the locations and shapes of their respective features of interest. The detection is done via a two-step modelling approach based on the Hough transform and the deformable template technique. The results show that facial features can be located very quickly with Adequate or better fit in over 80% of the images with the proposed system. We also report some preliminary experiments on chin-line following.

1 Introduction

In the research of face recognition, little work has been done so far on the automatic detection of facial features that is essential to the eventual recognition goal. The difficulty is the lack of formalism in defining what a face is. The problems we are faced with in automatic facial feature detection therefore exist on two levels:

(1) On the feature level, we need to derive a model for individual features that is descriptive enough to embody the common shape, but yet flexible enough to handle some degree of variation. As well for practicality, the technique employing such a model must be executable within a reasonably short time.

(2) On the face level, we are concerned with the overall system integrity. In specific, we would like to address the question of how individual components in the system should interact in order to maximize the performance.

In reference to the first problem, a number of different techniques with varying degrees of complexity had been employed in the past. Extraction techniques based on signature search [9][2][7], contour following [4], and fixed template matching [1] were found to be inadequate mostly because they had failed to satisfy the representational requirement. The description in each case was either too simple to be used as a reliable recognition measure or too restrictive to capture variations of the same general shape. The few successful attempts in this area [6] [8] [12] [10] [5] have been based on higher level modelling techniques, such as template and spring model, Hough transform, and deformable template model.

Even fewer attempts have been made to achieve a complete facial feature location system. Many of the research mentioned earlier [8][12][10] assumed that processing could be limited to a localized region without investigating how it could be done. Others [9][2][7] adopted a sequential approach so that the successful location of one feature would serve to limit the search of other features. While it is computationally attractive and intuitive, it will allow error to propagate and accumulate from module to module. The integrity of the overall system in such sequential design is highly questionable. The solution is therefore to adopt a *hypothesis and verify* approach such as that suggested by Govindaraju *et al.* [6] and Craw *et al.* [5]. However, the hypothesis genera-

*This research is supported in part by the Canadian Natural Sciences and Engineering Research Council under Grant OGP9198. In addition, the first author is supported by scholarships from the Province of Alberta and the Department of Computing Science at the University of Alberta.

tion mechanism needs to be simplified to be used effectively in a practical system.

In this paper, we report our experiments in extracting the shapes and locations of eyes and mouth from a *relatively unposed* head and shoulder picture of a clean shaven subject without spectacle. By relatively unposed, we mean that the facial image is assumed to be a front-view ID-type picture, but the face location, head size, lighting, and background are allowed to vary. The system is composed of three modules: a context module that generates hypothesized face locations (ie. face contexts), and eye and mouth modules, whose aims are to confirm as well as refine the locations and shapes of their respective features of interest. The context module is based on a speculation and confirmation concept which employs a simple morphological filter based segmentation technique that is very quick to execute and yield reliable results. The eye and mouth modules, on the other hand, are more complex and are based on a combined approach of the Hough transform and deformable template techniques. The important design objective here is to eliminate dependency among confirmation modules, so that if one fails, the rest can still continue. As well, the confirmation modules can be computed in parallel, possibly on a multiple processor system. We also report preliminary experiments on chin line detection. The development and results of this system is discussed in detail in the remainder of this paper.

2 Context Module

The design of the facial context module is based upon the observation that though the sizes and distance among facial features may vary, their overall spatial arrangement remains the same. A facial context is therefore simply a collection of distinct regions whose spatial arrangement resembles that of the eyes, eyebrows and mouth. The approach here is to first identify all distinct regions and then attempt to group these regions into plausible contexts. The eventual objective is to obtain a list of potential face contexts ordered by face likelihood.

Image Segmentation:

Since there is a separate confirmation step for verification and refinement, the emphasis of our segmentation step will not be in precision. Rather, we just want to locate regions of interest roughly and quickly. It is broken down into three steps as follows:

Resolution Reduction: The original 256 × 256 image is reduced to 64 × 64 by a 4 × 4 averaging operation.

Valley Detection: Morphological *opening residue* operation with a 5 × 5 circle mask is used to extract dark regions (ie. eyes, eyebrows, and mouth) in the reduced image [11].

Region Idenification: 8-connected blob coloring is employed to assemble the detected pixels into distinct regions.

In order to facilitate reasoning in the following facial context evaluation step, the segmented image is further condensed into a list of distinct regions with the following attributes: size, length and width of the bounding rectangle, average, maximum and minimum grey level, and centroid location.

Context Constraint Model:

Pair Analysis: The context constraint model is based primarily on the successful location of the eyes. The eyes in this case are modelled simply as a pair of similarly shaped horizontal regions rated and thresholded in terms of their shape resemblance and position correspondence. Shape resemblance is measured by overlapping the bounding rectangle of the two's and taking the proportion between their symmetric difference and intersection. And positional correspondence is derived by forming a vector between the centroid of the two regions and measuring the amount of deviation of this vector from the horizon.

Context Completion: Once the pair (ie. eye) regions are identified, the module then attempts to complete the context by locating *plausible* regions positioned at the estimated locations of the remaining parts (ie. eyebrows and mouth). The total cost of missing components is simply the sum of all the missing parts.

Context Evaluation: The final context evaluation is based on a combined cost of the two previous measure. This combined cost measure is a heuristic derived through conservative experiments. It can not guarantee that the correct context will be ranked first in the list, but it will certainly include it in the list for further confirmation. This cost measure can be further tuned

and modified to reduce the time requirement.

Experimental Results and Discussion:

The context module described has been implemented and tested with a set of 67 images, all of which are without spectacles. The correctness of each context is judged by whether the true eyes and mouth are included in their estimated positions. Figure 1 illustrates such example contexts. Although the estimate may seem to be

(a) Correct Context (b) Wrong Context

Figure 1: Example Contexts

over-generous, it is kept conservative in order to account for the potential error in locating the centroids of the eye regions.

The actual run time ranged from a low of 0.010 sec to a high of 2.117 sec. The average run time was 0.863 sec with a 0.604 sec standard deviation. Of the 67 test images, the context module correctly identified the real face as being one of the potential contexts in 64 images. Fifty-seven of these ranked it as being the first in the list, while the remaining 7 ranked it within the top 3 potential contexts. Of the 3 that failed, lighting was generally insufficient. As a result, the segmentation step was unable to pick up one or both of the eye regions. It was expected that if pictures were of better quality, better performance could be achieved with no modification to the module. The result is very promising since it places the module at a 96% hit rate with minimal execution time.

3 Eye Module

The objective of the eye module is to confirm and refine the eye location in the given region. Our approach is a two-step technique based on the use of Hough transform and deformable template modelling. In the first step, we shall attempt to locate the irises modelled as a pair of circles using the circle Hough transform technique, as in [8]. After the irises have been located, a template of only the bounding parabolas is used to correctly orient and complete the description of the eye pair.

Circle Hough Transform:

Our circle Hough transform routine has been designed to incorporate gradient direction information. The equation of a circle used is $(x-a)^2+(y-b)^2=(d/2)^2$ where (a,b) is the circle center, and d is the diameter. Given the coordinates of a valid edge point, (x,y), the gradient angle θ and its error $\delta\theta$, the centers of all potential circles will lie on an arc of $\theta+\pi\pm\delta\theta$ at a distance of $\frac{d}{2}$ away from the edge point. The parameter space is therefore a 3 dimensional one with axes a, b, and d. A modified Midpoint circle algorithm is used to generate these arcs efficiently. Details of this routine can be found in [3].

In order to account for potential edge location error, a local averaging operation over the $a-b$ plane within the parameter space is first performed. As well, a preliminary screening based on statistical measures is applied. For each diameter R, we calculate the standard deviation σ and mean μ of the accumulated counts. Only cells with accumulated count greater than $\sigma+\mu$ are screened through to be examined in detail. Finally, the evaluation of each circle is based on both its diameter and accumulated count: $SC_{circle} = count * \frac{count}{diameter}$.

However, the combination of edge inaccuracy and accidental alignment of edge points produces a *blurred* parameter space which makes distinction of good and bad circles difficult. To overcome this problem, we apply an additional constraint in circle selection, namely a pair of identically sized circles, one in each eye region, must be located. The circle pair is evaluated using the following equation: $SC_{pair} = SC_{circle1} + SC_{circle2} - tilt^2$ where $SC_{circle1}$ and $SC_{circle2}$ are the score computed from previous equation for the two circles, and *tilt* is the angular measure in radians between the centers of the two circles. The equation SC_{pair} is used in a ranking step to obtain an ordered list of potential iris pairs. In the absence of a better measure for the time being, only the first pair is used in the deformable template step.

This can be refined to an iterative search down the potential iris list, once a reliable measure has been established to distinguish a good fit from a bad.

Deformable Template for Eye Boundary:

The eye boundary template, modified from that proposed by Yuille [12], is composed of only two parabolic curves. In order to utilize the symmetry between the two eyes, the final template used is composed of two instances of that templates. The energy function of our combined template can be divided into five separate pieces: the upper and lower parabolas for the two eyes, and the combined shape function.

The energy function used is based only on edges and internal constraint force. The upper and lower parabola edge functionals are simply defined to be the average edge strength over the boundary of the upper and lower lids respectively for each eye. The shape functional, on the other hand, is composed of a number of shape and symmetry terms aimed at achieving certain desired proportionality among template parameters.

The evaluation and optimization of this template are done using the a modified Midpoint conic generation routine and Downhill Simplex respectively. The details of their implementations and actual system variables used can be found in [3].

Experimental Results and Discussion:

Our investigation was aimed at testing the detection capability of the eye module. The test was performed on proper eye regions, selected from the 64 images that were correctly identified by the context module. The objective was to decide whether given the proper region, the eye module could correctly extract the eyes. The final template fit was overlaid on the original image and subjectively rated for goodness-of-fit. The results are classified into four rating categories similar to that used by Shackleton *et al.* [10]. The results of this run are summarized in Table 1. An example of each is shown in Figure 2.

In general, it was found that a relatively **Good** fit can be attained if the image is well contrasted and of a bigger size. Yet, many images which were taken early on in this investigation have lower picture quality and less effective use of available image area (ie. the head size is relatively small).

Rating	Number	Description
Good	19 (29%)	well fitted
Adequate	35 (55%)	resaonable fit centered at the correct position.
Marginal	6 (10%)	some error in fit.
No Fit	4 (6%)	failed completely

Table 1: Eye Module Results

This contributed to the high percentage in the **Adequate** category. In those cases, the edge information was less defined, therefore the fitted template might be slightly off at some places, though the overall fit remained fairly close.

Among the **Marginal** and **No Fit** cases, the major source of error came from the iris size, an important reference measure for the deformable template step. This was partly due to the image quality and resolution problems mentioned earlier. Because of the absence of a measure to reliably distinguish a *good* fit from a *bad*, only the first iris pair is used in the deformable template step. Under poor image condition, the first iris pair ranked by heuristics was often not the real iris. Hence, this accounted for some of our **Marginal** and **No Fit** cases. We expect the percentage of **Good** to **Adequate** cases to improve, once a reliable distinguishing criteria is established. However, we do note that there is an important problem inherent to our eye module. And it lies on the use of the circle Hough transform to locate the iris. In doing so we assume that the iris will appear as a circular region. However, for many individuals with small eyes (or more precisely, narrow eye openings), a large portion of this circular iris is occluded.

4 Mouth Module

The objective of the mouth module is to confirm and locate precisely where the mouth is within the given region. The conceptual design of the mouth module parallels that of the eye module and is divided into two stages. The first stage uses the classical Hough transform technique to identify long horizontal edge segments within the subregion supplied by the context module. Based on this refined estimate, the 2nd stage will employ the deformable template technique as proposed

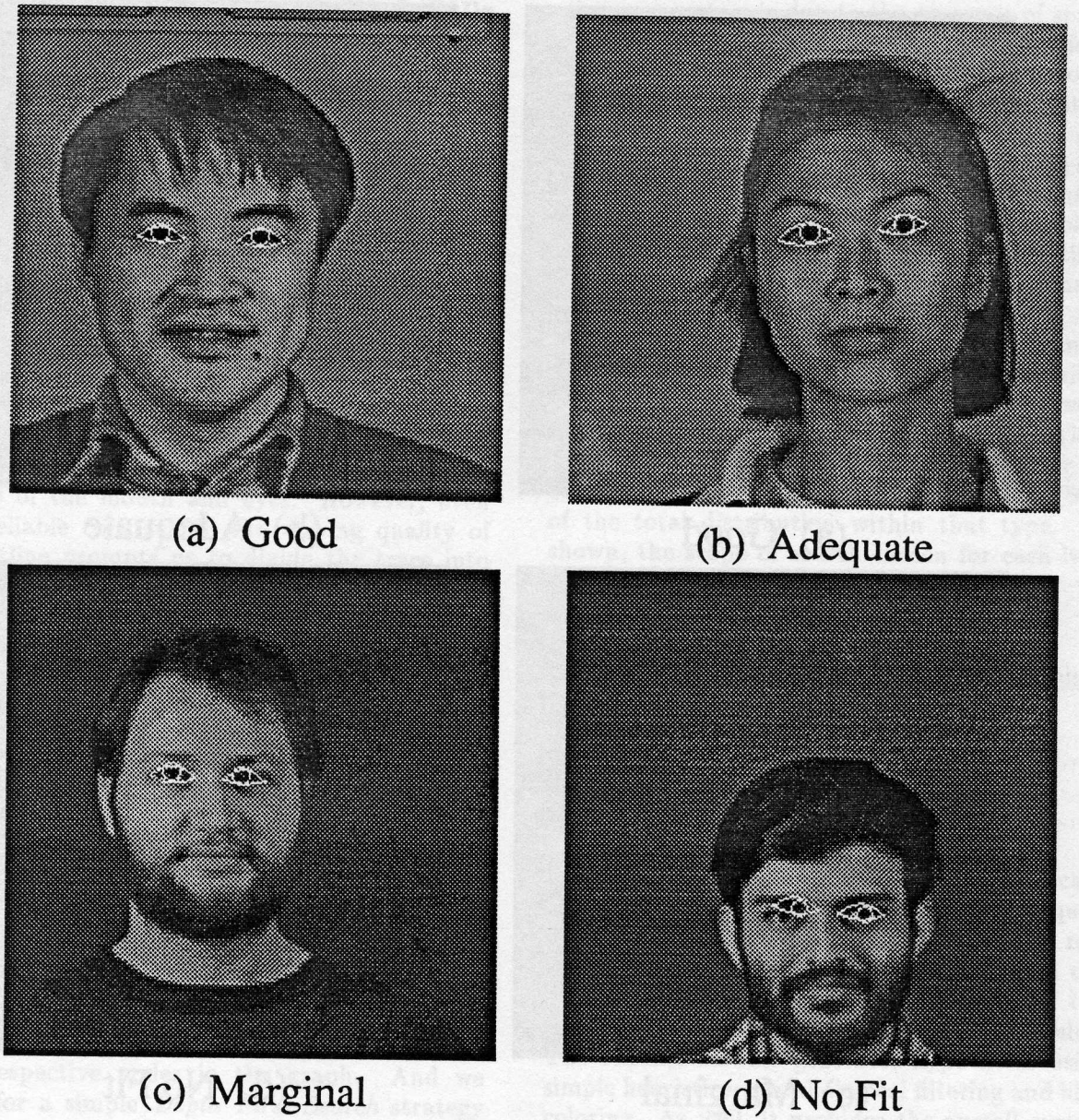

Figure 2. Eye Template Fit

by Yuille to identify the precise location and shape of the mouth.

The experimentation for this module is set up in the same format as that of the eye module. We tested the mouth module with 42 images. This is a subset of the images we used in a similar experiment for the eye module. The reason is that our mouth template is designed to handle closed mouths and clean shaven individuals only, and therefore some of the images were excluded. The results are again evaluated subjectively using the same rating categories as that of the eye module. An example of each is shown in Figure 3. The results are summarized in Table 2.

Rating	Number
Good	15 (36%)
Adequate	19 (45%)
Marginal	5 (12%)
No Fit	3 (7%)

Table 2: Mouth Module Results

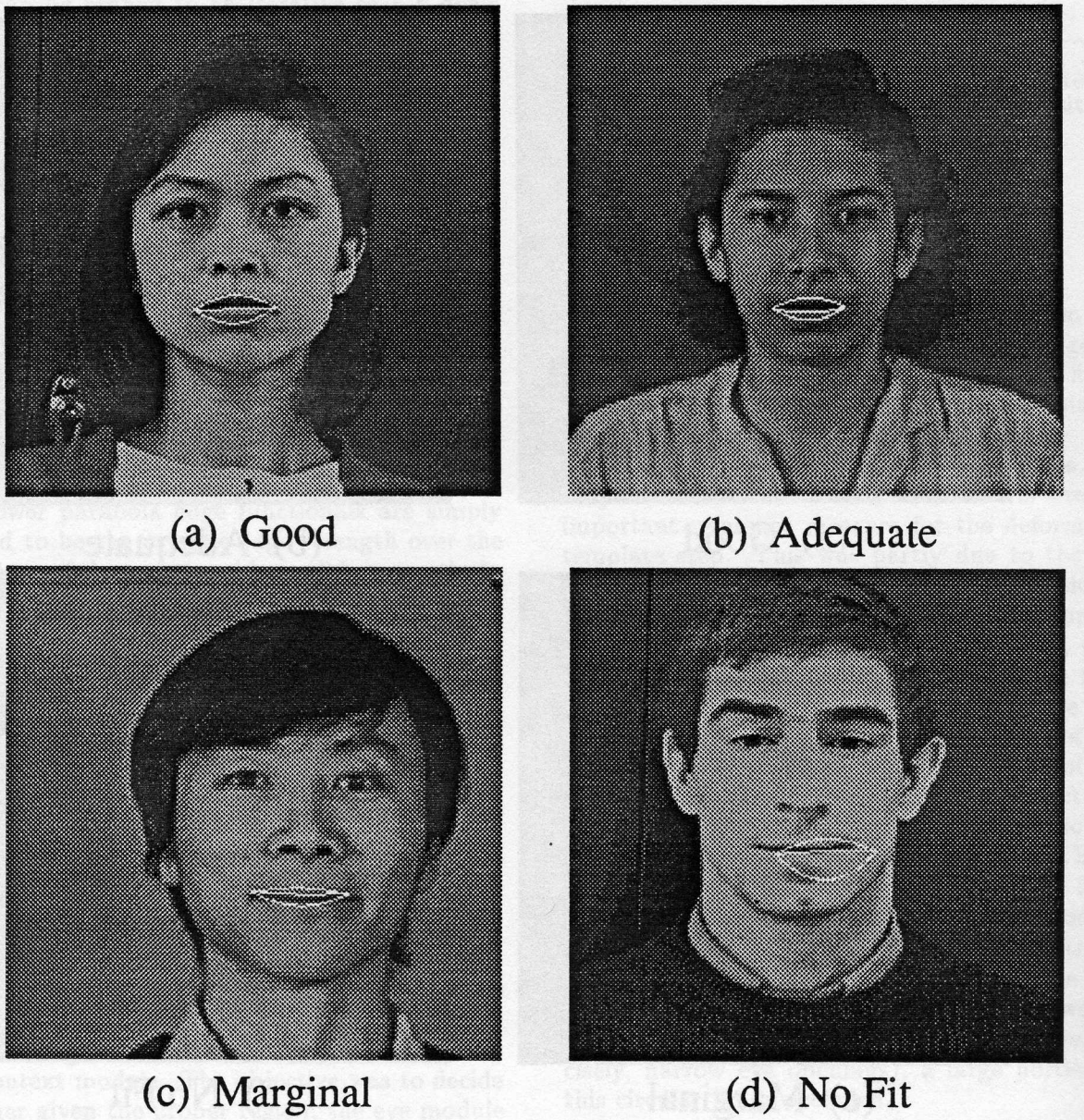

Figure 3. Mouth Template Fit

The findings here are generally the same as that for the eye module: the mouth module is capable of extracting a **Good** to **Adequate** fit in over 80% of the 42 test images. Because only the first horizontal line is used from the list generated by the Hough transform step, there is no guarantee that it is in fact part of the mouth. Most **No Fit** cases were generated by those with the incorrectly selected mouth line. This can be rectified once a robust qualitative measure is derived for distinguishing valid mouths from the invalid ones.

However, an important issue to note is the general lack of reliable edge information in the mouth image. Unlike the eyes, lips do not always appear as a clearly outlined object in an intensity image. This is particularly true for subjects with very thin and/or light colored lips.

5 Preliminary Study on Chin-line following

To complete the facial description, we also attempt to identify the lower face outline. Unlike the eyes and mouth, faces come in a variety of different shape, some are round, some are more oval, there are yet some which are more angular. Therefore, it will be very difficult to model it completely using for example a parabola, or a partial ellipse. Therefore, the strategy here is to extract it by line following instead.

Since the displacement of various components on a human face are highly constrainted, the area within which the face outline must pass through can be estimated from the already confirmed locations of the mouth and eyes. However, even with reliable constraints, the varying quality of the outline prompts us to divide the trace into three segments: left, right and bottom. The left segment outlines the left margin of the face from the eye level down to the mouth level. The right segment does for the other side. The bottom segment is the boundary between the face and the neck. The left and right segments are typically straight and more well-defined, therefore, can be used to confirm we are on the right track before attempting the more difficult trace through the bottom segment.

The trace of each segment is done separately using the same generic line follower which treats the edge elements as nodes in an undirected graph. The process of finding the best edge connecting two distinct points in an image is equivalent to finding the minimum cost path through their respective nodes in the graph. And we opted for a simple *Depth First Search* strategy for its ease of implementation and runtime efficiency. The successors for each accepted node is ranked using the following criteria to determine the search order:

(1) edge strength,
(2) proximity to the previous nodes,
(3) curvature,
(4) distance from goal, and
(5) past point performance.

The behavior of this generic line follower can then be toned by changing system parameters, ie. gap and curvature tolerance to fit the specific needs of each segment. The need for such relaxation parameters is due to the presence of spurious gaps and noises in the edge image. Therefore, instead of being in the immediate neighborhood, the successors of a node can be anywhere within a cone-shaped area projected from it. The gap and angle tolerances will then determine the size of this search cone. Furthermore, given the maximum gap and angle tolerance allowed, a lookup table of neighbor offsets indexed by discretized distance and angle can be precomputed to maximize runtime performance.

By examining the chin line generated from 30 different images, we can determine the distributions of gap size and angular difference between consecutive points for each segment type (left, right, and bottom). The optimal parameter for each type is then chosen so that it will cover 90% of the total distribution within that type. As shown, the scope of search chosen for each is as follows :

Type	Gap tolerance	Angular tolerance
left	5	30 degree
right	5	30 degree
bottom	10	60 degree

6 Conclusion

In order to construct a system to automatically locate facial features, one must address the questions of how individual features should be represented and how they should interact with one another. Our proposed system maps these two requirements onto distinct levels of processing. The context module generates hypotheses using simple heuristics, morphological filtering and blob coloring. As well, it provides the overall system control. It was capable of capturing the correct face location in 96% of the test images with an average run time of less than 1 second. The eye and mouth modules have been designed to confirm the existence and extract the precise shape and location of their respective features of interest through a combined Hough transform and deformable template technique. The results show that these facial features can be located with **Adequate** or better fit in over 80% of the images within less than 12 sec. Unlike [8][12][10], the results reported here do not depend on unknown pre-processing units or *a-priori* knowledge

of the rough location of the desired features, as well the runtime performance is far better. Future research will include enhancement to the current feature modules, addition of an alternative context module for bespectacled individuals, as well as additional modules for features, such as eyebrows, noses, or even moustaches. As well, we have to investigate how the context module should integrate the results from the various feature modules and the possibility of refining these results in an iterative process. And with the current system design, these modifications can be easily added without imposing drastic impact on the other components. Hopefully, this will eventually lead to a completely automated facial recognition system.

References

[1] Robert J. Baron. Strengths and weaknesses of computer recognition systems. In Andrew W. Young and Hadyn D. Ellis, editors, *Handbook of Research on Face Processing*, chapter 10, pages 507–18. Elsevier Science Publishers B.V., P.O. Box 1991, 1000 BZ Amsterdam, The Netherlands, 1989.

[2] S.R. Cannon, G.W. Jones, R. Campell, and N.W. Morgan. A computer vision system for identification of individuals. *IEEE IECON'86 Proceedings*, 1:347–51, 1986.

[3] Gloria Chow. Automatic extraction of facial features. Master's thesis, University of Alberta, 1992.

[4] I. Craw, H. Ellis, and J.R. Lishman. Automatic extraction of face-features. *Pattern Recognition Letters 5*, pages 183–87, 1987.

[5] I. Craw, D. Tock, and A. Bennett. Finding face features. *ECCV92*, May 1992.

[6] V. Govindaraju, D.B. Sher, R.K. Srihari, and S.N. Srihari. Locating human faces in newspaper photographs. *Proceeding of CVPR*, pages 549–554, 1989.

[7] L.C. Lambert. Evaluation and enhancement of the afit autonomous face recognition machine. Master's thesis, Air Force Institute of Technology, 1987.

[8] Mark Nixon. Eye spacing measurement for facial recognition. *SPIE Applications fo Digital Image Processing VIII*, 575:279–85, 1985.

[9] Toshiyuki Sakai, Makoto Nagao, and Takeo Kanade. Computer analysis and classification of phototgraphs of human faces. *First USA-Japan Computer Conference Proceedings*, pages 55–62, October 1972.

[10] M.A. Shackleton and W.J. Welsh. Classification of facial features for recognition. *Proceedigns of the IEEE Conference on Computer Vision and Pattern Recognition (CVIP-91)*, pages 573–79, 1991.

[11] Robert C. Vogt. *Automatic Generation of Morphological Set Recognition Algorithms.* Springer-Verlag New York Inc., 1989.

[12] Alan Yuille, David Cohen, and Peter Hallinan. Facial feature extraction by deformable templates. Technical Report 88-2, Harvard Robotics Laboratory, 1988.