Detection of Human Faces under Scale, Orientation and Viewpoint Variations

Kin Choong Yow and Roberto Cipolla
Department of Engineering
University of Cambridge
Cambridge CB2 1PZ, England

Abstract

Many current human face detection algorithms make implicit assumptions about the scale, orientation or viewpoint of faces in an image and exploit these constraints to detect and localize faces. The algorithm may be robust for the assumed conditions but however it becomes very difficult to extend the results to general imaging conditions. In an earlier paper, we proposed a feature-based face detection algorithm to detect faces in complex background. In this paper, we will examine its ability to detect faces under different scale, orientation and viewpoint. The results show that the algorithm can indeed cope with a good range of scale, orientation and viewpoint variations that is typical of a subject sitting in front of a computer terminal.

1. Introduction

In the computer vision community, there is no shortage of face detection and localization algorithms for human face processing. However, there is still a serious deficiency in algorithms which are robust to different imaging conditions, such as variations in scale, orientation, and viewpoint. Many face detection algorithms make assumptions about such conditions, and use this constraint to restrict the search space for improved results. However, this makes it difficult to extend the algorithm to situations where the assumptions are not true. Hence, it is important to find a generic algorithm that is robust under such conditions.

2. Scale, orientation and viewpoint invariant approaches

Many present approaches to solve the face detection problem have some invariance to scale, orientation and viewpoint changes, though usually some or all are assumed. The common approach to deal with scale variations is by examining the image at different scales and finding a match to a face template at each scale. Fixed-shape regions at different scales from the image are extracted, subsampled to the size of the template or filter, and then matched to the template. Experiments using this approach have been carried out by Chen *et.al.* [2], Dai *et.al.* [3], Lanitis *et.al.* [7], Sung and Poggio [14], Yang and Huang [16]. The common problem with this approach is that the face template (or filter) is restricted to only a single view of the face.

Orientation variations can be handled by a feature-based approach where facial features in the image are extracted using matched filters of the right scale and orientation (efficiently implemented by using "steerable-scalable" basis filters - Perona [11], Freeman and Adelson [4]). The detected facial features are then grouped into face candidates according to their geometrical relationships (Leung *et.al.* [8], Yow and Cipolla [17]). The difficulty, however, in using this approach is that the feature extraction stage is either not robust enough or it detects too many candidate features.

Chen *et.al.* [2] cope with viewpoint variations by performing matching in 3 views (1 fronto-parallel and 2 profile views) using a fuzzy pattern matcher based on colour. However, this approach is sensitive to the face shape and colour.

3. A feature-based face detection system

In an earlier paper (see Yow and Cipolla [19] this volume), we proposed a feature-based face detection system that is able to cope with a reasonable amount of scale, orientation and viewpoint variation. In this paper, we will review the key ideas and examine how we can extend the algorithm to cope with more general imaging conditions.

3.1. The face model

The face is modelled as a plane with 6 oriented facial features. To cope with occlusion or missing features (eyebrows, in general), the face model is decomposed into partial face groups (or PFGs) consisting of 4 features. These PFGs are further subdivided into components consisting of

2 features (horizontal and vertical pairs - Hpair and Vpair) for the purpose of perceptual grouping and evidence propagation. Fig. 1 shows the different component groups.

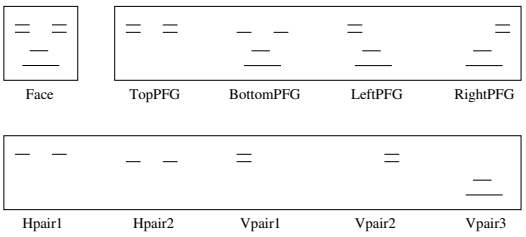


Figure 1. The face model and the face groups.

At low resolutions, all the 6 facial features will appear only as dark elongated blobs against the light background of the face. Hence, the 6 facial features are modelled as pairs of oriented edges as shown in fig. 2. The image is smoothed before the feature detection process so that any high-resolution features will take the form of the lower resolution ones. The vertical edges in the eye and nose model are important only for the labelling of the feature.

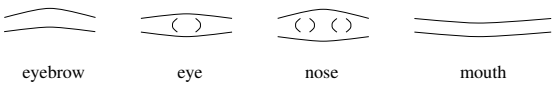


Figure 2. The facial feature models.

3.2. Perceptual grouping

The face detection process is modelled by a two stage model of perception based on Triesman [15]). The first stage extracts image information into points and regions of interest, and the second stage perform grouping and reasoning activities on these points and regions based on the Gestalt laws (Kohler [6], Koffka [5]), and on the model of the face.

The first stage, the preattentive feature selection stage, finds a list of interest points by filtering the image with a matched bandpass filter (Yow and Cipolla [17]) and then searching for local maxima. Next, the edges around each interest point are linked using a standard boundary following algorithm. A feature point is found if there are two roughly parallel edge segments with opposite polarity on both sides of the point. The extent of the feature region is then defined by finding a boundary box around the two edges (fig. 3).

Measurements of the region's image characteristics (such as edge length, edge strength, grey-level variance) are then made and stored into a feature vector $\mathbf{x}$. A facial feature candidate i is a valid facial feature j if the Mahalanobis distance $\mathcal{M}_{ij}$ of the feature vector $\mathbf{x}_i$ is within an admission threshold τ_j from the class mean μ_j, i.e.

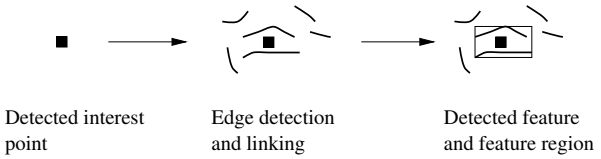


Figure 3. Preattentive feature selection.

$$\mathcal{M}_{ij} < \tau_j \text{ , where } \mathcal{M}_{ij} = (\mathbf{x}_i - \mu_j)^T \, {\Sigma_j}^{-1} (\mathbf{x}_i - \mu_j)$$

where μ_j and $\mathbf{\Sigma}_j$ are the mean vector and covariance matrix respectively for facial feature j obtained from images in the training set. This procedure is repeated for all 4 classes of facial features, and the candidate is discarded if it does not belong to any of the 4 feature class.

The second stage, the attentive feature grouping stage, groups these detected feature candidates into face groups using our model knowledge of the face. Single features are grouped into vertical and horizontal pairs, pairs are grouped into partial face groups, and partial face groups are grouped into face candidates (fig. 4).

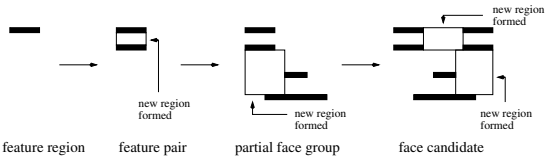


Figure 4. Attentive feature grouping.

The rules for grouping the facial components are divided into 2 groups. One group encodes geometric information such as length, orientation, inter-feature distance, etc., and the other group encodes spatial information about whether there should be edges of a particular strength and orientation at some spatial location in the feature region. These rules are represented by values in two separate vectors and the Mahalanobis distance of these two feature vectors are used to determine the component's membership in its class.

This grouping process is effective in removing false positives because a lot of geometric and spatial evidence are used, in particular the edge and spatial information about the new region formed in the component groups (see fig. 4). One important advantage of this process is that though the spatial region to be analyzed gets larger at higher levels, there are fewer of these regions to process, so the processing time is kept small throughout the whole algorithm.

3.3. Probabilistic Framework

Probabilities are assigned to each feature or face group and these probabilities are updated using Bayesian networks (or belief networks). Belief networks have nodes represent-

ing random variables and arcs signifying direct dependencies specified in terms of conditional probabilities (Sarkar and Boyer [13]). Each node can take either of 2 values, True or False, and has a conditional probability table (or CPT) describing the conditional probability of each value given each possible combination of the values of the parent nodes.

The entries in the CPT can be estimated directly by using the statistics of the set of examples (Russell *et.al.* [12]). The crucial value in the belief network is the prior probability of the "face" node, which is often hard to estimate. The choice of an appropriate prior depends on the complete space of hypothesis, and an uniform prior is assumed for our case.

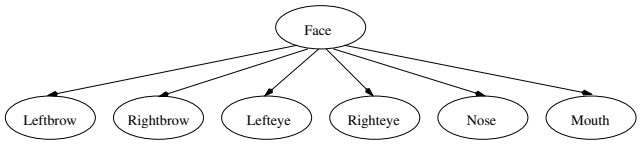


Figure 5. Belief network.

The belief network used is shown in fig. 5. The belief network has 6 child nodes for each of the 6 facial features. Fronto-parallel views will have 6 pieces of evidence while profile views will have 4. To fully exploit contextual evidence and model knowledge, a second belief network (fig. 6) is used to reinforce the belief of each feature based on the presence of neighbouring features.

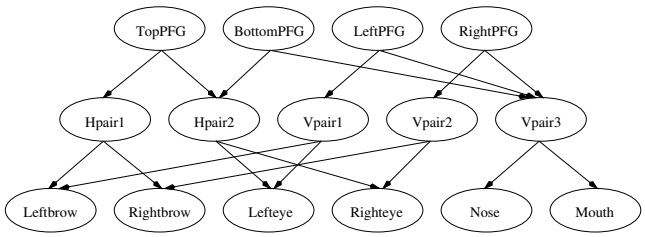


Figure 6. Reinforcement belief network.

A propagation algorithm for singly connected networks given by Pearl [10] is used to propagate the evidence through the reinforcement network. Each node when instantiated with a piece of evidence will modify its parent or child nodes based on the conditional probabilities between the nodes. These parent or child nodes will further modify their parent and child nodes, thus propagating the evidence throughout the network. The details of the evidence propagation is given in Yow and Cipolla [18].

The evidence for each facial feature or face group i is related to its Mahalanobis distance, $\mathcal{M}_{ij}$, and the admission threshold for the jth feature class, τ_j, by :

$$P_i = (1 - \frac{\mathcal{M}_{ij}}{\tau_j}),$$

Each facial feature that is detected is assigned 4 probability values, P_{brow}, P_{eye}, P_{nose} and P_{mouth}, using the above equation. When a higher level group is formed, only the probability of the corresponding feature is propagated. For example, if a vertical brow-eye pair (Vpair1) is formed, only P_{brow} of the upper facial feature and P_{eye} of the lower feature is propagated. Likewise, only these values are updated in the propagation process. As a result, only true positive faces are updated to a high confidence level.

4. Detection of faces under scale variations

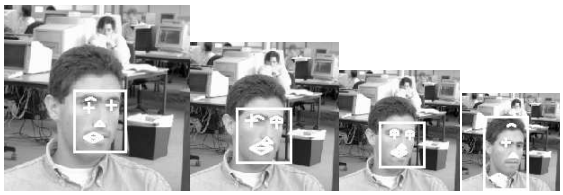
In this section, we will look at the effects of varying scale on the detection of faces. In our approach, two types of filters are used, the preattentive filter and the edge detection filter. The preattentive filter is constructed from a second derivative of Gaussian, and is elongated at an aspect ratio of 3:1 for better orientation selectivity. The edge detection filter is a standard Canny filter, which is a first derivative of Gaussian. Both filters are Gaussian derivative filters and we can thus examine the scale dependency (the σ of the Gaussian functional) in the same way.

(a) (b) (c) (d)

Figure 7. Varying the scale of the preattentive filter. The facial features detected by the preattentive feature selection stage is shown. (a) σ = 3.0 (81 points). (b) σ = 2.0 (177 points). (c) σ = 1.0 (332 points). (d) σ = 0.7 (408 points).

Fig. 7 shows the result of varying the scale of the preattentive filter from $\sigma = 3.0$ to $\sigma = 0.7$ while keeping the scale of the edge detection filter fixed at $\sigma = 3.0$. We observe that at a scale ($\sigma = 0.7$) smaller than the one required for matched filtering ($\sigma = 3.0$), the correct facial features are still detected, though there is a much larger number of false detected feature points. A large σ cause greater smoothing to take place and thus less feature points are detected, but the accuracy of localization of these points is lost.

We subsequently vary the size of image while keeping the scale of the preattentive filter constant at $\sigma = 1.0$. The scale of the edge detection filter is also reduced and kept at $\sigma = 1.0$. Fig. 8 shows the result for the image being reduced to 80%, 60%, 40% and 20% of the original size. We fail to detect the facial features when the face is too small because the image structure of these facial features are corrupted by quantiza-



(a) (b) (c) (d)

Figure 8. Varying the size of the face in the image. (a) percentage size = 80%. (b) percentage size = 60%. (c) percentage size = 40%. (d) percentage size = 20%.

tion noise. However, we are successful in detecting the facial features of large faces even though our preattentive filter is small. This is because the size of the facial features are actually determined by the edge detection and edge linking process, and not by the scale of the preattentive filter.

(a) (b) (c)

Figure 9. Varying the aspect ratio of the preattentive filter. (a) aspect ratio = 3:1 (81 points). (b) aspect ratio = 2:1 (110 points). (c) aspect ratio = 1:1 (201 points).

We do a further test by varying the aspect ratio of the preattentive filter. Fig. 9 shows the result of varying the aspect ratio from 3:1 to 1:1. We observe that the facial features are still detected even though we reduce the aspect ratio to 1:1. The significance of this is that we can steer a 1:1 second derivative Gaussian exactly by using only 3 basis filters (Freeman and Adelson [4]), instead of using 16 basis filters to give a 1% error approximation for a 3:1 filter (Perona [11]) - a huge saving in computational requirements.

The Gaussian functional minimizes the product of localization in space and frequency (Marr and Hildreth [9]), but its trade-off between the signal-to-noise ratio and the accuracy of localization is well studied (Canny [1]). Since the edge detection filter is a first derivative Gaussian, choosing a small σ will result in noisy edges that are difficult to link. A large σ, however, may blur the image features or even cause two separate edges to be smoothed into one. For an application of detecting the face of a person sitting in front of a computer terminal, σ = 1.0 is found to be sufficient.

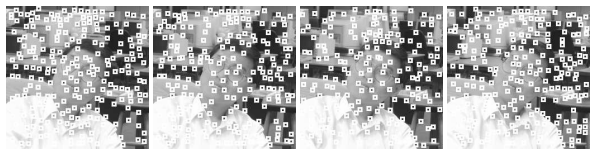
5. Detection of faces under orientation variations

We will now look at the effects of varying the orientation of the filter on the detection of faces. We use an image in which the subject's head is rotated approximately $30°$ to the right, i.e. at an orientation of $-30°$ from vertical. We keep the preattentive filter at the 3:1 aspect ratio and rotate the filter from $-60°$ to $30°$ in $30°$ increments.

(a) (b) (c) (d)

Figure 10. Varying the orientation of the preattentive filter. (a) orientation = -60° (99 points). (b) orientation = -30° (89 points). (c) orientation = 0° (90 points). (d) orientation = 30° (89 points).

The results in fig. 10 show that though the correct orientation is $-30°$, the facial features can still be detected by the filter at orientations of $-60°$ and $0°$. Thus, the algorithm can tolerate an orientation variation of about $30°$.



(a) (b) (c) (d)

Figure 11. Varying the orientation with aspect ratio = 1.1. (a) orientation = -60° (225 points). (b) orientation = -30° (205 points). (c) orientation = 0° (283 points). (d) orientation = 30° (223 points).

We again reduce the aspect ratio of the preattentive filter. Fig. 11 shows the result of varying the orientation at an aspect ratio of 1:1. We observe that the facial features are detected in all the different orientations of the filter. The significance of this is that we can make do without steerable filters completely. We can simply use only one single orientation of the preattentive filter and just examine the vicinity of the attention points for pairs of edges that are roughly parallel and have the correct polarity.

6. Detection of faces under viewpoint variations

In Yow and Cipolla [17] we have shown that the Gaussian derivative filter (the preattentive filter described in this paper) is able to detect facial features under different viewpoints, even under profile view.

Fig. 12 shows the features detected by the preattentive filter in profile views of faces. The scale of the preattentive filter used is σ = 3.0 and the aspect ratio is 1:1. We observe that all the facial features which can be seen in the image are detected by the preattentive filter.

(a) (b) (c)

Figure 12. Detecting features in profile views. (a) 195 points. (b) 174 points. (c) 185 points.

However, the difficulty in detecting faces under such viewpoints is that the facial features which we have chosen in our model can all be seen from only a limited range of viewpoints (mainly fronto-parallel). Thus, for general viewpoints (especially profile view), some of these features may be occluded and the evidential support becomes low.

To overcome this, we look for additional features when we have a face hypothesis at a different viewpoint (e.g. profile view). We observe that in a profile view, there is a large region of the cheek that is roughly featureless. Hence we can mark out additional regions in the image that are the likely location of the cheeks, and examine the number, strength and orientation of edges in it. Face candidates that are formed from a single partial face group (indicating a possible profile view face candidate) are examined for these additional regions. The same Class space - Mahalanobis distance method is used to verify the group of features and regions as a valid profile view of a face.

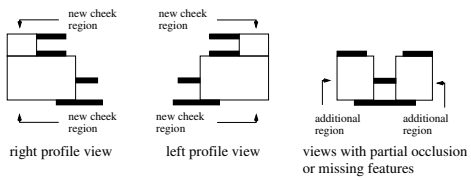


Figure 13. The additional cheek regions used under different viewpoints and when facial features are missing or occluded.

The same situation applies to views which some of the facial features are occluded. The different cheek regions that are used for the different views are shown in fig. 13.

7. Results

We implement the face detection algorithm as described in [19] but we use only one single scale and orientation of the filter for the preattentive feature selection process. The scale of the preattentive filter is chosen to be the same as the edge detection filter ($\sigma = 1.0$) so that we only need to smooth the image once. The intermediate results are shown in fig. 14.

(a) (b) (c) (d)

Figure 14. (a) Interest points (466 points). (b) Feature regions (180 points). (c) Partial Face Groups (28 top, 10 bottom, 5 left, 1 right). (d) Face candidates (2 faces).

Comparing the results of fig. 14 with that of fig. 7a, we observe a large increase in the number of feature points ($\frac{466}{81} \approx 5.75$ times increase in the number of points). However, the perceptual grouping process is able to reduce the number of face candidates down to only two (fig. 14d.).

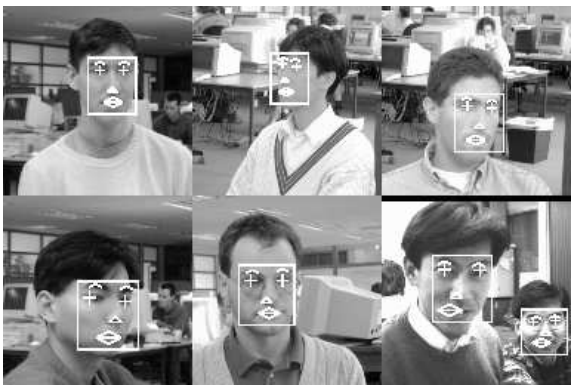


Figure 15. Result of face detection on various face images at different scales.

Our algorithm is implemented on a SUNSparc20 workstation. The images are taken from subjects sitting in front of a workstation mounted with a Pulnix monochrome CCD camera. The images used are 256x256 pixel resolution. The

time taken to process the images is about 90 seconds due to the large number of feature points detected.

We achieved a successful face detection rate of 85% on a database of 110 images of faces at different scale, orientation and viewpoint. Some of the results of testing the algorithm are shown in fig. 15. We observe that the algorithm is able to handle a good range of scale variations.

Figure 16. Result of face detection on face images at different orientation.

We also show the results of the algorithm when tested on images with different face orientation, as well as on profile views of the face. The results in figs. 16 and 17 show that the algorithm is able to cope well with variations in orientation and viewpoint.

Figure 17. Result of face detection on face images at profile views.

8. Conclusion

We have investigated the effects of varying the scale and orientation parameters of a feature-based face detection algorithm, and have proposed an extension of the algorithm to work under different imaging conditions. The algorithm is shown to be able to work well in detecting faces under different scale, orientation and viewpoint.

References

[1] J. Canny. A computational approach to edge detection. *IEEE Trans. Patt. Analy. and Machine Intell.*, 8(6):679–698, 1986.

[2] Q. Chen, H. Wu, and M. Yachida. Face detection by fuzzy pattern matching. In *Proc. 5th Int. Conf. on Comp. Vision*, pages 591–596, MIT, Boston, 1995.

[3] Y. Dai, Y. Nakano, and H. Miyao. Extraction of facial images from a complex background using SGLD matrices. In *Proc. Int. Conf. on Pattern Recognition*, volume A, pages 137–141, Jerusalem, 1994. IEEE Computer Society Press.

[4] W. T. Freeman and E. H. Adelson. The design and use of steerable filters. *IEEE Trans. Patt. Analy. and Machine Intell.*, 13(9):891–906, 1991.

[5] K. Koffka. *Principles of Gestalt Psychology*. Harcourt, Brace and World, 1935.

[6] W. Kohler. *Gestalt Psychology*. Liveright, New York, 1947.

[7] A. Lanitis, C. J. Taylor, and T. F. Cootes. An automatic face identification system using flexible appearance models. *Image and Vision Computing*, 13(5):393–401, 1995.

[8] T. Leung, M. Burl, and P. Perona. Finding faces in cluttered scenes using labelled random graph matching. In *Proc. 5th Int. Conf. on Comp. Vision*, pages 637–644, MIT, Boston, 1995.

[9] D. Marr and E. C. Hildreth. Theory of edge detection. In *Proc. Royal Soc. of London*, volume B207, pages 187–217, 1980.

[10] J. Pearl. *Probabilistic Reasoning in Intelligent Systems*. Morgan Kaufman, San Mateo, Calif., 1988.

[11] P. Perona. Steerable-scalable kernels for edge detection and junction analysis. In G. Sandini, editor, *Proc. 2nd European Conf. on Comp. Vision*, pages 3–18, Italy, 1992. Springer–Verlag.

[12] S. Russell, J. Binder, D. Koller, and K. Kanazawa. Local learning in probabilistic networks with hidden variables. In *Proc. Int. Joint Conf. on Artif. Intell.*, 1995.

[13] S. Sarkar and K. L. Boyer. Integration, inference, and management of spatial information using bayesian networks: Perceptual organization. *IEEE Trans. Patt. Analy. and Machine Intell.*, 15(3):256–273, 1993.

[14] K. K. Sung and T. Poggio. Example-based learning for view-based human face detection. Technical Report A.I. Memo 1521, CBLC Paper 112, MIT, Dec. 1994.

[15] A. Triesman. Perceptual grouping and attention in visual search for features and objects. *Journal of Experimental Psychology: Human Perception and Performance*, 8(2):194–214, 1982.

[16] G. Yang and T. S. Huang. Human face detection in a complex background. *Pattern Recognition*, 27(1):53–63, 1994.

[17] K. C. Yow and R. Cipolla. Finding initial estimates of human face location. In *Proc. 2nd Asian Conf. on Comp. Vision*, volume 3, pages 514–518, Singapore, 1995.

[18] K. C. Yow and R. Cipolla. Feature-based human face detection. Technical Report CUED/INFENG/TR249, University of Cambridge, August 1996.

[19] K. C. Yow and R. Cipolla. A probabilistic framework for perceptual grouping of features for human face detection. In *Proc. 2nd Int. Conf. on Auto. Face and Gesture Recog.*, Vermont, USA, 1996.