Human Face Detection with Neural Networks and the DIRECT Algorithm

Yau-Zen Chang   Kao-Ting Hung   Shih-Tseng Lee

Medical Augmented Reality Research and Development center
Department of Mechanical Engineering
Chang Gung University, Tao-Yuan 33302, Taiwan
zen@mail.cgu.edu.tw   kattin@st.cgu.edu.tw   joe1181o@adm.cgmh.org.tw

ABSTRACT

Based on Rowley's approach [2], *this paper aims at proposing a new architecture that uses a specific optimization technique, the DIRECT (DIviding RECTangle) algorithm, to improve the efficiency of face detection in images. The system consists of two main parts: a neural network-based face detection arbitrator and a search strategy based on an integer-handling DIRECT algorithm. By the architecture, the number of arbitration is dramatically reduced, and human faces, if they are present in an image, are not restricted to predetermined resolutions and aspect ratios. Experimental results show that the proposed architecture is efficient in terms of both speed and robustness.*

INDEX TERMS: face detection, neural network, image processing, Image pattern recognition, DIRECT algorithm

1. INTRODUCTION

Human face detection from complex background and in different positions of an image is an important step for surveillance and intelligent human-machine interface.

Rowley's approaches [2] [3] are typical image-based face detection methods in which upright and frontal views of human face are detected using artificial neural networks (ANN). Owing to its conceptual simplicity, the approach of [2] has been followed by many researchers, such as [4]-[6] which use ANNs. In our previous work [11], we adopted both the image and feature-based techniques to design two ANNs to estimate the position as well as planar orientation of faces.

However, the approach of [2] is restricted to a fixed training and classification template, which is of 20×20 pixels. And the original image must be re-sampled at several preset resolutions due to the so-called image pyramid arrangement. Furthermore, the ANN must check all the extracted windows in all the sampled images for the potential existence of human face. These restrictions can lead to prolonged processing time and overlook of faces that are difficult to identify in these resolutions and aspect ratios.

With these shortages of [2] in mind, this paper aims at proposing a new architecture that uses a specific optimization technique, the DIRECT (DIviding RECTangle) algorithm, to improve the efficiency of face detection in images.

2. THE INTEGER DIRECT ALGORITHM

We propose to apply the DIRECT algorithm [13] to improve the face detection architecture described in Section 2. The name of the optimization algorithm is an abbreviation of DIviding RECTangles, describing a specific manipulation procedure. Developed by Jones et al. in 1993, the procedure is deterministic, guaranteed to converge, and has a very fast convergence rate without trapping in local extremes [14]-[16].

In this work, the images are of finite resolution, and we need an integer version of the DIRECT algorithm. There are two parts to be considered: center definition and trisecting manipulation. The center point between two integers, say a to b, is defined as *floor*($(a+b)/2$), where output of the *floor* function is the greatest possible integer.

In order to divide boxes into 3 parts, the trisected widths are calculated according to the following equation:

$$\Delta = floor\left(\frac{(b-a+1)}{3}\right)$$

The above equation would lead to two specific conditions of Δ:

$$\text{If } \Delta \geq 1, \begin{cases} Child_L : [a, a+\Delta-1] \\ Child_C : [a+\Delta, b-\Delta] \\ Child_R : [b-\Delta+1, b] \end{cases}$$

$$\text{If } \Delta = 0, \begin{cases} Child_C = a \\ Child_R = b \end{cases}$$

where $Child_C$, $Child_L$ and $Child_R$ denote the center, left, and right decedents after participation, respectively. That is, when $\Delta \geqq 1$, the box will be divided into 3 parts: $Child_L$ ranges from a to $a+\Delta-1$ with width Δ, $Child_C$ ranges from $a+\Delta$ to $b-\Delta$ with width $b-a-2\Delta+1$, and $Child_R$ ranges from $b-\Delta+1$ to b with width Δ. When $\Delta = 0$, the box is divided into 2 parts: $Child_C$ and $Child_R$, each is of unit length.

3. PROPOSED FACE DETECTION METHOD

Our system consists of two main parts: a neural network-based face detection arbitrator and a search strategy based on an integer-handling DIRECT algorithm. As shown in Fig. 1, the position, size, and aspect ratio of the windows to crop sub-images for face detection are based on the DIRECT algorithm. After cropping and preprocessing, the feature is extracted at predetermined 49 locations. The feature values are then sent to the ANN to decide how close they represent a human face.

In other words, the face detection problem is converted into an optimization problem to find out the best position and size of the cropping window that contains a human face.

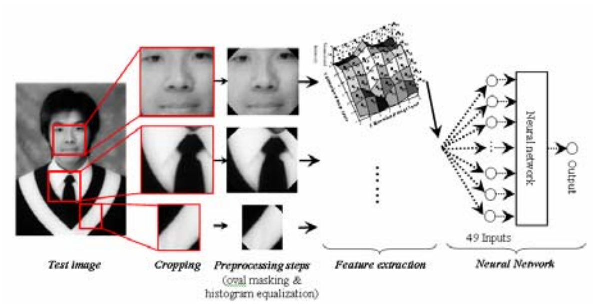


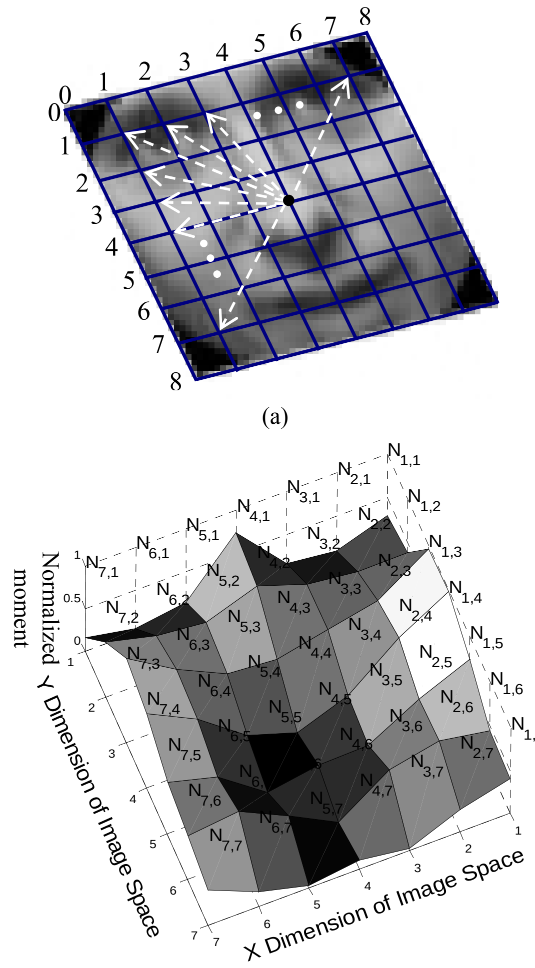
Fig. 1 The proposed face detection architecture

4.1. Neural network-based face arbitrator

The training and testing images for ANN in [2] are based on a fixed window (20×20 pixels). This approach not only restricts its applicability to face images with different aspect ratios, but also requires the computationally intensive resizing operation.

Based on these observations, we propose a moment-based feature extraction strategy that can alleviate the restrictions on aspect ratio and resolution.

The strategy takes the symmetrical property of human face into consideration. As Fig. 2 (a) illustrates, we segment the face image into 8 by 8 grids, this would result in 64 nodes from $N_{0,0}$ to $N_{8,8}$. Note that only the intensities at the node are used, thus no resize operation is required, and the template images are not required to be square.



(a)

(b)

Fig. 2 (a) Location of the 49 nodes. (b) Distribution of these 49 normalized values.

Based on the first moment values of these nodes, 49 quantities are obtained:

$$M_{i,j} = I(p_{i,j}) \times \| p_{i,j} - p_C \|, i = 1 \sim 7, j = 1 \sim 7$$

where $p_{i,j}$ denotes the position of node $N_{i,j}$, p_C are the position of central node of this grid map, i.e., $N_{4,4}$, and $I(p_{i,j})$ represents the intensity of node $N_{i,j}$. Fig. 9 (b) shows distribution of these 49 normalized values corresponding to the example image. Obviously, the central moment would be smallest, and the distribution has a symmetrical characteristic along horizontal axis.

The ANN uses these 49 values as inputs to judge for the existence of human face. There are two output neurons, and the ANN is trained such that the output is [1, 0] when a human face is present, and [0, 1] when none.

There are at least four variables to be decided in the face detection problem: the X and Y coordinates, x and y, width, w, and height, h. These variables are summarized in to an unknown vector:

$$\theta = \{x, y, w, h\}.$$

The ranges of these variables are defined within the following boundaries:

$$1 \le x \le \text{width of the image}$$

$$1 \le y \le \text{height of the image}.$$

$$24 \le w \le 90$$

$$24 \le h \le 90$$

The problem is then formulated as finding the parameter vector, θ, by the DIRECT algorithm, such that the corresponding output of the ANN approximates the output of the arbitrator when a human face is present, which is the value [1, 0] in our current design. Fig. 3 shows the flow chart of the face detection architecture using the integer-handling DIRECT algorithm

In general, DIRECT algorithm proposes a searching strategy to obtain all possible solution. We extract all solution according to a specific threshold, i.e. if the values of objective function less than the predefined threshold, the corresponding solution would be considered.

5. EXPERIMENTAL RESULTS

We randomly selected 340 frontal-view face images and 306 non-face images from the MIT-CBCL face recognition database, CMU face database, and our own database to train the ANN. Besides, we prepare 138 images contains no face in order to improve the robustness of the ANN using the so-called "bootstrap" method [12].

The network is of the two-layered Multi-layered Perceptron (MLP) architecture with 20 neurons in the first hidden layer and 15 neurons in the second hidden layer. A standard back-propagation algorithm [17] was used to train this 49-20-10-2 network for 500 cycles. Fig. 4 shows the positions in the search space visited by the integer-handling DIRECT algorithm. Note that, for a 3-dimensional illustration, only the X and Y coordinates, x and y, and width, w are shown.

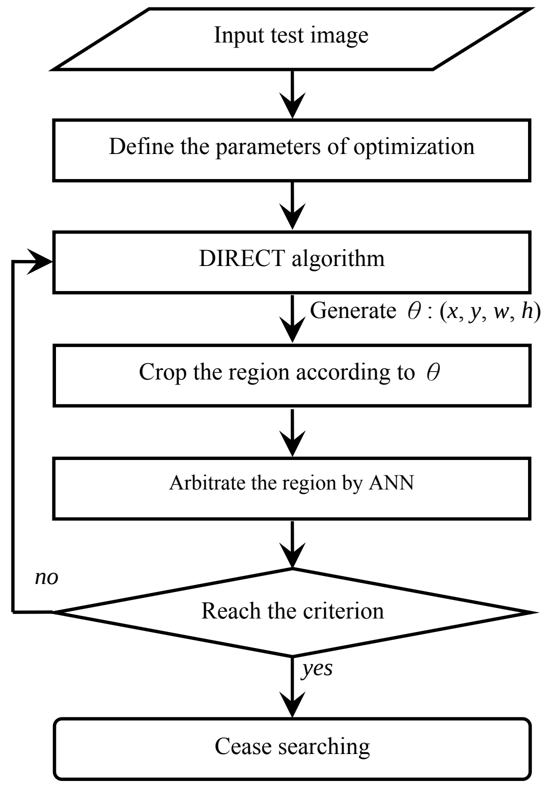


Fig. 3 Flow chart of the face detection architecture using the integer-handling DIRECT algorithm

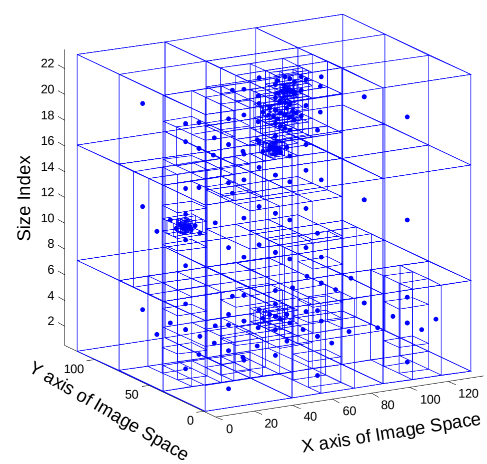


Fig. 4 Positions in the search space visited by the integer-handling DIRECT algorithm

Based on 50 detection results, the system based on the DIRECT algorithm can efficiently locate the human face within 1000 arbitrations using the ANN. As a comparison, when a typical 600 by 400 image is encountered, the number of arbitration in the approach of [2] can easily exceed the amount of 300,000.

Also found in the experiments, the DIRECT algorithm could cause repetitive detection. This problem can be easily solved by integrating all the neighboring positions into one by, for instance, the fuzzy c-mean method. Furthermore, we can integrate other image processing techniques, such as the skin color segmentation, to reduce the search area to further reduce the computing effort.

6. CONCLUSION

Based on the architecture of [2], we proposed a new human face detection architecture using the integer-handling DIRECT algorithm. The newly designed face classifier is computationally efficient and invariant to image size. Furthermore, the architecture is robust in terms of correctness, since the aspect ratio of face images is not faxed.

As for future work, we will address the issues of rotated faces and side-view faces. Skin color segmentation may also be included to provide further acceleration.

ACKNOWLEDGEMENT

This study is supported by Technology Development Program for Academia (TDPA) of the Ministry of Economic Affairs, R.O.C., under contract no. 94-EC-17-A-19-S1-035.

REFERENCES

[1] E. Hjelmås and B. K. Low, "Face Detection: A Survey," *Computer Vision and Image Understanding*, Vol. 83, pp. 236-274, 2001.

[2] H. A. Rowley, S. Baluja and T. Kanade, "Neural Network-based Face Detection," *IEEE Transaction Pattern Analysis and Machine Intelligence*, Vol. 20, pp. 23-38, January, 1998.

[3] H. A. Rowley, S. Baluja and T. Kanade, "Rotation Invariant Neural Network-based Face Detection," *Proceedings of Computer Vision and Pattern Recognition*, pp. 38-44, 1998.

[4] C. Garcia, M. Delakis, "Convolutional Face Finder: A Neural Architecture for Fast and Robust Face Detection", *IEEE Transaction Pattern Analysis and Machine Intelligence*, Vol. 26, No. 11, pp. 1408-1423, 2004.

[5] K. Curran, X. Li, and N. M. Caughley, "The Use of Neural Networks in Real-time Face Detection," *Journal of Computer Science*, Vol. 1(1), pp. 47-62, 2005.

[6] P. Hartono, S. Hashimoto, "A Robust Face Detector Algorithm Utilizing Neural Networks and Partial Template Matching," *Proceedings of the SPIE*, Vol. 5603, pp. 119-127, 2004.

[7] S. Birchfield, "Elliptical head tracking using intensity gradients and color histograms," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp.232-237, 1998.

[8] C.C. Han, H.Y. Liao, G.J. Yu, and L.H. Chen, "Fast face detection via morphology-based pre-processing," *Pattern Recognition*, Vol.33, pp. 1701-1712, 2000.

[9] C.H. Lin and K.C. Fan, "Triangle -based approach to the detection of human face," *Pattern Recognition*, Vol.34, pp.1271-1284, 2001.

[10] R.L. Hsu, "Face detection in color images," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol.24, pp.696-706, 2002.

[11] Y. Z. Chang and G. T. Hung, "Detection of Face Pose with Neural Networks," *International Symposium on Artificial Life and Robotics*, B-Con Plaza, Beppu, Oita, Japan, January 23-25, 2006.

[12] K. K. Sung, "Learning and Example Selection for Object and Pattern Detection," PhD Thesis, *MIT AI Lab*, January 1996.

[13] D. R. Jones, C. D. Perttunen, and B. E. Stuckman, "Lipschitzian optimization without the Lipschitz constant," *Journal of Optimization Theory and Application*, Vol. 79(1), pp. 157-181, 1993.

[14] C. A. Floudas, and P. M. Pardalos, *Encyclopedia of Optimization*, Kluwer Academic Publishers, London, 2001.

[15] H. Zhu and D. B. Bogy, "DIRECT Algorithm, and Its Application to Slider Air-Bearing Surface Optimization," *IEEE Transactions on Magnetics*, Vol. 38, No. 5, September, 2002.

[16] H. Zhu and D. B. Bogy, "Hard disc drive air bearing design: modified DIRECT algorithm and its application to slider air bearing surface optimization," *Tribology International*, Vol. 37, pp. 193-201, 2004.

[17] F. M. Ham, I. Kostanic, *Principles of Neurocomputing for Science & Engineering*, McGraw-Hill, 2001.