DETECTING HUMAN FACE FROM MONOCULAR IMAGE SEQUENCES BY GENETIC ALGORITHMS

Hertog Nugroho†* *Shigeyoshi Takahashi*‡ *Yoshiteru Ooi*‡ *Shinji Ozawa*†

†Department of Electrical Engineering, Faculty of Science and Engineering
Keio University, Yokohama Kanagawa 223, JAPAN
{ hertog, ozawa } @ozawa.elec.keio.ac.jp
‡Advanced Research Center, INES Corporation, Yokohama Kanagawa 224, JAPAN
{ taka, ooi } @ines.co.jp

ABSTRACT

This work investigates a new approach to detect human face from monocular image sequences. Our method consists of two main search procedures, both using Genetic Algorithms. The first one is to find a head inside the scene and the second one is to identify the existence of face within the extracted head area. For this purpose, we have developed two models to be used as a tool to calculate the fitness for each observation in the search procedure: One is a head model which is approached by an ellipse and the other one is a face template which size is adjustable. The procedures work sequentially. The head search is activated first, and after the head area is found, the face identification is activated. The experiment demonstrates the effectiveness of the method.

1. INTRODUCTION

Human face recognition has attracted many researches because of its potential application in areas such as criminal identification, surveillance, security, etc. In real world, the task of face recognition includes a task to detect and locate a face(s) in a complex environment such as in the busy corridor, in the office, in front of entrance gates, etc. Thus, to realize a fully automatic system in human face recognition, the first task is to find and locate a face and then, after the face has been located, the identification is performed.

So far, most researches concentrated only on identification task assuming that the face has been already well defined or using a simple background so as to make picking the face would not take much effort [1].

Recently, the first problem had been addressed. However, since both the location and the size of a face in a picture are unknown, the problem is very difficult. So far, two approaches has been proposed: One using model-template matching [2], and the other one using coarse-fine analysis (finding the head or face and then analyzing the details of facial features) [3], [4].

In this paper, we attempt to develop an automatic face location system by using the first approach and exploiting head/face color information. The distinctive feature of our approach is the use of Genetic Algorithms which has been well known for its robustness from noise and always converge to its global optimal value. We assume that the camera is fixed-installed in a strategic location to cover a wide area such as the above of entrance gate, the upper side of the wall's corner, etc. For the case of a moving person, by applying image subtraction method, we can isolate the overall posture area of the person and then concentrate on that person to find his head. Finding the head is done by using a simple ellipse model and then activated the search procedure based on Genetic Algorithms. Then, the histogram of the extracted head is analyzed. The main property of human head (including face) which we exploited in the analysis is that of the color of human hair. In grey scale image, the human hair is usually darker than the skin of his face. Also some salient features of human face such as eyes, nose and mouth look darker than the skin. By exploiting this property, we applied thresholding technique with automatic threshold to isolate the dark portion and then analyzing it. In case of frontal view face, we activated the second search procedure (the first one is the search for a head), also based on Genetic Algorithms, to locate eyes and mouth.

2. EXTRACTION OF THE HEAD OF MOVING PERSON

The first process in finding the head is the detection of moving object. We used an image subtraction method which is simple and yet powerful enough to discriminate between moving objects and another non-moving ones. Three frames are required for the computation. The main reason for using more than 2 frames is both to improve the estimation confidence of the object motion and to effectively exclude the occlusions caused by object motion.

First, the sensing system inputs a series of image I from a camera sensor. Let $S(t_n-\Delta t)$, $S(t_n)$, and $S(t_n+\Delta t)$ be edge images extracted by Sobel operation from $I(t_n-\Delta t)$, $I(t_n)$, and $I(t_n+\Delta t)$, respectively. Δt is the temporal difference in the images. Then, the difference image D_- can be obtained by subtracting two successive edge images as follows:

$$D_- = pos\big(S(t) - S(t-\Delta t)\big) \tag{1}$$

where

$$pos(x) = \begin{cases} x & \text{for } x > 0 \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

With the same procedure, the difference image D_+ can be obtained by subtracting $S(t)$ with $S(t+\Delta t)$. Please notice

*The author is supported by Hitachi Scholarship Foundation.

that the subtrahend in both procedures is similar, i.e: $S(t)$. It is necessary when we need to get moving object portion in current state. By multiplying D_- and D_+ and thresholding, the edges of moving person at current frame Ω_t is obtained, while the remaining background noise is greatly reduced. In our implementation, we introduced a blurring process between edge extraction and image subtraction. The idea is inspired by that proposed in [5], while, instead of exploring the edge images, they used the ramp of edges. This has an advantage of being robust against nearby noise while retaining the salient edges.

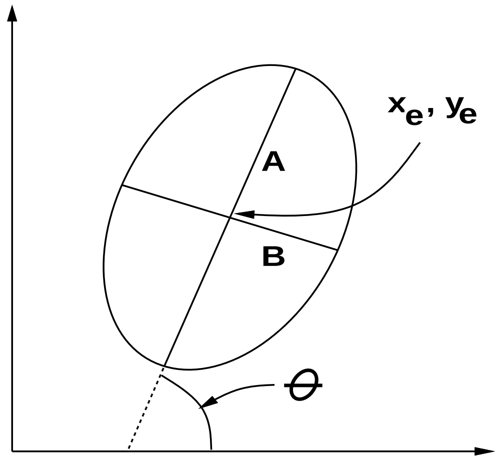

Figure 1. Head Model

To find a head of this person, we used an ellipse model with 5 parameters: two parameters x_e and y_e for the center of the ellipse, two principal axis A and B, and orientation θ as was shown in Fig.1. These 5 ellipse parameters are converted into 1 string member of population whose the cost function will be evaluated by the algorithms. By using these 5 parameters and concentrating on the upper-half of the extracted moving person, the head search algorithms is activated. For a lack of space, details of the algorithms is left out in this paper. We emphasize here that the basic functions in Genetic Algorithms, i.e: reproduction, crossover, and mutation are utilized. (For those readers who are interested in the algorithms, please refer to [6].) The cost function Υ for this search is as follows:

$$\Upsilon = \Big(\sum_{\Omega_t} h(x,y)\Big)/l_e \tag{3}$$

where $(x,y) \in \Psi(x_e, y_e, A, B, \theta)$ and

$$\Psi(x_e, y_e, A, B, \theta) = \tag{4}$$
$$\frac{((x-x_e)\cos\theta + (y-y_e)\sin\theta)^2}{A^2} + \frac{(-(x-x_e)\sin\theta + (y-y_e)\cos\theta)^2}{B^2} = 1$$

and l_e is the contour length of the ellipse.

To ensure convergency, it is necessary that information under evaluation should have a ramp-like characteristic around its salient features. To fulfill this requirement, before the search is activated, the image Ω_t is blurred.

The process is halted when the cost function gets its optimum value. In terms of genetic algorithms, it means that from all string population under evaluation, we have at least several members which have the same highest cost function value or differ only a little fraction from the highest cost function value.

(a) (b)

Figure 2. (a) Edge image Ω_t and (b) the extracted head area.

Figure 2 shows the edge image Ω_t and the search result represented by ellipse which is superimposed on source image. From the edge image we can see noise around the head and some other location. But the procedure has demonstrated that even with this ill-condition, it can still accurately locate the head and estimate its size which proves our claim in the previous section.

3. HEAD ANALYSIS

After the assumed head area $\Phi(x,y)$ has been extracted, we examined the histogram of all pixels within the area. We expect 3 conditions: it is mostly dominated by dark pixels, it consists of a rather balanced dark and light pixels, and it is mostly dominated by light pixels. In the first condition, we didn't expect to find a face, and hence the procedure stops here. In the second and third condition, we can expect a face, and hence we apply a face search procedure concentrating on this area. Again, we used Genetic Algorithms to perform this function.

To isolate dark pixels from the light ones, we used thresholding technique proposed by Otsu[7], and each pixel in $\Phi(x,y)$ is labeled as 1 if its value is less than th_1 (determined automatically) and 0 otherwise. To check the above mentioned condition, at first we count pixels in $\Phi(x,y)$ whose labeled value are 1 and compared the result to the whole number of pixels in $\Phi(x,y)$. If the ratio is less than th_2 (somewhere between 0.45 and 0.65) we can expect a face, otherwise we stop.

Assume we have a ratio less than th_2. Concentrating on $\Phi(x,y)$, we try to decide whether we have a face or not. For this purpose, we have developed a simple face model as was shown in fig.3. Our model is based mostly on 3 salient face features: left and right eyes and mouth. From these features, we defined 10 parameters: displacement from center of template x_τ, y_τ with respect to center of ellipse x_e, y_e, distance between eyes(eew), distance between eyes axis and mouth(em), eyes and mouth widths $w_k, k = 1,2,3$, and heights h_k, all normalized with respect

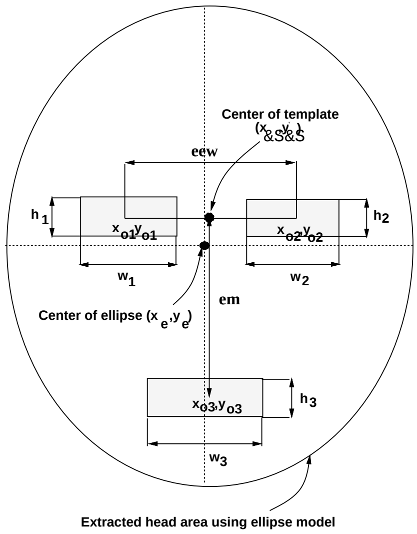

Figure 3. Template used to search eyes and mouth

to ellipse size. These 10 parameters form 1 string member of the population whose the cost function will be evaluated by the algorithm. The cost function for template Γ is as follows:

$$\Gamma = a_1 \times (\sum_{k=1}^{3} e_k) - a_2 \times s_e - a_3 \times s_m \quad (5)$$

where

$$e_k = f(p_k/A_k), \quad A_k = w_k \times h_k \quad (6)$$

$$p_k = \sum_{y=y_{ok}-h_k/2}^{y_{ok}+h_k/2} \sum_{x=x_{ok}-w_k/2}^{x_{ok}+w_k/2} \phi(x,y) \quad (7)$$
$$\phi(x,y) \in \Phi(x,y)$$

$$S_e = \sum_{y_k=y_{ok}-h_k/2}^{y_{ok}+h_k/2} \sum_{x_k=x_{ok}-w_k/2}^{x_{ok}+w_k/2} \quad (8)$$
$$\mid \phi(2x_{o1} - x_1, y_1) - \phi(x_2, y_2) \mid$$
$$k = 1, 2$$

$$S_m = \sum_{y=y_{o3}-h_3/2}^{y_{o3}+h_3/2} \sum_{x=x_{o3}}^{x_{o3}+w_3/2} \mid \phi(2x_{o3} - x, y) - \phi(x,y) \mid (9)$$

We used equal sized template for the eyes template and for function $f(p_k/A_k)$ in Eq.6, we used a simple function as was shown in fig.4, with maximum δ value decided by experiment. This function represents our confidence about finding a facial feature "inside" the template. In this case, we count a portion of a number of pixels labeled 1 within the predetermined area A_k with respect to the area. A zero portion means that the feature isn't found. A maximum 1 portion doesn't mean anything, because it could be a part of hair. So, if we want to catch a facial feature, the optimal portion should be between the extreme values (represented by δ). We found from the experiment that the value is between 0.75 and 0.85. The second and the third terms in Eq. 5 (S_e and S_m) measure the symmetry between left and right eyes and the shape of mouth, respectively. Zero value of these terms gives a high probability that the feature is found, and, hence, the overall fitness value is increased.

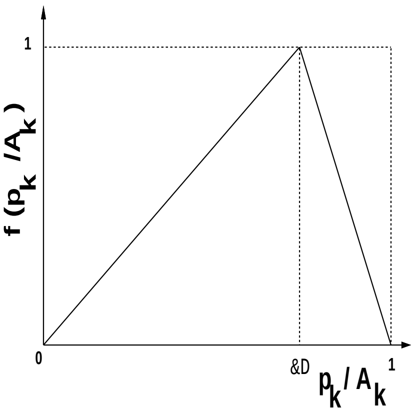

Figure 4. Function used in cost calculation.

Similar with the search procedure mentioned in the previous section, this process is also halted when the cost function gets its maximum value.

4. EXPERIMENT

The experiment was conducted indoor. A person was required to walk arbitrarily within the scope of camera view and his face wasn't necessarily facing the camera. The scene was recorded into a video format and then converted into 256 × 240 sized frames. The detailed result of one frame during the process is illustrated in Figure 2 and 5. Figure 2 illustrates the head area which had been found from edge image by the head search, while Figure 5 illustrates the extracted dark pixels region in $\Phi(x,y)$ as a result of automatic thresholding and the identification result.

(a) (b)

Figure 5. (a) The extracted head region $\Phi(x,y)$ and (b) the detected facial features.

Figure 6 reflects a blow version of the extracted head area (magnified by interpolation). It demonstrates that the procedure can correctly locate the salient facial features. When the eyes can accurately be located, the region catched by

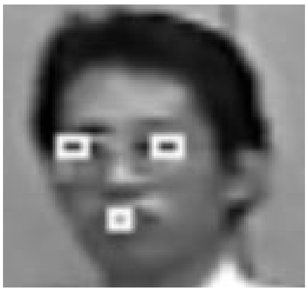

Figure 6. The magnified face area.

the lower template still needs further investigation whether it represents mouth or nose. But the ultimate aim, which is finding a face, is accomplished.

In the near future, we plan to add zooming function in term of hardware installation (by using zooming camera) and compare the quality of the result with that obtained by the image processing. This is necessary when the task of face identification needs to be addressed.

5. CONCLUSION

This paper has investigated a new approach to the issue of detecting a human face from a sequence of monocular images. We have presented a new model-based searching method based on Genetic Algorithms which detects a head and then, identify the face. Preliminary experiment demonstrates the effectiveness of the detector. In the near future, we plan to add the detector with zooming capability.

REFERENCES

[1] Osamu Nakamura, Shailendra Mathur, Toshi Minami, *"Identification of Human Faces Based on Isodensity Map"*, Pattern Recognition Vol 24, No. 3, 1991, pp 263-272.

[2] Akitoshi Tsukamoto, Chil-Woo Lee, and Saburo T-suji, *"Detection and Pose Estimation of Human Face with Multiple Model Images"*, IEICE Trans. Information and Systems Vol. E77-D. No. 11 November 1994.

[3] Guangzheng Yang and Thomas S. Huang, *"Human Face Detection in a Complex Background"*, Pattern Recognition Vol 27, No. 1, 1994, pp 53-63.

[4] Liyanage C. De Silva, Kiyoharu Aizawa and Mitsutoshi Hatori, *"Detection and Tracking of Facial Features by Using Edge Pixel Counting and Deformable Circular Template Matching"*, IEICE Trans. Information and systems Vol. E78-D. No. 9 September 1995.

[5] Marie-Pierre Dubuisson and Anil K. Jain, *"Contour Extraction of Moving Objects in Complex Outdoor Scenes"*, Int. Journal of Computer Vision, 14, 1995, pp 83-105.

[6] David E. Goldberg, *"Genetic Algorithms in Search, Optimization, and Machine Learning"* , Addison-Wesley Pub. Co., 1989.

[7] N. Otsu, *"A threshold selection method from gray-level histogram"*, IEEE Trans. SMC, SMC-8, 1978, pp. 62-66.

[8] Demas Sanger, *"Automatic Extraction of Face from Color Reversal Film Using Statistical Multistep Filtering Technique"*, ACCV,95 Second Asian Conference on Computer Vision, Singapore, Vol II pp 465-468.

[9] Qian Chen, Haiyuan Wu, and Masahiko Yachida, *"Real-time Face Detection"*, ACCV,95 Second Asian Conference on Computer Vision, Singapore, Vol II pp 479-483.