

Short Paper

Modified Hough Transforms for Object Feature Extraction*

JAR-FERR YANG AND SHU-SHENG HAO⁺

*Department of Electrical Engineering
National Cheng Kung University
Tainan, Taiwan 701, R.O.C.*

E-mail: jfyang@ee.ncku.edu.tw

⁺E-mail: n2883125@ccmail.ncku.edu.tw

In this paper, we propose the use of modified Hough transforms to efficiently extract object feature parameters, which are usually contaminated by heavily noisy corugation and discontinuity. The modified HT (MHT) is developed by introducing spatial and parameter weighting functions to improve the detection performance for the traditional Hough transform (HT), which generally fails to robustly detect natural object parameters. Using designed test patterns and real images, simulations show that the proposed weighting functions are helpful in detecting noise-corrupted object features. Due to its robustness, the MHT can be easily figured with a coarse-to-fine adaptive search mechanism to reduce the huge amount of computation for feature parameters extraction.

Keywords: modified Hough transform, model-based coding, feature parameters extraction, coarse-to-fine search, facial object estimation

1. INTRODUCTION

Model-based video coding has been recognized as one of potential techniques for achieving video communication below 10 kbps. With few parameters to describe the features of the object, model-based video coding is different from conventional coding schemes, which need to compress the image pixels by removing statistical or spatial redundancies. However, precise estimation of feature parameters with very little computation represents success for model-based approaches. Once the feature parameters are correctly extracted, they can easily be used to characterize the video objects. Thus, parameterized video object planes have been adopted as the visual presentation in MPEG-4 standardization [1]. In order to extract feature parameters, many algorithms, such as the template matched method [2], active contour method [3], mosaic method [4], and Hough transform methods [5, 6], have been proposed recently. Many facial features can gener-

Received July 2, 1999; revised November 16, 1999; accepted December 30, 1999.

Communicated by Mark H. Y. Liao.

* This research was supported by National Science Council under Contract #NSC-85-2221-E-006-046, Taiwan, R.O.C.

ally be described by analytic curves generally. For example, a circle can specify the pupil, and parabolas or ellipses can also characterize the mouth [7, 8]. The Hough transform (HT) method, which has been used widely in many applications [9-12], is an efficient method for detecting analytic curves. Since they use statistic models, the HT algorithms can be categorized as probabilistic and non-probabilistic methods [13]. Since the HT approach requires heavy computation and memory occupation, many modified methods in adaptive, randomized, and fast realizations have been proposed [14-19]. As for natural features, the HT approach without any modification seldom succeeds in detecting natural shapes. In this paper, we propose several modified HT (MHT) algorithms which can improve the performance in detecting facial features. Furthermore, the MHT can easily add a coarse-to-fine search to reduce computation and memory. Simulation results demonstrate that the suggested MHT algorithms achieve satisfactory results while the traditional HT fails to correctly solve the problems studied here.

2. STANDARD HOUGH TRANSFORM (HT)

For a gray image, we can use $d_i(x, y)$ to represent the intensity function at the (x, y) position. Before the Hough transform, is executed, $d_i(x, y)$ should be pre-processed and binarizes as

$$d_p(x, y) = T\{d_i(x, y)\} \quad (1)$$

and

$$d_b(x, y) = B\{d_p(x, y)\}, \quad (2)$$

respectively. In (1) and (2), $T\{ \}$ and $B\{ \}$ denote the operators of pre-processing transform and binarized functions, respectively. The pre-processing, $T\{ \}$, may include edge enhancement and thinning processes. The binarized operator, $B\{ \}$, is usually expressed by

$$B\{d_p(x, y)\} = \begin{cases} 1, & \text{if } d_p(x, y) > \delta \\ 0, & \text{if } d_p(x, y) \leq \delta \end{cases}, \quad (3)$$

where δ is the threshold in binarization. The proper selections for the pre-processing transform, $T\{ \}$, and the threshold value δ are usually image-dependent. Several methods have been proposed for effectively retrieving the threshold value [17-19].

The HT translates the object-enhanced pixels, $d_b(x, y)$, which represent the shape of the original image, into show-up counts in the parameter space by fitting an assumed analytic curve equation. A given analytic curve function can be characterized by

$$F(x, y, \mathbf{a}) = 0, \quad (4)$$

where \mathbf{a} is a vector composed of elements in the parameter space. The parameter space is an N -dimensional accumulation space if \mathbf{a} is composed of N parameters. To describe ana-

analytic curves, for example, the circles can be given by

$$(x-u)^2 + (y-v)^2 = r^2, \quad (5)$$

where r and (u, v) denote the radius and the center of the circle in the image space, respectively. Thus, the vector \mathbf{a} composed of (u, v) and r forms a 3-dimensional parameter space which characterizes the circle. A parabola curve can be described as $(y-v) = (x-u)^2 / \lambda_x$, $\theta = \tan^{-1}(v/u)$. Four coefficients, which are (u, v) , λ_x , and θ , can be used to describe an arbitrary parabola curve. Thus, the vector \mathbf{a} composed of (u, v) , λ_x , and θ forms a 4-dimension parameter space. If we divide the possible parameter region into many small cells, the HT performs mapping from edge pixels to parameter cells. For example, the resolution of circles in the parameter space can be defined as $\Delta u = u_{i+1} - u_i$, $\Delta v = v_{j+1} - v_j$, and $\Delta r = r_{k+1} - r_k$, where u_i , v_j , and r_k denote the parameters of the (i, j, k) th cell and Δu , Δv , and Δr represent the resolution of cells in u , v and r axes, respectively. For the highest resolution, we may choose $\Delta u = \Delta v = \Delta r = 1$. Considering the memory space and computation time, we may choose values of Δu , Δv , and Δr that are greater than one. However, coarse parameters usually can not be used to achieve precise detection of features. In the image space, of course, we may also apply the same concept by down sampling the image as $\Delta x = x_{i+k} - x_i$ and $\Delta y = y_{j+k} - y_j$, where x_i and y_j represent the pixel location in the image space, and $k \geq 1$ is the down-sampled step. However, the number of edge pixels will be approximately reduced by a factor of k^2 . Hence, the HT method may not obtain sufficient information for detecting the peak of counts. In summary, reducing the resolution in the parameter space or down sampling pixels in the image space can reduce the amount of computation required. However, the accuracy of parameter estimation for the HT will be correspondingly degraded.

For each pixel, we can substitute its location information, x and y , into (4) to obtain a hypercurve depicted by $\mathfrak{S}(\mathbf{a}) = 0$. Once the parameter hypercurve, $\mathfrak{S}(\mathbf{a}) = 0$, passes through the cells, the HT will increase their corresponding accumulators by one. After applying all detected $\mathbf{d}_B(x, y)$, the HT then employs show-up accumulation to extract analytic curves. In a mathematical expression, for a given parameter vector \mathbf{a} , the Hough transform of $\mathbf{d}_B(x, y)$ is expressed by

$$C_S(\mathbf{a}) = H_S \{ \mathbf{d}_B \mid F(x, y, \mathbf{a}) = 0 \} = \sum_{F(x, y, \mathbf{a})=0} \mathbf{d}_B(x, y) \quad (6)$$

Then, the most-likely set of feature parameters can be detected by searching the cell, which is with the maximum accumulated value

$$\mathbf{a}_{opt} = \max_a \{ C_S(\mathbf{a}) \}. \quad (7)$$

For the second possible set of parameters, we can choose the second maximum accumulated count, and so on for the other sets.

3. MODIFIED HOUGH TRANSFORM (MHT)

Observing the human facial features, most of the shapes of the features do not show complete analytic curves. For example, the pupil could possibly have some corrugation around it and be covered by eyelids in the upper or lower parts. Usually, noisy will cause the detection performance achieved by the HT to deteriorate greatly. In order to take the neighborhood pixels around the designated pixel into account, the first modification of the HT is suggested as

$$C_M(\mathbf{a}) = H_\varepsilon \{d_B(x, y) | F(x, y, \mathbf{a}) \leq \varepsilon\} = \sum_{F(x, y, \mathbf{a}) \leq \varepsilon} d_B(x, y), \quad (8)$$

where ε characterizes the size of the neighborhood which confines the calculating range of the pixels. With (8), we can achieve robust detection, which allows deviation of pixels irritated by noise. Thus, we can not only reduce the computation required in the thinning process for our modified HT, but also improve the detection performance. Hence, the first modified HT can be expressed by

$$C_B(\mathbf{a}) = H_\varepsilon \{d_E(x, y) | F(x, y, \mathbf{a}) \leq \varepsilon\} = \sum_{F(x, y, \mathbf{a}) \leq \varepsilon} d_E(x, y), \quad (9)$$

where

$$d_E(x, y) = d_B(x, y)w_\varepsilon(x, y, \mathbf{a}) \quad (10)$$

denotes that the edge pixels are weighted. If we choose $w(x, y, \mathbf{a}) = 1$, then we have $C_B(\mathbf{a}) = C_M(\mathbf{a})$. If we linearly give more weight to the pixels which are closer to the feature, then the linear weighting function becomes

$$w_\varepsilon(x, y, \mathbf{a}) = m \cdot |r - r_c| + w_0, \quad (11)$$

where $m = \pm w_0 / b\Delta r$ denotes the slope and r_c is the desired radius relative to the same circle center. Hence, w_0 is the highest weighting factor, Δr is the resolution and b is a constant. In order to reduce the required computation, we can consider a step-weighting function, for example,

$$w(x, y, \mathbf{a}) = \begin{cases} w_1 & |\mathbf{a} - \mathbf{a}_c| \leq a_1 \\ w_2 & a_1 < |\mathbf{a} - \mathbf{a}_c| \leq a_2, \\ w_3 & |\mathbf{a} - \mathbf{a}_c| > a_2 \end{cases} \quad (12)$$

where a_1 , a_2 , and a_3 are the thresholds and \mathbf{a}_c is the desired center parameter vector.

Since the binarized image may still contain long curvature curves, the HT generally has great tendency to detect them as the target curves. The longer curves have more edge pixels. The HT based on the maximum show-up count definitely fails to detect some

natural objects, such as the pupil, which is more like a circle but shorter in length than the eyelid. In order to clearly identify the circle and avoid the long-length arches, we further propose a second modification for the HT:

$$C_L(\mathbf{a}) = \rho_1(\mathbf{a}) H_\varepsilon \{d_B(x, y) / F(x, y, \mathbf{a}) = 0\} = \rho_1(\mathbf{a}) \sum_{F(x, y, \mathbf{a})=0} d_B(x, y), \quad (13)$$

where $\rho_1(\mathbf{a})$ is the parameter weighting function. To avoid long arches, for example, we may choose an inverse proportion to the radius parameter:

$$\rho_1(\mathbf{a}) = \begin{cases} 1/r & r \geq r_0 \\ \rho_0 & r < r_0, \end{cases} \quad (14)$$

where r_0 is a constant. In (14), limiting of the parameter weighting function to ρ_0 avoids a large weighting for small radii, which is undesirable for detecting the pupil's parameters. In order to reduce the required computation, we can also change the parameter weighting function, similar to (12), into a step weighting function:

$$\rho_2(\mathbf{a}) = \begin{cases} \rho_1 & r \leq R_1 \\ \rho_2 & R_1 < r \leq R_2 \\ \rho_3 & r > R_2 \end{cases}. \quad (15)$$

The radius between R_1 and R_2 is the most likely parameter so we should give it the largest weighting value ρ_2 while assigning a smaller weight to the radius under R_1 or over R_2 . Hence, the weighting factors satisfy $\rho_2 > \rho_1$ and $\rho_2 > \rho_3$. Actually, we can combine both the spatial and parameter weighting processes together to achieve better detection performance as follows:

$$\begin{aligned} C_R(\mathbf{a}) &= H_\varepsilon \{d_R(x, y) / F(x, y, \mathbf{a}) \leq \varepsilon\} = \sum_{F(x, y, \mathbf{a}) \leq \varepsilon} \rho_2(\mathbf{a}) w_\varepsilon(x, y, \mathbf{a}) d_B(x, y) \\ &= \sum_{F(x, y, \mathbf{a}) \leq \varepsilon} d_R(x, y), \end{aligned} \quad (16)$$

where

$$d_R(x, y) = d_B(x, y) \rho_2(\mathbf{a}) w_\varepsilon(x, y, \mathbf{a}). \quad (17)$$

4. DETECTION OF FEATURE PARAMETERS

When used to fit the analytic equation, the HT and modified HT (MHT) algorithms require the same amount of computation. Because weighting factors are introduced, the modified HT algorithms must determine and accumulate non-integer factors while the standard HT only needs to accumulate only ones. To detect a circle from a binarized image containing 863 edge pixels, we applied the HT and the MHT algorithms to compute a

100×100×100 parameter space. Table 1 shows the CPU times required by the HT and MHT, which were executed on Pentium-233 personal computer. The computation required by the modified HT slightly increased more than the computation required by the standard HT did. However, the modified HT algorithm achieved greater robustness than the HT did. For the detection of the pupil's parameters, the counts versus the 3-D parameter cells could not be plotted. Here, we reduce the radius, r -axis by selecting the maximum count from those ones, who has the same circle center (u, v) as

$$C_R^{\max,r}(u, v) = \max_r \{C_R(r, u, v)\}. \quad (18)$$

Table 1. A CPU time comparison between the HT and MHT.

Algorithms	Computation Time (sec)	Remarks
HT(Eq.6)	7.85	Conventional HT
MHT(Eq.8)	10.00	$\epsilon=3$
MHT(Eq.14)	9.94	$r_0=15$
MHT(Eq.16) ⁺	12.85	$\epsilon=3, r_0=15$

⁺ with continuous weighting $\rho_2(\mathbf{a})$

Fig. 1 shows the accumulated maximum-radius counts versus the circle center (u, v) parameter space obtained by using the standard HT. We find the HT shows three significant peaks, and that it is very difficult to identify the true circle center. As for the MHT, however, Fig. 2 shows a significant and sharp peak in the target center, so we can easily locate the feature parameters. Furthermore, we also find that the accumulated counts obtained by using the MHT are smoother than those obtained by using the HT.

Due to its robust and smooth performance, we further apply a coarse-to-fine search mechanism to the MHT to reduce the required computation. The coarse MHT search was performed on a down-sampled image and with a lower resolution parameter space to roughly resolve an evident peak in the first stage. Around the evident peak, we then applied a fine-resolution MHT search to precisely detect the desired feature parameters. If the total number of edge pixels $d_B(x, y)$ is assumed to be N_I , and if the number of accumulation cells in the parameter space is $n_r \times n_u \times n_v$, then applying the HT and MHT, we need 4 additions and 3 multiplications to determine one vote in the parameter space for detecting a circle. To complete the transformation, we need in total $4N_I n_r n_u n_v$ additions and $3N_I n_r n_u n_v$ multiplications. If we execute the MHT on the down-sampled image with step k and a low-resolution parameter space with $n'_r \times n'_u \times n'_v$ cells in the coarse search step, then the MHT requires $4N_I n'_r n'_u n'_v / k^2$ additions and $3N_I n'_r n'_u n'_v / k^2$ multiplications. Due to the located evident peak, we can narrow the parameter space and give the finest parameter resolution. We assume that the number of accumulation cells is $n''_r \times n''_u \times n''_v$, then the MHT requires $4N_I n''_r n''_u n''_v$ additions and $3N_I n''_r n''_u n''_v$ multiplications in the fine search step. If the image is down-sampled by 2 (i.e., $k = 2$), if the parameter space resolution is scaled by 3 (i.e., $n_r = 3n'_r$, $n_u = 3n'_u$, and $n_v = 3n'_v$), and if we choose the same number of parameter cells for both coarse and fine searches (i.e., $n''_r = n'_r$, $n''_u = n'_u$, and $n''_v = n'_v$), then the direct MHT requires $108N_I n'_r n'_u n'_v$ additions and

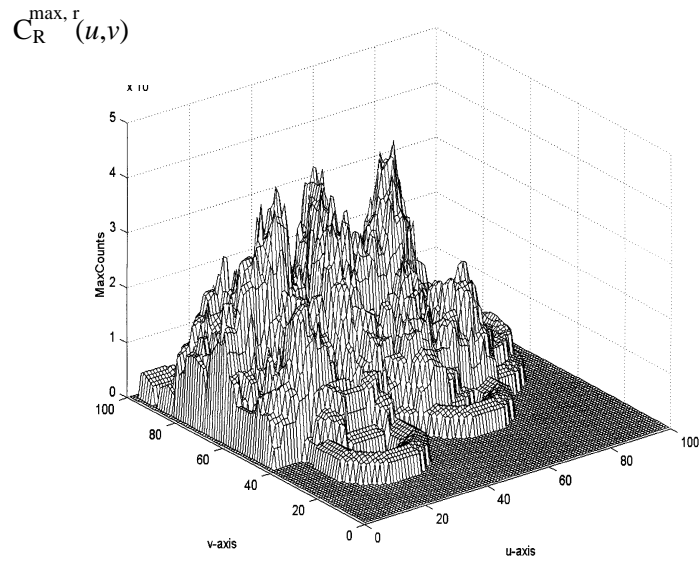


Fig. 1. Accumulated maximum-radius counts versus the circle center (u, v) parameter space for the traditional HT method.

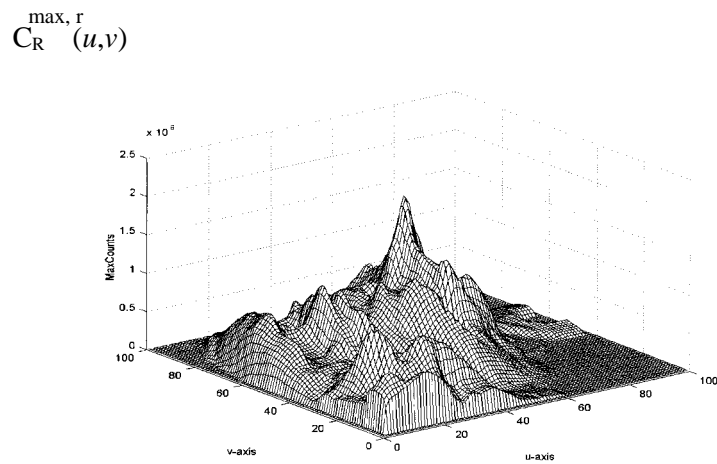


Fig. 2. Accumulated maximum-radius counts versus the circle center (u, v) parameter space for the proposed MHT method.

$81N_p n'_u n'_v$ multiplications to achieve the same parameter resolution as the coarse-to-fine MHT. However, the MHT needs about $N_p n'_u n'_v$ additions and $3/4 N_p n'_u n'_v$ multiplications in the coarse search step, and needs $4N_p n'_u n'_v$ additions and $3N_p n'_u n'_v$ multiplications in the fine search step. The coarse-to-fine MHT reduces the amount of computation required by the MHT to achieve the similar detection performance by about 22 times. Fig. 3 shows the accumulated counts obtained by using the MHT in the fine search step. The MHT exhibits a cleaner shape peak for the desired feature parameters.

By using the traditional HT, Fig. 4 depicts the detected results from four eye's images. The results show that the extracted feature parameters by the HT are far away from the true pupil curves. The HT does not achieve satisfactory results. With spatial and parameter weighting functions, Fig. 5 shows that the coarse-to-fine MHT exhibits visible improvement in detection of pupil parameters. Although the MHT in the coarse step show slightly variation in (a1) to (a4), the MHT in the fine search step successfully detects the feature parameters. It is noted that the coarse-to-fine MHT achieves much better detection performance and much lower computation than the traditional HT. To compare other modified HT algorithm, Fig. 6 shows the detected results achieved by the proposed MHT and the RWHT [19]. To detect other facial features, for example the upper lips, Fig. 7 shows the detected results by the HT, the MHT and the RWHT. Since the upper lips do not have any other competitive curve, all three algorithms can track the main curves of the upper lips. However, the proposed MHT achieves the best match. All results demonstrate that the proposed MHT exhibits a better detection performance than the HT and the RWHT algorithm [19].

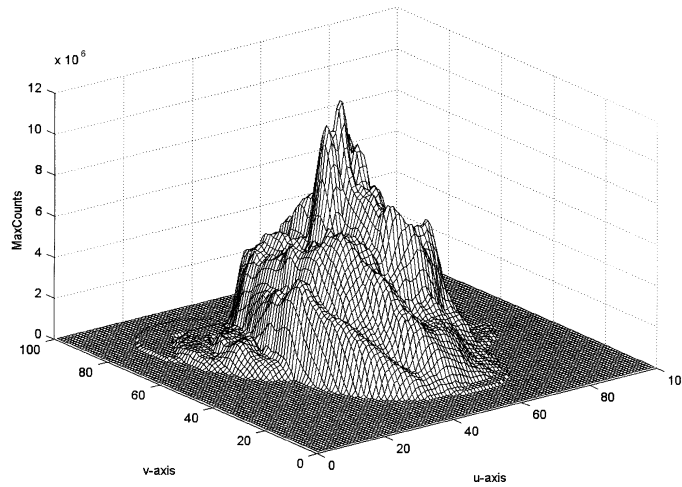


Fig. 3. Accumulated maximum-radius counts versus the circle center (u, v) parameter space for the proposed coarse-to-fine MHT in the fine search step.

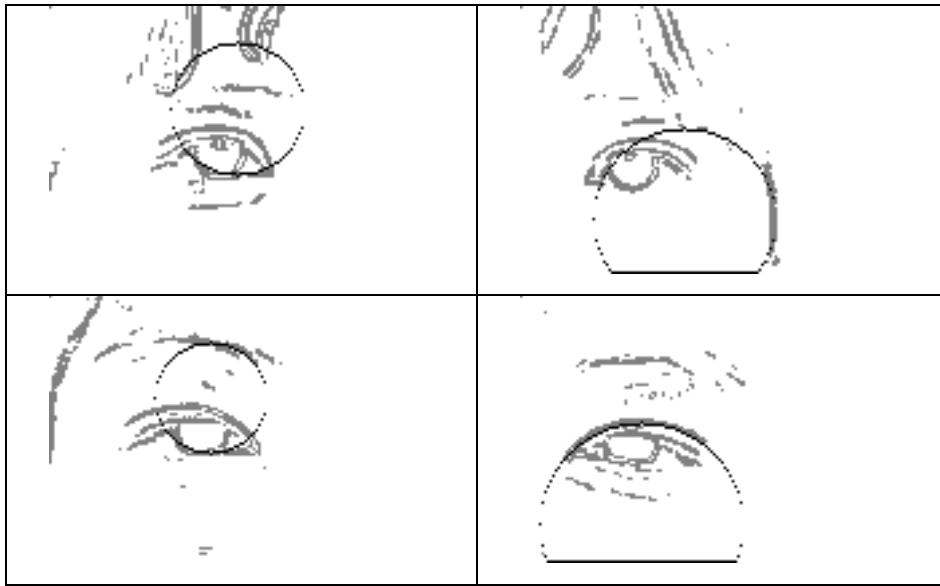
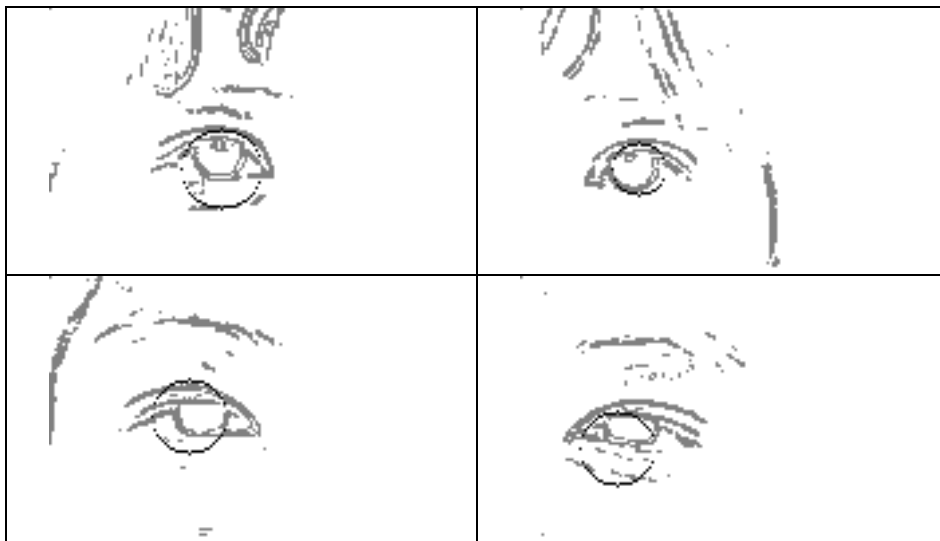
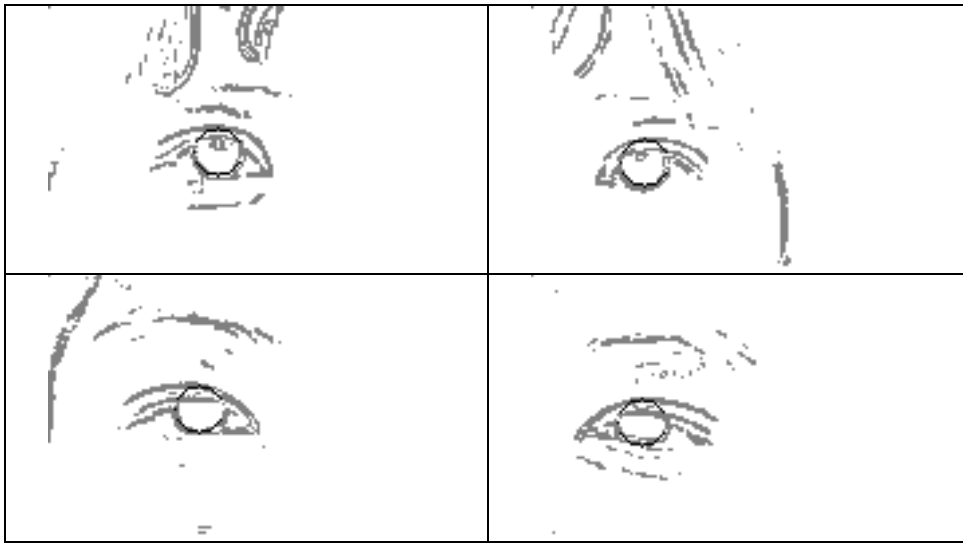


Fig. 4. Pupil features extracted by the original HT.



(a)

Fig. 5. Pupil features extracted by the MHT using the coarse-to-fine searching algorithm: (a) in the coarse search step; (b) in the fine search step.



(b)

Fig. 5. (Cont'd) Pupil features extracted by the MHT using the coarse-to-fine searching algorithm: (a) in the coarse search step; (b) in the fine search step.



Fig. 6. Pupil features extracted by the proposed MHT (upper row) and the RWHT [19] (lower row).

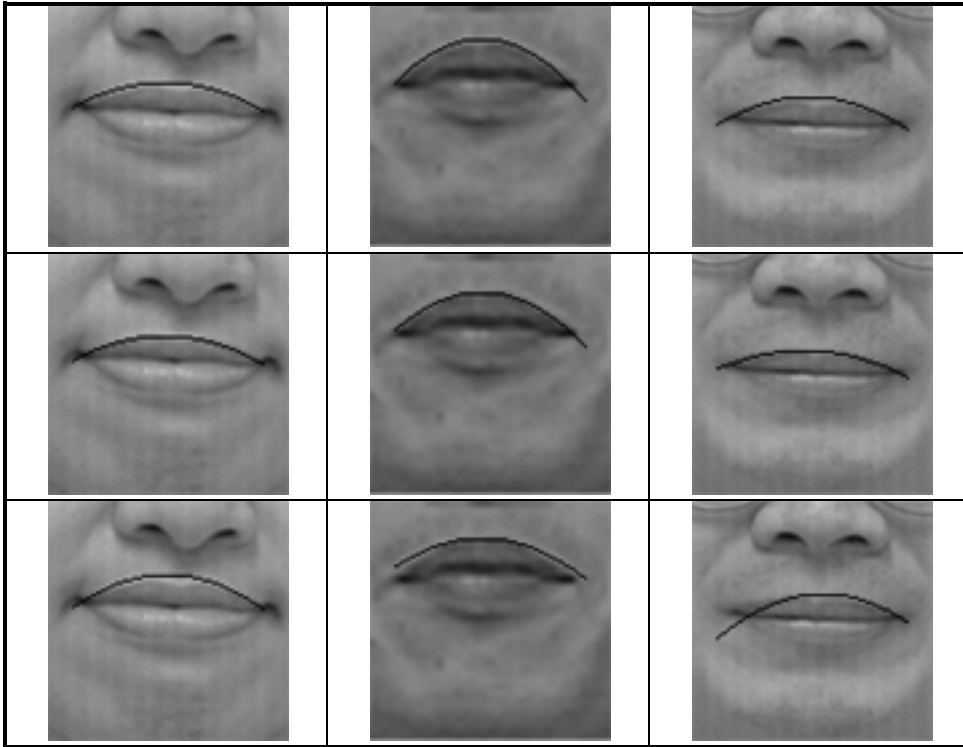


Fig. 7. Upper lips detected by the HT(1st row), MHT (2nd row), and the RWHT [19] (3rd row).

5. CONCLUSIONS

In this paper, we proposed modified HT algorithms to effectively extract facial feature parameters. By introducing spatial and parameter weighting functions, the MHT with slightly increasing the computation complexity shows robust and precise detection in extracting the facial parameters. The spatial weighting results in an averaging approach for the noisy pixels while the parameter weighting removes the unlikely curves from the parameter space. Due to the robustness, we can further adopt a coarse-to-fine search procedure to reduce the computation time and memory occupation of the MHT. Simulations show that the coarse-to-fine MHT achieves better detection performance and requires lower computation complexity than the traditional and other modified HT algorithms.

ACKNOWLEDGEMENT

The authors would like to thank the anonymous referees whose careful reviews and detailed comments helped improve the readability of this paper.

REFERENCES

1. R. Talluri, K. Oehler, T. Bannon, J. D. Courtney, A. Das, and J. Liao, "A robust, scalable, object-based video compression technique for very low bit-rate coding," *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 7, No. 1, 1997, pp. 221-233.
2. A. Yullie, D. Cohen, and P. Hallian, "Facial feature extraction by deformable templates," Technical Report 88-2, Harvard Robotics Laboratory, 1988.
3. C.-L. Huang and C.-W. Chen, "Human facial feature extraction for face interpretation and recognition," *Pattern Recognition*, Vol. 25, No. 12, 1992, pp. 1435-1444.
4. G. Yang and T. S. Huang, "Human face detection in a complex background," *Pattern Recognition*, Vol. 27, No. 1, 1994, pp. 53-63.
5. G. Chow and X. Li, "Towards a system for automatic facial feature detection," *Pattern Recognition*, Vol. 26, No. 12, 1993, pp. 1739-1755.
6. X. Li and N. Roeder, "Face contour extraction from front-view images," *Pattern Recognition*, Vol. 28, No. 8, 1995, pp. 1167-1179.
7. X. Xie, R. Sudhakar, and Z. Zhuang, "On improving eye feature extraction using deformable templates," *Pattern Recognition*, Vol. 27, No. 6, 1994, pp. 791-799.
8. K.-M. Lan and H. Yan, "Locating and extracting the eye in human face images," *Pattern Recognition*, Vol. 29, No. 5, 1996, pp. 771-779.
9. J. Sklansky, "On the Hough technique for curve detection," *IEEE Transactions on Computers*, Vol. 27, No. 10, 1978, pp. 923-926.
10. P.-K. Ser and W.-C. Siu, "Memory compression for straight line recognition using the Hough transform," *Pattern Recognition Letters*, Vol. 16, No. 2, 1995, pp. 133-145.
11. G. Eichmann, "Topologically invariant texture description," *Computer Vision Graphics & Image Processing*, Vol. 41, No. 3, 1987, pp. 267-281.
12. M. Kushnir, K. Abe, and K. Matsumoto, "Recognition of handprinted Hebrew characters using features selected in the Hough transform space," *Pattern Recognition*, Vol. 18, No. 2, 1985, pp. 103-114.
13. H. Kalviainen, P. Hirvonen, L. Xu, and E. Oja, "Probabilistic and non-probabilistic Hough transforms: overview and comparisons," *Image and Vision Computing*, Vol. 13, No. 4, 1995, pp. 239-252.
14. J. Illingworth and J. Kittler, "The adaptive Hough transform," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 9, No. 5, 1987, pp. 690-698.
15. L. Xu, E. Oja, and P. Kultanen, "A new curve detection method: randomized Hough transform (RHT)," *Pattern Recognition Letters*, Vol. 11, No. 5, 1990, pp. 331-338.
16. N. Guil, J. Villalba, and E. L. Zapata, "A fast Hough transform for segment detection," *IEEE Transactions on Image Processing*, Vol. 4, No. 11, 1995, pp. 1541-1548.
17. S.-S.Hao, M.-F.Lee, and J.-F. Yang, "Pupil detection methods for low-bit rate videophone," *1997 Workshop on Consumer Electronics: Digital Video and Multimedia*, B2-3, 1997, pp. 12-17.
18. S.-S. Hao, M.-F. Lee, and J. F. Yang, "Modified Hough transforms for extracting pupil features," in *Proceedings of the 1997 International Symposium on Communications*, 1997, pp. 308-312.

19. J. Forsberg, U. Larsson, and A. Wernersson, "Mobile robot navigation using the range-weighted Hough transform," *IEEE Robotics & Automation Magazine*, Vol. 11, No. 3, 1995, pp. 18-26.

Jar-Ferr Yang (楊家輝) was born in Keelung, Taiwan, on September 15, 1954. He received the B.S. degree from Chung-Yuan Christian University in 1977, the M.S. degree from National Taiwan University in 1979, and the Ph.D. degree from the University of Minnesota Minneapolis, U.S.A., in 1988, all in Electrical Engineering. He was an instructor at the Chinese Naval Engineering School from 1979 to 1980, and he worked for the Data Transmission and Network Design Research group, Chunghwa Telecommunication Laboratories, from 1981 to 1984. Currently, he is a professor in the Department of Electrical Engineering and the Chairman of the Computer and Communication Research Center, National Cheng Kung University. His research interests are in the areas of signal processing and compression, fast algorithm design, neural networks, and array signal processing.

Shu-Sheng Hao (郝樹聲) was born in Taipei, Taiwan, on October 12, 1959. He received the B.S. and M.S. degrees in Electrical Engineering from National Cheng Kung University, Taiwan, in 1983 and 1985, respectively. He is a Ph.D. student at NCKU now. His research interests include model-based image coding, signal processing, and low-bit-rate video compression.