Correspondence

Facial Expression Recognition From Line-Based Caricatures

Yongsheng Gao, Maylor K. H. Leung, Siu Cheung Hui, and
Mario W. Tananda

Abstract—**The automatic recognition of facial expression presents a significant challenge to the pattern analysis and man-machine interaction research community. Recognition from a single static image is particularly a difficult task. In this paper, we present a methodology for facial expression recognition from a single static image using line-based caricatures. The recognition process is completely automatic. It also addresses the computational expensive problem and is thus suitable for real-time applications. The proposed approach uses structural and geometrical features of a user sketched expression model to match the line edge map (LEM) descriptor of an input face image. A disparity measure that is robust to expression variations is defined. The effectiveness of the proposed technique has been evaluated and promising results are obtained. This work has proven the proposed idea that facial expressions can be characterized and recognized by caricatures.**

Index Terms—**Caricature, classification, facial expression recognition, line edge map, static image.**

I. INTRODUCTION

In recent years, the demand for human-machine interaction systems has increased significantly. Consequently, the automatic identification of facial expression has become an active research area. Facial expressions, as a reflection of emotions, have been extensively studied in psychology [1]–[4]. The study of Mehrabian [4] indicated that in face-to-face communication of human, 7% of the communication information is transferred by linguistic language, 38% by paralanguage, and 55% by facial expressions. This implies that facial expression plays an important role in human communication, and in a human-like communication between man and machine as well.

Compared with face recognition, there is relatively a small amount of work on facial expression recognition. The existing approaches include facial motion analysis using image sequences [5]–[10], [17], neural network approaches [9], [14], spatial pattern analysis [11]–[14], and models of facial features [7], [10], [13], [15], [16].

The facial action coding system (FACS) [3] is the most widely used method to measure facial movement. The optical flow analysis [5], [7]–[10], [18], [19] is the major technique for facial motion measurement. Mase [7] used optical flow to estimate the activity of the facial muscles. A recognition accuracy of 80% was reported on 30 test cases. Essa and Pentland [10] extended this approach using an anatomical and physical model of the face. They refined the facial motion estimation from optical flow using recursive estimation. The estimated forces were used to recognize the facial expression. A recognition accuracy of up to 98% was achieved in their research. Yacoob and Davis [8] constructed a mid-level representation of facial motion directly from optical flow

using heuristic rules instead of a physical model. Black and Yacoob [6] extend this method, using local parameterized models of image motion. They achieved 89% accuracy at recognizing expressions in their database of 70 image sequences containing a total of 145 expression records. Donato *et al.* [5] analyzed the facial actions through optical flow, holistic spatial analysis, Gabor wavelet representation and independent component representation. Their system can correctly classify up to 96% of the test expressions. Tian, Kanade, and Cohn [42] developed an automatic face analysis (AFA) system for analyzing facial expressions based on both permanent and transient facial features in image sequences. The AFA system recognizes fine-grained changes in facial expression into actions units of the FACS. Multistate facial component models were used for tracking and modeling the various facial features, such as lips, eyes, brows, cheeks, and furrows. The average recognition rates of up to 96% were reported. Fasel and Luettin [43] reported a system that recognizes asymmetric FACS action unit activities and intensities without using markers. The performance of their system is comparable to marker-based systems or human FACS experts. Pantic and Rothkrantz [20] presented a face action recognition strategy based on their proposed face model and multiple feature detection algorithms applied in parallel. They introduced the side-view of human faces into the expression classification. The reported integrated system for facial expression recognition (ISFER) is a completely automated system that has been developed to convert the face geometry into a description of face actions, and then into weighted emotion labels.

At present, the recognition of facial expression by computer mainly focuses on studying the facial motion from image sequences. From these methods, the motion characteristic of human faces is analyzed in detail. The existing works show that expressions can be recognized from image sequences and satisfactory recognition rates have been reported [5], [10]. Facial action coding is a more detailed analysis of facial behavior than discriminating static "mug-shot" pictures of expressions. Psychological study [21] showed that facial expressions were more accurately recognized from dynamic images than from a single static "mug-shot" image. However, the information in a single image is sometimes enough to recognize the expression. The fact that humans can recognize facial expressions from just a single static image is an example. In many multimedia and man-machine interface applications, such as multimedia data retrieval over WWW, expression-based face recognition and interactive Internet games, image sequences of detailed facial actions are not available but a static image. Therefore, how to analyze facial expressions through static images is of great importance. However, it still remains an unsolved problem. Cottrell and Metcalfe [11] recognized facial expressions using holistic representations based on principal components. Padgett and Cottrel [12] extended the work on feature regions using principal component analysis. Lanitis [13] used flexible models to interpret the facial expressions from static image. They achieved 74% correct classification using shape-free model, 53% accuracy using shape model and 70% using combined model. Zhang [14] applied multiplayer perceptron on geometric positions of a set of fiducial points and Gabor wavelet coefficients extracted from the face image at the fiducial points. The best performance obtained was 92.3%. However, the 34 fiducial points used in this approach were selected manually. Automatic detection of these points is an open question.

One of the key issues in expression recognition from a static image is the choice of appropriate image descriptors and corresponding sim-

ilarity measure, which must be robust to the relatively large variations of facial geometric deformations for identical expression among human populations. In this paper, we proposed a facial expression recognition method from line-based caricatures. The proposed approach uses line edge map (LEM) [22] as expression descriptor. Similarity is obtained by computing the proposed directed line segment Hausdorff distance (dLHD) between the query face LEM and the caricature models of expressions. The significance and challenge of this work are that the recognition is conducted from simple line-drawing caricatures characterizing facial expressions. No specific correspondence of feature points is needed except automatically normalizing the input image by aligning the positions of the two eyes. The whole recognition process is completely automatic.

The paper is organized as follows. Section II presents the extraction of facial expression descriptor. In Section III, a directed line segment Hausdorff distance for line matching is proposed and described in detail. The idea of caricature models characterizing facial expressions and a splitting process are introduced in Section IV. Experimental results on system evaluation are reported in Section V. Finally, the conclusion of the paper is given in Section VI.

II. EXTRACTION OF FACIAL EXPRESSION DESCRIPTOR

Edges are the most fundamental features of objects in the three–dimensional (3-D) world. The edges in an image reflect substantial local intensity changes that are caused by the geometrical structure of the object, the characteristics of surface reflectance of the object, and viewing direction. The actions of facial expressions produce 3-D physical feature deformations of the face, which are reflected in geometrical structure changes of its edge map. Edge detection is an important technique in computer vision and pattern recognition systems since the edges usually correspond to some important properties of objects, such as object shape and boundary.

Edge extraction has been studied most extensively, and many reliable algorithms have been proposed and implemented [23]–[34]. A popular scheme for edge detection is to find the extrema in the blurred spatial derivates of the intensity images. The idea of blurred Laplacian filtering is extensively elaborated by Marr *et al.* [23], [24]. The Canny edge detection algorithm [28] is considered a "standard method" used by many researchers. Based on the assumption that intensity edges are ideal step edges corrupted by Gaussian noise, the Canny edge detector uses linear filtering with a Gaussian kernel to smooth noise and then computes the edge strength and direction for each pixel in the smoothed image. The edge detection algorithms proposed by Chen and Yang [29], Nevatia *et al.* [30], Nalwa *et al.* [31], Iverson *et al.* [32], Bergholm [33] and Rothwell *et al.* [34] are only a few examples from a large volume of these algorithms that had been presented to the vision community. The continuing development of edge detectors is producing increasingly complex edge detection algorithms. However, the "increased sophistication (of newer algorithms) is not producing a commensurate improvement in performance" [25] since the early algorithms of Roberts [26] and Sobel [27]. Heath *et al.* [35], [36] had investigated the performance of different edge detectors. They compared the edge detectors based on experimental psychology and statistics, in which humans rated the outputs of low level vision algorithms. One of their clear results is that no one single edge detector was best overall and for any given image it is difficult to predict which edge detector will be best [36].

The proposed facial expression recognition method does not rely on any specific edge detector. It is a general method that any edge detector can be used. In this study, an edge detector based on the algorithm of Nevatia [30] is used followed by a thinning process to generate one-pixel wide edge curves.
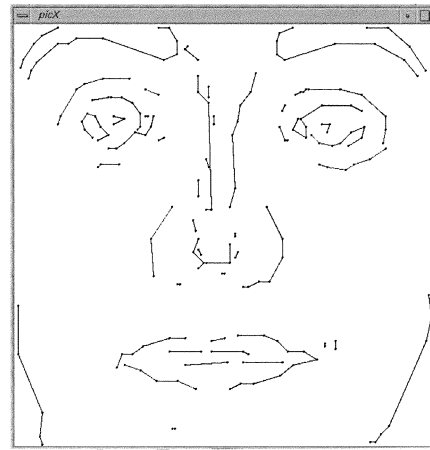

Fig. 1. Illustration of the LEM of a face.

To generate the LEM, the dynamic two strip algorithm (Dyn2S) [38] is utilized to detect dominant points on the facial edge curves. In Dyn2S algorithm, a strip is fitted to the left and right of each point on the curve, and the points inside each strip are approximated as a straight line. The orientation and width of the strip are adjusted automatically. Longer and narrower strips are favored. A measure of "merit" based on the strips lengths and widths, and the angle between the strips is calculated. This merit provides an objective evaluation of the prominent strength of each point. The result of applying these processes on a face is illustrated in Fig. 1. The LEM representation enhances the geometrical structure properties in the edge map. By grouping low level pixels into higher level line features, the oriented structural features are extracted to characterize the facial expressions.

III. LINE-BASED DISPARITY MEASURE

It is noticed that facial expressions distort the facial edge curves in position and orientation. The measurement of the degree of the distortion could provide a solution to the facial expression recognition task. The Hausdorff distance is a shape comparison metric based on binary images. It is a distance measure defined between two point sets. Since the Hausdorff distance uses spatial information of edge curves but lacks structural information, which is critical in representing facial expressions, its discriminative capability can be strengthened and enhanced by incorporating orientation information. A disparity measure, line segment Hausdorff distance (LHD), was proposed by Gao and Leung [22] to measure the similarity of human faces. The LHD is a global shape comparison measure between two LEM images. However, it is not suitable for recognizing facial expressions characterized by caricatures. A line-based caricature depicts a facial expression by a few lines, which are the common features of that particular expression and independent of individual subjects. Other edge information or lines, which are not generated by the expression, such as the lines of the nose and the outline of the face, are not necessary to appear in the caricature. This can improve the computational efficiency of the algorithm. In other words, an expression caricature characterizes only part of the face LEM lines (i.e., a subset of LEM) that are unique to the expression. Fig. 2 shows three examples of the line-based caricatures used in this study. Therefore, the expression recognition task in this research is to identify the corresponding lines of the caricature in the input face LEM and measure the similarity between the caricature and the matched part of the LEM. Obviously, it is not a problem of measuring the global similarity between two line sets that LHD can address. In this study, a directed line segment Hausdorff distance (dLHD) is proposed to measure the similarity from the line-based caricature of the facial
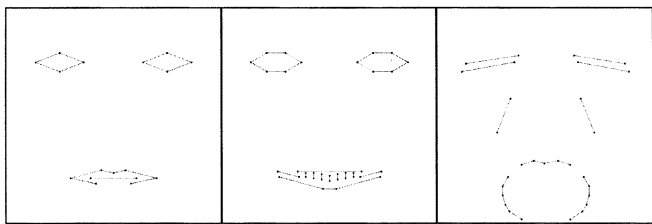
Fig. 2. Manually sketched facial models of neutral, smiling and screaming expressions.

Fig. 3. Facial expression models after splitting process.

expression to the LEM of the input face image. dLHD is a disparity measure defined between two line sets to measure the difference from one line set (A) to a subset (B_{sub}) of the other line set (B). The subset B_{sub} is automatically identified during the matching process, which provides the smallest value of dLHD. Unlike most shape comparison methods that build a one-to-one line correspondence between a model and a test image, dLHD can be calculated without explicit line correspondence to deal with the adding, missing and broken line problem caused by segmentation error. The dLHD for LEM image matching is more tolerant to perturbations in the locations of lines than correlation techniques since it measures proximity rather than exact superposition. This property is crucial for recognizing facial expressions from static images because the degrees of physical deformations of identical expression are different from person to person. Moreover, there is only one caricature for each expression model, which cannot be an exact replica of any input face.

Given an input LEM $T = \{t_1, t_2, \ldots, t_p\}$ and a line set of caricature model $M = \{m_1, m_2, \ldots, m_q\}$, dLHD is built on the disparity vector $\vec{d}(m_i, t_j)$ between lines m_i and t_j

$$\vec{d}(m_i, t_j) = \lfloor d_\theta(m_i, t_j) \quad d_{//}(m_i, t_j) \quad d_\perp(m_i, t_j) \rfloor$$

where $d_\theta(m_i, t_j), d_{//}(m_i, t_j), d_\perp(m_i, t_j)$ are angular, parallel and perpendicular distances, respectively. The angular distance measures the structural/directional difference between m_i and t_j. The parallel and perpendicular distances measure the differences of geometrical locations.

Definition 1: $d_\theta(m_i, t_j) = f(\theta(m_i, t_j))$. $\theta(m_i, t_j)$ computes the smaller intersecting angle between lines m_i and t_j. $f(\ )$ is a nonlinear penalty function to map the angle to a scalar. It is desirable to encourage small angular variation (which is most likely the intra-class variation) but penalize heavily on large deviation (which is most likely the inter-class difference). In this study, the tangent function is used.

Definition 2: $d_{//}(m_i, t_j) = \min[\text{diag}(\mathbf{L}_{//mt})]$. Given two horizontal lines m_i and t_j, the horizontal distances between any two of the four ends of lines m_i and t_j along the horizontal direction are defined by the matrices

$$\mathbf{L}_{//mt} = \begin{bmatrix} l_{m_l t_l}, & l_{m_l t_r} \\ l_{m_r t_l}, & l_{m_r t_r} \end{bmatrix} \quad \text{and} \quad \mathbf{L}_{//tm} = \begin{bmatrix} l_{t_l m_l}, & l_{t_l m_r} \\ l_{t_r m_l}, & l_{t_r m_r} \end{bmatrix}$$

where the subscripts l and r stand for left and right ends of line m_i or t_j. Since $l_{m_l t_r} = l_{t_r m_l}$, we have $\mathbf{L}_{//mt} = \mathbf{L}_{//tm}^T$. diag $(\mathbf{L}_{//mt})$ denotes the set of diagonal elements of matrix $\mathbf{L}_{//mt}$.

$d_{//}(m_i, t_j)$ is defined as the minimum distance of $\text{diag}(\mathbf{L}_{//mt})$, which means dLHD only reflects the smallest shift of the line end-points. If one of the end-points of m_i is located exactly at the same horizontal location of the corresponding end-point of t_j and the other one shifts, the parallel distance remains zero no matter how far the other end-point of m_i shifts. This helps to alleviate the uncertainty problem of caricature lines drawn by users.
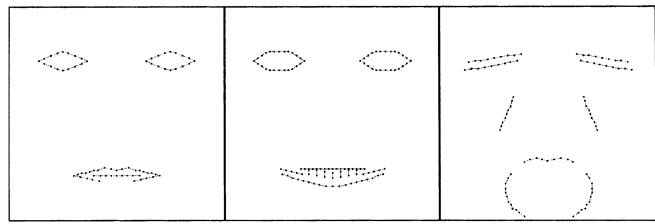
Definition 3: $d_{//}(m_i, t_j) = 0$, if $(m_l < t_l \cap m_r > t_r) \cup (m_l > t_l \cap m_r < t_r)$. In order to cater for the effect of shifting and uncertainty of line end-points, $d_{//}(m_i, t_j)$ is reset to zero if one line is within the range of the other.

Definition 4: $d_\perp(m_i, t_j)$ is the vertical distance between the lines m_i and t_j.

In general, m_i and t_j would not be parallel. The shorter line is then rotated with its midpoint as rotation center to make it parallel to the other line. Subsequently, the coordinate system is rotated such that the two lines are horizontal before computing $d_{//}(m_i, t_j)$ and $d_\perp(m_i, t_j)$. The distance between two lines m_i and t_j is defined as

$$d(m_i, t_j) = \sqrt{(W \cdot d_\theta(m_i, t_j))^2 + d_{//}^2(m_i, t_j) + d_\perp^2(m_i, t_j)} \tag{1}$$

where W is the weight of the angular distance to balance the contributions of the angle and the displacement. The directed line segment Hausdorff distance (dLHD) from M to T is defined as

$$\text{dLHD}(M, T) = \frac{1}{\sum_{m_i \in M} l_{m_i}} \sum_{m_i \in M} l_{m_i} \cdot \min_{t_j \in T} d(m_i, t_j) \tag{2}$$

where l_{m_i} is the length of line m_i. Every $\min_{t_j \in T} d(m_i, t_j)$ for the line m_i is weighted by the normalized length of m_i (i.e., $\frac{l_{m_i}}{\sum_{m_i \in M} l_{m_i}}$) because its contribution to dLHD is assumed to be proportional to the length of the line. When the two lines to be matched are perpendicular, i.e., $\theta(m_i, t_j) = 90°$, which is the most unlikely match, the searching of t_j with minimum distance $d(m_i, t_j)$ will never match such a perpendicular pair. Hence, the calculation of $\tan 90°$ is avoided by skipping it over.

IV. EXPRESSION CHARACTERIZATION USING CARICATURES

In this work, the user characterizes, on the screen, the approximate caricatures of the pictorial facial expressions. The recognition is reduced to matching of the LEM descriptor of the input face against the line-drawing expression models. The expression caricature consists only of straight lines. Fig. 2 displayed a set of facial expression caricatures drawn by a user using a prototype system which implements the proposed approach. Once a line is drawn, a symmetric pair is automatically generated on the screen by the system. The faces are assumed symmetric in this implementation to simplify the drawing process of users. This is a reasonable assumption because a picture of a human face and most human facial expressions are considered symmetric by human observers. Nevertheless, the proposed approach is not limited to symmetric expression models.

In order to increase the resolution of the dLHD matching process, a split scheme is used. If a line drawn by a user is longer than a preset threshold T_l, it is split into two at the midpoint of the line. The splitting process is applied recursively until all the lines are shorter than T_l. Fig. 3 shows the results of applying the splitting process on the original caricatures in Fig. 2.

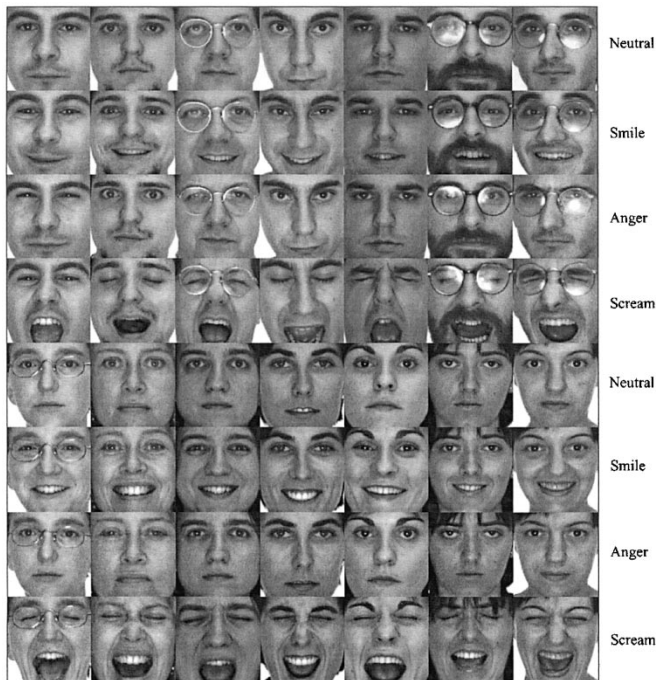Fig. 4.   Some results of the expression recognition demo system.

Fig. 5.   Examples of facial expressions in the database.

Fig. 6.   Examples of LEM descriptors of neutral (left), smiling (middle), and screaming (right) faces.

TABLE I
RECOGNITION RESULTS (%)

	Neutral	Smile	Scream	**Average**
Male	91.8	68.9	88.5	**83.1**
Female	96.1	90.2	86.3	**90.9**
Average	**93.8**	**78.6**	**87.5**	**86.6**

Fig. 7.   Effect of W.

V. EXPERIMENTAL RESULTS

A prototype facial expression recognition system for still images using the proposed techniques was implemented (Fig. 4). The facial expression images from the AR face database [39] of Purdue University were used to evaluate the system performance. Original images were normalized (in scale and orientation) and cropped such that the two eyes are aligned roughly at the same position with a distance of 80 pixels in the final images. Many reliable automatic face detection and eye location algorithms are available, such as [6], [44]–[48]. A few examples of the cropped faces are shown in Fig. 5. For the details on the collection of the images in the AR face database, readers can refer to [40]. The anger faces in the database were not used in this experiment as many subjects did not make a "typical" angry expression (see Fig. 5) and our users can not draw an anger caricature based on the characteristics of the anger faces in the database. In this experiment, a total of 336 images of 112 individuals (61 males and 51 females) were tested. The numbers of images corresponding to neutral, smile and screaming expressions are the same.
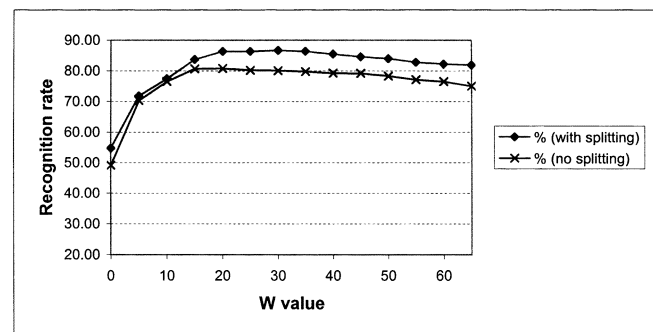
Some examples of the LEM descriptors generated by the system are displayed in Fig. 6. The experimental results on facial expressions are summarized in Table I. The effects of W and splitting threshold are illustrated in Figs. 7 and 8. It is found that the system without using structural/orientation information, i.e., $W = 0$ in Fig. 7, could only

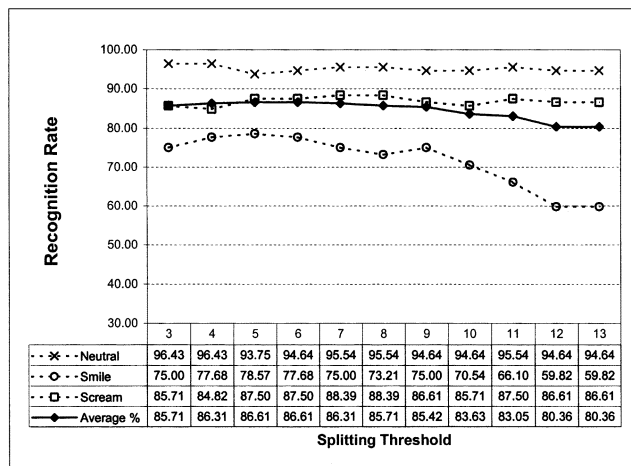|	3	4	5	6	7	8	9	10	11	12	13
- ✕ - Neutral	96.43	96.43	93.75	94.64	95.54	95.54	94.64	94.64	95.54	94.64	94.64
- ○ - Smile	75.00	77.68	78.57	77.68	75.00	73.21	75.00	70.54	66.10	59.82	59.82
- ☐ - Scream	85.71	84.82	87.50	87.50	88.39	88.39	86.61	85.71	87.50	86.61	86.61
—◆— Average %	85.71	86.31	86.61	86.61	86.31	85.71	85.42	83.63	83.05	80.36	80.36

Fig. 8.   Effect of splitting process with $W = 30$.

correctly recognize 55% of the expressions. It improved quickly by introducing the LEM orientation features into the dLHD calculation and reached the optimal value of 86.6% when W ranged from 20 to 30. The system performance then degraded gracefully. The splitting process increased the system performance by 6.6% (Fig. 7).

It is interesting to note that the average recognition accuracy on the expressions of females is higher than that of males by 7.8% (Table I). The system correctly recognized 90.9% of women's facial expressions while 83.1% of men's expressions were correctly recognized. This might imply that females make more standard and expressive expressions than males. It is also found that the recognition rate of smiling face is the lowest, particularly for males. This could be due to the fact that smiling expressions are relatively more different from person to person. This could be improved by providing more detailed caricatures in future investigation.

It is difficult to compare directly our results to those of reported techniques because of differences in testing data. The techniques using facial motion information from image sequences achieved recognition rates from 80% [7] up to 96% [5], [42] and 98% [10] on different test sets. However, it is much more challenging to recognize facial expressions from a single static image than an image sequence containing details of the action. Lanitis [13] achieved 74% correct classification using shape-free model, 53% accuracy using shape model and 70% using combined model. The best performance obtained by Zhang [14] was 92.3%. However, the 34 fiducial points used in this approach were selected manually. Automatic detection of these points is an open question. Compared to these very few reported techniques on facial expression recognition from static images, the average recognition rate of 86.6% is a very promising result. Noted that the whole recognition process of the proposed method can be fully automatic. Our results are obtained by using an independent image database, which is publicly available. Thus, it can be used as a benchmark for direct performance comparison by other researchers.

The computational complexity of dLHD is of the order $O(kN_m N_t)$. N_m and N_t are the line numbers of the model and the test LEM. k is the time to compute $d(m_i, t_j)$ in (1). A time efficiency test was conducted on a SGI Octane workstation with 300 MHz CPU and 512 MB RAM. The average real computational time for one match is 6.992 ms. This result demonstrates that the proposed approach is a computational efficient technique, which is suitable for real-time applications.

VI. CONCLUSION

The automatic recognition of facial expression presents a significant challenge to the pattern analysis research community because the ex-

pression is generated by nonrigid object deformations and these deformations vary from person to person. The expression recognition from a static image is, particularly, a more difficult problem compared to the recognition from an image sequence due to lack of information during expression actions. In this paper, we present a technique for facial expression recognition from a single static image using line-based caricatures. The whole recognition process is completely automatic. The proposed technique also addresses the computational expensive problem of most existing facial expression techniques and thus it can be applied to real-time applications.

The proposed approach uses line matching of user sketched expression models over the LEM descriptor of the input face image. The defined dLHD has shown good tolerance to expression variations of identical expressions from a large data set (the AR face database) with 336 static expression images of over 100 subjects. The results of our study reveal that the facial expressions could be characterized and recognized by line-drawing caricatures, which is in accordance with the general belief that facial caricatures with simple lines catch the subjects' feature well [41]. Further study on appropriate caricatures to model the expressions will be carried out. In particular, the richness of facial expression requires the development of more sophisticated models to characterize the expression features.

REFERENCES

[1] J. N. Bassili, "Emotion recognition: The role of facial movement and the relative importance of upper and lower areas of the face," *J. Personality Social Psychol.*, vol. 37, pp. 2049–2059, 1979.

[2] P. Ekman and W. Friesen, *Unmasking the Face.* Englewood Cliffs, NJ: Prentice-Hal, 1975.

[3] ——, *The Facial Action Coding System.* San Francisco, CA: Consulting Psychologists, 1978.

[4] A. Mehrabian, "Communication without words," *Psychol. Today*, vol. 2, pp. 53–56, 1968.

[5] G. Donato, M. S. Bartlett, J. C. Hager, P. Ekman, and T. J. Sejnowski, "Classifying facial actions," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 21, pp. 974–989, Oct. 1999.

[6] M. J. Black and Y. Yacoob, "Recognizing facial expressions in image sequences using local parameterized models of image motioon," *Int. J. Comput. Vis.*, vol. 25, pp. 23–48, 1997.

[7] K. Mase, "Recognition of facial expression from optical flow," *Proc. IEICE Trans. Special Issue Computer Vision and Its Applications*, vol. 74, pp. 3474–3483, 1991.

[8] Y. Yacoob and L. Davis, "Recognizing human facial expression from long image sequences using optical flow," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 16, pp. 636–642, June 1996.

[9] M. Rosenblum, Y. Yacoob, and L. Davis, "Human expression recognition from motion using a radial basis function network architecture," *IEEE Trans. Neural Networks*, vol. 7, pp. 1121–1138, Sept. 1996.

[10] I. A. Essa and A. P. Pentland, "Coding, analysis, interpretation, and recognition of facial expressions," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 19, pp. 757–763, 1997.

[11] G. Cottrell and J. Metcalfe, "Face, gender and emotion recognition using holons," in *Advances in Neural Information Processing Systems*, D. Touretzky, Ed., 1991, vol. 3, pp. 564–571.

[12] C. Padgett and G. Cottrell, "Representing face images for emotion classification," in *Advances in Neural Information Processing Systems.* Cambridge, MA: MIT Press, 1997, vol. 9.

[13] A. Lanitis, C. J. Taylor, and T. F. Cootes, "Automatic interpretation and coding of face images using flexible models," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 19, pp. 743–756, July 1997.

[14] Z. Zhang, "Feature-based facial expression recognition: Sensitivity analysis and experiments with a multilayer perceptron," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 13, pp. 133–137, 1999.

[15] D. Terzopoulus and K. Waters, "Analysis and synthesis of facial image sequences using physical and anatomical models," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 15, pp. 569–579, June 1993.

[16] H. Li, P. Roivainen, and R. Forchheimer, "3-D motion estimation in model-based facial image coding," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 15, pp. 545–555, June 1993.

[17] N. Oliver, A. Pentland, and F. Bérard, "LAFTER: A real-time face and lips tracker with facial expression recognition," *Pattern Recognit.*, vol. 33, pp. 1369–1382, 2000.

[18] J. F. Cohn, A. J. Zlochower, J. J. Lien, Y. T. Wu, and T. Kanade, "Automated face coding: A computer-vision based method of facial expression analysis," *Psychophysiology*, vol. 35, pp. 35–43, 1999.

[19] J. J. Lien, T. Kanade, J. F. Cohn, and C. C. Li, "A multi-method approach for discriminating between similar facial expressions, including expression intensity information," *Proc. IEEE Conf. Computer Vision Pattern Recognition*, 1998.

[20] M. Pantic and L. J. M. Rothkrantz, "Expert system for automatic analysis of facial expressions," *Image Vis. Comput.*, vol. 18, pp. 881–905, 2000.

[21] J. N. Bassili, "Emotion recognition: The role of facial movement and the relative importance of upper and lower areas of the face," *J. Personality Social Psychol.*, vol. 37, pp. 2049–2059, 1979.

[22] Y. Gao and M. K. H. Leung, "Human face recognition using line edge maps," in *Proc. IEEE 2nd Workshop Automatic Identification Advanced Technologies*, NJ, Oct. 1999, pp. 173–176.

[23] D. Marr, *Vision: A Computational Investigation Into the Human Representation and Processing of Visual Information*. San Francisco, CA: Freeman, 1981.

[24] D. Marr and E. C. Hildreth, "Theory of edge detection," in *Proc. Royal Soc. London*, vol. B207, 1980, pp. 187–217.

[25] K. L. Boyer and S. Sarkar, "Assessing the state of the art in edge detection," *SPIE, Applicat. Artif. Intell. X: Mach. Vis. Robot.*, vol. 1708, pp. 353–362, 1992.

[26] L. G. Roberts *et al.*, "Machine perception of three-dimensional solids," in *Optical and Electro-Optical Information Processing*, J. T. Tippett *et al.*, Eds. Cambridge, MA: MIT, 1965, pp. 159–197.

[27] I. E. Sobel, "Cameral Model and Machine Perception," Ph.D. dissertation, Stanford Univ., Stanford, CA, 1970.

[28] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-8, pp. 679–698, 1986.

[29] G. Chen and Y. H. Yang, "Edge detection by regularized cubic B-spline fitting," *IEEE Trans. Syst., Man, Cybern.*, vol. 25, pp. 636–643, Apr. 1995.

[30] R. Nevatia and K. R. Babu, "Linear feature extraction and description," *Comput. Graph. Image Process.*, vol. 13, pp. 257–269, 1980.

[31] V. S. Nalwa and T. O. Binford, "On detecting edges," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-8, pp. 699–714, 1986.

[32] L. A. Iverson and S. W. Zucker, "Logical/linear operators for image curves," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 17, pp. 982–996, 1995.

[33] F. Bergholm, "Edge focusing," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 9, pp. 726–741, 1987.

[34] C. A. Rothwell, J. L. Mundy, W. Hoffman, and V. D. Nguyen, "Driving vision by topology," in *Intl. Symp. Computer Vision*, Coral Gables, Fla., Nov. 1995, pp. 395–400.

[35] M. D. Heath, S. Sarkar, T. Sanocki, and K. W. Bowyer, "A robust visual method for assessing the relative performance of edge-detection algorithms," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 19, pp. 1338–1359, Dec. 1997.

[36] ——, "Comparison of edge detectors: A methodology and initial study," *Comput. Vis. Image Understand.*, vol. 69, pp. 38–54, 1998.

[37] B. Takács, "Comparing face images using the modified hausdorff distance," *Pattern Recognit.*, vol. 31, pp. 1873–1881, 1998.

[38] M. K. H. Leung and Y. H. Yang, "Dynamic two-strip algorithm in curve fitting," *Pattern Recognit.*, vol. 23, pp. 69–79, 1990.

[39] AR Face Database, Purdue Univ., West Lafayette, IN. [Online]. Available: http://rvl1.ecn.purdue.edu/~aleix/aleix_face_DB.html

[40] A. M. Martinez and R. Benavente, "AR Face Database," CVC Tech. Rep. #24, Purdue Univ., West Lafayette, IN, June 1998.

[41] A. J. Calder, A. W. Young, P. J. Benson, and D. I. Perrett, "Self priming from distinctive and caricatured faces," *Brit. J. Psychol.*, pp. 141–162, 1996.

[42] Y. Tian, T. Kanade, and J. F. Cohn, "Recognizing action units for facial expression analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, pp. 97–115, Feb. 2001.

[43] B. Fassel and J. Luettin, "Recognition of asymmetric facial action unit activities and intensities," in *Proc. 15th Int. Conf. Pattern Recogniton*, 2000, pp. 1100–1103.

[44] E. Saber and A. M. Tekalp, "Frontal-view face detection and facial feature extraction using color, shape and symmetry based cost functions," *Pattern Recognit. Lett.*, vol. 19, pp. 669–680, 1998.

[45] K. M. Lam and H. Yan, "Locating and extracting the eye in human face images," *Pattern Recognit.*, vol. 29, pp. 771–779, 1996.

[46] H. A. Rowley, S. Baluja, and T. Kanade, "Neural network-based face detection," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 20, pp. 23–38, Jan. 1998.

[47] C. H. Lee, J. S. Kim, and K. H. Park, "Automatic human face location in a complex background using motion and color information," *Pattern Recognit.*, vol. 29, pp. 1877–1889, 1996.

[48] K. K. Sung and T. Poggio, "Example-based learning for view-based human face detection," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 20, pp. 39–51, Jan. 1998.

Automated Web Navigation Using Multiagent Adaptive Dynamic Programming

Joby Varghese and Snehasis Mukhopadhyay

Abstract—Today a massive amount of information available on the WWW often makes searching for information of interest a long and tedious task. Chasing hyperlinks to find relevant information may be daunting. To overcome such a problem, a learning system, cognizant of a user's interests, can be employed to automatically search for and retrieve relevant information by following appropriate hyperlinks. In this paper, we describe the design of such a learning system for automated Web navigation using adaptive dynamic programming methods. To improve the performance of the learning system, we introduce the notion of multiple model-based learning agents operating in parallel, and describe methods for combining their models. Experimental results on the WWW navigation problem are presented to indicate that combining multiple learning agents, relying on user feedback, is a promising direction to improve learning speed in automated WWW navigation.

Index Terms—Adaptive dynamic programming, multi-agent learning, relevance feedback, vector-space model, Web navigation.

I. INTRODUCTION

Since the World Wide Web (WWW) began as a wide-area hypermedia information sharing initiative aiming to give universal access to documents, it has become the world's largest collection of documents without any consistent organization other than the linked nature of documents. Navigating through the linked document space to access relevant documents without getting lost in the "maze" is indeed a daunting task for the average user. Hence, there is a clear need for automated personalized supporting agents that can automatically traverse through the Web of documents to retrieve documents of interest and relevance to a particular user.

Presently, different Web search engines can be used to locate and retrieve information on the Web. However, the problem with these search services is that they are rarely personalized to the user. Further, they locate information available on the Web based on keywords and leave a large number of hits to the user, making it impossible for the user to go through each and every document and their hyperlinks, searching for the relevant information.

J. Varghese and S. Mukhopadhyay are with the Indiana University –Purdue, IN 46202 USA (e-mail: smukhopa@cs.iupui.edu).