An Unsupervised Technique for Automatic Face Detection and Extraction

Muhammad Hameed Siddiqi
Ubiquitous Computing Lab, Kyung Hee University, Korea
siddiqi@oslab.khu.ac.kr

Rahman Ali
Ubiquitous Computing Lab, Kyung Hee University, Korea
rahmanali@oslab.khu.ac.kr

Ibrahiem M. M. El Emary
Dept. of Information Science, King Abdulaziz University, Jeddah, Saudi Arabia
omary57@hotmail.com

Sungyoun Lee
Ubiquitous Computing Lab, Kyung Hee University, Korea
sylee@oslab.khu.ac.kr

***Abstract*—Automatic face detection is the essential part of the facial expression recognition (FER) systems. Before investigating the facial expressions, it is compulsory to detect and extract the faces first from the expression frames. Existing methods often involve modeling of the face detection that normally necessitates huge amount of training data and cannot efficiently tackle changes over time. In this paper, an unsupervised technique based on active contour (AC) model is adopted in order to detect and extract the human faces automatically from the expression frames. In this model, the combination of two energy functions like Chan-Vese (CV) energy and Bhattacharyya distance functions were exploited that not only minimize the dissimilarities within the object (face) but also maximize the distance between the object (face) and background. The developed method is more robust to noise and illumination that are typical issues in FER systems. The proposed AC model is an unsupervised technique; means no training data is required. The developed approach achieved best results than of conventional CV AC model.**

Keywords-component; Facial expressions; face dectection; active contour; level set.

I. Introduction

Essentially, face detection is the first-step for automatic FER systems, with the purpose of localizing and extracting the face region from the expression frames. It has several applications such as content-based image retrieval, video coding, video conferencing, crowd surveillance, and intelligent human computer interfaces. The human face is a dynamic object and has a high degree of variability in its appearance, which makes face detection a difficult problem in computer vision. A wide variety of techniques have been proposed, ranging from simple edge-based algorithms to composite high-level approaches utilizing advanced pattern recognition methods.

Though, there lots of works have been done for face detection; however, most of them have their own limitations. In [1], the author exploited Neural Network (NN) for face detection that divided the frame in small windows in order to find the location of the face. However, the performance of NN is completely reliant on the number of hidden layers and nodes and learning rates, and also to get an optimum performance, the network has to be comprehensively adjusted [2]. A robust geometrical method has been proposed by [3] for face detection; however, this method does not has the capability to represent the global face structure.

The authors of [4–10] employed different appearance-based methods for face detection. However, appearance-based methods have some limitations, like these methods are very sensitive to scale, their accuracy rate decreases under varying pose, illumination and in complex environments. Also it is almost impossible for these methods to find an optimal way of projection that has the capability to instantaneously discriminate multiple face classes [11].

Some factors such as illumination, head pose and occlusion might reduce robustness of the FER systems; therefore, the authors of [12] and [13] employed the combination of skin-color and template matching methods in order to detect the face. However, still the skin-color methods have difficulties to get the faces robustly under varying lighting conditions and especially in the presence of complex background [14]. Moreover, template methods are correlation-based, which computational wise much expensive and also a huge amount of storage are required [15].

Therefore, the objective of this paper is to propose an unsupervised face detection AC model that automatically detects and extracts the human face from the facial frame. The proposed AC model based on level set and is the combination of two energy functions like Chan-Vese [16] energy and Bhattacharyya distance [17] functions. The proposed AC model is most robust to noise and illuminations, which not only minimizes the dissimilarities within the object (face) but also maximizes the distance between the two regions such as face and the background.

We already described some related work about this field. The rest of the paper is organized as follows. Section II delivers an overview of the proposed AC model. Section III provides some experimental results of the proposed model with some discussion on the results and a comparison with the conventional CV AC model. Finally, the paper will be concluded after some future directions in Section IV.

II. Material and Methods

Mostly, the accuracy of the FER systems reliant on the performance of automatic face detection. In the field of image segmentation, since it was first introduced by [18], active contour (AC) model has achieved much attention.

An AC model is a deformable spline influenced by constraint and image forces that pull it towards object contours. It tries to move into a position where its energy is minimized. Active contour tries to improve by imposing desirable properties such as continuity and smoothness to

978-1-4799-3197-2/14/$31.00 ©2014 IEEE

the contour of the object, which means that the active contour approach adds a certain degree of prior knowledge for dealing with problem of finding the object contour. Recently in [16], the Chan-Vese (CV) proposed a novel form of active contour for object segmentation based on level set framework. Unlike other active contour models which rely much on the gradient of the image as the stopping term and thus have unsatisfactory performance in noisy images, the CV active contour model does not use the edge information but utilizes the difference between the regions inside and outside of the curve, making itself one of the most robust and thus widely used techniques for image segmentation, especially, in the area of face detection. Its energy function is defined by

$$F(C) = \int_{in(c)} \left| I(X) - c_{in} \right|^2 dx + \int_{out(c)} \left| I(X) - c_{out} \right|^2 dx \quad (1)$$

where $x \in \Omega$ (the image plane) $\subset$R, $\Omega \rightarrow Z$ is a certain image feature such as intensity, color, or texture, and c_{in} and c_{out} are respectively the mean values of the image feature inside [in(c)] and outside [out(c)] the curve *C*. Considering the image segmentation as a clustering problem, we can see that this model forms two segments (clusters) such that the differences within every segment are minimized. However, the global minimum of the above energy functional does not always guarantee the desirable results. The unsatisfactory result of the CV AC in this case is due to the fact that it is trying to minimize the dissimilarity within each segment but does not take into account the distance between different segments. The proposed methodology is to incorporate an evolving term based on the Bhattacharyya distance to the CV energy functional that minimizes the dissimilarities within the object and maximizes the distance between the two regions. The proposed energy function is:

$$E_0(C) = \beta F(C) + (1-\beta) B(C) \quad (2)$$

where β∈[0,1], B(C) is the Bhattacharyya coefficient [17] with the Eq. 3 and 4.

$$f_1(x) = \frac{k_\sigma(x) * \left[H(\phi(x)) I(x) \right]}{k_\sigma(x) * \left[H(\phi(x)) \right]} \quad (3)$$

$$f_2(x) = \frac{k_\sigma(x) * \left[1 - H(\phi(x)) I(x) \right]}{k_\sigma(x) * \left[1 - H(\phi(x)) \right]} \quad (4)$$

where $f_1(x)$ and $f_2(x)$ are the local fitting functions [19] which depend on the level set function ϕ, and need to be updated in each contour evaluation and H(•) and δ'(•)≡ H′(•) respectively the Heaviside and Dirac functions [20]. Note that the Bhattacharyya distance is defined by *[–log B(C)]* and the maximization of this distance is equivalent to the minimization of *B(C)* that also to be comparable to the *F(C)* term, *B(C)* is multiplied by the area of the image because its value is always within the interval [0, 1] whereas *F(C)* is calculated based on the integral over the image plane. In general, we can regularize the solution by constraining the length of the curve and the area of the region inside it. Therefore, the energy functional is defined by

$$E(C) = \gamma \int_\Omega \left| \nabla H(\phi(x)) \right| dx + \eta \int_\Omega H(-\phi(x))\, dx \quad (5)$$

where $\gamma \geq 0$ and $\eta \geq 0$ are constants.

The intuition behind the proposed energy functional is that we seek for a curve which (1) is regular, i.e., the first two terms and (2) partitions the image into regions such that the differences within each region are minimized, i.e., the *F(C)* term and the distance between the two regions is maximized, i.e., the *B(C)* term. The level set implementation for the energy functional in 4 can be derived as.

$$\frac{\partial \phi}{\partial t} = |\nabla \phi| \left\{ \begin{array}{l} \gamma k + \eta + \beta \left[(I - c_{in})^2 + (I - c_{out})^2 \right] \\ -(1-\beta) \left[\begin{array}{c} \delta_0(z-1) \\ \frac{B}{2}\left(\frac{1}{A_{in}} - \frac{1}{A_{out}} \right) + \frac{1}{2} \int_z \left(\begin{array}{c} \frac{1}{A_{out}} \sqrt{\frac{f_1}{f_2}} \\ - \frac{1}{A_{in}} \sqrt{\frac{f_2}{f_1}} \end{array} \right) dz \end{array} \right] \end{array} \right\} \quad (6)$$

where A_{in} and A_{out} are respectively the areas inside and outside the curve *C*. Thus, the proposed AC model overcame the limitation of conventional CV AC model in the area of face detection.

III. Experimental Results and discussion

The accuracy of the FER systems are completely reliant on the automatic face detection that is considered one of the challenging tasks for the existing works. Therefore, we proposed an unsupervised and a robust AC model that is the combination of the two energy functions that detects and extracts the faces automatically from the expression frames. The proposed AC model has been tested and validated on publicly available standard dataset named Cohn-Kanade dataset [21] in order to assess its performance.

In the facial expression videos, the active contour evolution in a certain frame is performed independently of the other frames, means that the face detection in a video is done frame-based. The only utilized information is the final contour obtained in the previous frame which will be used to determine the initial position of the active contour in the current frame. First, an ellipse with major axis along y-axis of length 10 and minor axis along x-axis of length 10 is selected as the initial contour. In the experimental results of this paper, this initial shape will be same for all the expression frames, but only the center location varies. In each video, the first frame is segmented using manual initialization such that the initial contour is closer to the human face.

Then from the second frame, the position of the initial contour's center in the current frame is the mean value of the points along the final contour in the previous frame. Suppose, there are *M* points $(x_i^{(n)}, y_i^{(n)})$, i=1…M along the final contour of the facial frame *n(n≥1)*. Then, the center

$(c_x^{(n+1)}, c_y^{(n+1)})$ of the initial contour in the frame *(n+1)* is calculated as

$$\left(c_x^{(n+1)}\right) = \frac{1}{M}\sum_{i=1}^{M} x_i^{(n)}, \left(c_y^{(n+1)}\right) = \frac{1}{M}\sum_{i=1}^{M} y_i^{(n)}$$

Some experimental results of the proposed AC model with the above scheme on six different types of facial expressions are given in Fig. 1. Also, some sample results of the CV AC model are represented in Fig. 2.

(a) **(b)** **(c)**

Fig. 1. Sample results of the proposed AC model (a) represents the initial contour (b) shows the final contour and (c) presents the extracted face.

Fig. 2. Sample results of CV AC model (a) represents the initial contour (b) shows the final contour and (c) presents the extracted face.

IV. CONCLUSION

Before investigating the facial expressions, automatic face detection is the essential part of the FER systems. In this research, we have proposed an active contour model based on level set that is the combination of two energy functions (1) Chan-Vese energy function that has been exploited in order to minimize the dissimilarities within the human face, and (2) Bhattacharyya distance function that is used to maximize the distance between the face and background. At first step, the initial contour should be closer to the human face. The proposed AC model detects and extracts the face from the current frame by employing this initial contour and the final contour of this frame is shifted to the next frame. Then, the mass center of the shifted contour is used as the center of the initial contour in the next frame and by this way, the process will be repeated till the last frame of the expression video. As a result, the proposed AC model with this initialization scheme can correctly detects human face from the expression videos. The overall results of the proposed AC model and the conventional CV AC are shown in Fig. 1 and 2 respectively.

In this paper, we presented the qualitative results of the face detection based on the proposed model. In next step, we will incorporate this model with feature extraction and recognition models in order to present the quantitative results in-terms of classification for the typical FER systems.

ACKNOWLEDGMENT

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (No. 2011–0030823.

REFERENCES

[1] H. A. Rowley, S. Baluja, and T. Kanade, "Neural network-based face detection," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 20, no. 1, pp. 23–38, 1998.

[2] P. Kakumanu, S. Makrogiannis, and N. Bourbakis, "A survey of skin-color modeling and detection methods," *Pattern recognition*, vol. 40, no. 3, pp. 1106–1122, 2007.

[3] O. Rezq and A. El-Sayed, "Geometrical approach of face detection and recognition," 2011.

[4] Z. S. Tabatabaie, R. W. Rahmat, N. I. B. Udzir, and E. Kheirkhah, "A hybrid face detection system using combination of appearance-based and feature-based methods," *International Journal of Computer Science and Network Security*, vol. 9, no. 5, pp. 181–185, 2009.

[5] S. Singh, M. Sharma, and N. S. Rao, "Accurate face recognition using pca and lda," *International Journal of Management, IT and Engineering*, vol. 2, no. 4, pp. 101–121, 2011.

[6] N. Nallammal and V. Radha, "Performance evaluation of face recognition based on pca, lda, ica and hidden markov model," *in Data Engineering and Management*, pp. 96–100, Springer, 2012.

[7] C. A. Waring and X. Liu, "Face detection using spectral histograms and svms," *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, vol. 35, no. 3, pp. 467–476, 2005.

[8] J. Olivares-Mercado, G. Aguilar-Torres, K. Toscano-Medina, M. Nakano-Miyatake, and H. Perez-Meana, "Gmm vs svm for face recognition and face verification," Reviws, Refinements and New Ideas in Face Recognition, 29–48,2011.

[9] G. A. Ramirez and O. Fuentes, "Face detection using combinations of classifiers," *in Computer and Robot Vision, 2005. Proceedings. The 2nd Canadian Conference on*, pp. 610–615, IEEE, 2005.

[10] G. Mahalingam and C. Kambhamettu, "Video based face recognition using graph matching," *in Computer Vision–ACCV 2010*, pp. 82–94, Springer, 2011.

[11] "Advantages and limitations." http://www.scourge.fr/mathdesc/documents/facerecog/AdvantagesLimitations.htm, 1999. (Last visited in May 2013).

[12] S. Tripathi, V. Sharma, and S. Sharma, "Face detection using combined skin color detector and template matching method," *International Journal of Computer Applications*, vol. 26, no. 7, 2011.

[13] J. Song, B. Chen, Z. Chi, X. Qiu, and W. Wang, "Face recognition based on binary template matching," *in Advanced Intelligent Computing Theories and Applications. With Aspects of Theoretical and Methodological Issues*, pp. 1131–1139, Springer, 2007.

[14] R.-L. Hsu, M. Abdel-Mottaleb, and A. K. Jain, "Face detection in color images," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 24, no. 5, pp. 696–706, 2002.

[15] C. Garcia and G. Tziritas, "Face detection using quantized skin color regions merging and wavelet packet analysis," *Multimedia, IEEE Transactions on*, vol. 1, no. 3, pp. 264–277, 1999.

[16] T. F. Chan and L. A. Vese, "Active contours without edges," *Image Processing, IEEE Transactions on*, vol. 10, no. 2, pp. 266–277, 2001.

[17] T. Kailath, "The divergence and bhattacharyya distance measures in signal selection," *Communication Technology, IEEE Transactions on*, vol. 15, no. 1, pp. 52–60, 1967.

[18] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *International journal of computer vision*, vol. 1, no. 4, pp. 321–331, 1988.

[19] L. He, W. G. Wee, S. Zheng, and L. Wang, "A level set model without initial contour," *in Applications of Computer Vision (WACV), 2009 Workshop on*, pp. 1–6, IEEE, 2009.

[20] D. Feltell, L. Bai, and R. Soar, "Level set brain segmentation with agent clustering for initialisation," *in International Conference on Bio-inspired Systems and Signal Processing (BIOSIGNALS), Funchal, Madeira, Portuga*l, 2008.

[21] T. Kanade, J. F. Cohn, and Y. Tian, "Comprehensive database for facial expression analysis," *in Automatic Face and Gesture Recognition, 2000. Proceedings. 4th IEEE International Conference on*, pp. 46–53, IEEE, 2000.