Available online at www.sciencedirect.com
ScienceDirect
Pattern Recognition Letters 28 (2007) 64–74
Pattern Recognition Letters
www.elsevier.com/locate/patrec

Facial boundary detection with an active contour model

Jae Sik Chang [a], Eun Yi Kim [b,*], Hang Joon Kim [c]

[a] *Computer Science Department, University of California, Santa Barbara, CA, USA*
[b] *Department of Internet and Multimedia Engineering, Konkuk University, 1 Hwayang-dong, Gwangjin-gu, 143-750 Seoul, South Korea*
[c] *Department of Computer Engineering, Kyungpook National University, Daegu, South Korea*

Received 11 May 2004; received in revised form 8 April 2006
Available online 23 August 2006

Communicated by T.K. Ho

Abstract

This paper presents an active contour model for extracting accurate facial regions in complex environments. In the model, a contour is represented by a zero level set of level function ϕ, and evolved via level set partial differential equations. Then, unlike general active contours, skin-color information that is represented by 2D-Gaussian model is used for evolving and stopping a curve, which allows the proposed method to be robust to noise and varying pose. To assess the effectiveness of the proposed method, it was tested with several natural scenes, and the results were compared with those of geodesic active contours. Experimental results demonstrate the superior performance of the proposed method.

Keywords: Facial boundary detection; Active contour model; Level sets; Skin color model; Markov random field

1. Introduction

With the rapid development of ubiquitous computing environments, there is a growing need for human computer interaction system. Consequently, automatic recognition systems for face or facial expression have become an important research area because face plays an important role in man–machine communication. For accurate human computer interaction, face or facial expression recognition systems should ideally work well even when varying appearance due to changing pose and the relatively low resolution of images at a distance. Then to overcome these problems, an accurate facial boundary can be used as important information. For example, pose-invariant face recognition algorithms using 3D face models have been developed (Hsu and Jain, 2001). However, it is difficult to align 3D face structure with 2D images and acquire 3D face shapes (Hsu and Jain, 2003). Facial boundaries provide sufficient information to reconstruct 3D face structure with various face pose. And in case of facial expression recognition, at present, the recognition of facial expression mainly focuses on studying the facial motion from image sequence. However, in many multimedia and human computer interaction applications, image sequences of detailed facial actions are not available but a static image. To recognize facial expression in static image, sketches or caricatures based methods have been proposed (Gao et al., 2003). And to generate the sketches or caricatures, facial boundary detection is very important task.

So far many techniques and algorithms for face detection have been proposed, then they can be roughly classified into four categories (Yang et al., 2002): knowledge-based approaches, feature invariant approaches, template-matching approaches, and appearance-based approaches. Among these approaches, template-matching approaches have been widely used due to their simplicity. This approach stores several templates describing the face, facial features, and their relations. When given an input image, it computes

* Corresponding author. Tel.: +82 2 450 4135; fax: +82 2 450 4072.
E-mail addresses: tani12@hanmail.net (J.S. Chang), eykim@konkuk.ac.kr (E.Y. Kim), hjkim@ailab.knu.ac.kr (H.J. Kim).

0167-8655/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.patrec.2006.06.020

the correlation values between the input image and the templates to detect facial regions (Yang et al., 2002). The approach has been proved to be very efficient in detecting frontal facial images; however, it cannot work well when no constraints are imposed on the scene content due to the rigidity of templates. To overcome this limitation, deformable templates were used. The method deforms the predefined templates and matches the deformed template with a given image. These operations are iteratively executed until the best fit for the template is found (Yang et al., 2002). Although these methods perform well, they are still hampered by noise sensitivity and intensive computational costs.

Recently, active contour models have been used as a good solution to object detection because of their elasticity. An active contour model is a description of an object boundary which is iteratively adjusted until it matches the object of interest (Fenster and Kender, 2001). Active contours are successfully used for object detection, but also for track non-rigid objects (Kass et al., 1988; Caselles et al., 1993; Chan and Vese, 2001).

The initial active contour model was proposed by Kass et al. (1988), which is called SNAKES. They used maker particles to describe contours, so it cannot accurately detect the boundaries of non-rigid objects and provide automatic topological changes. As an alternative to this, a geometric active contour model based on a level set method was proposed by Osher et al. (Caselles et al., 1993; Osher and Sethian, 1988). They used the image gradient to stop the curve, which is represented by a zero level set. Then, a geometric active contour model is described by the following differential equation:

$$\frac{\partial \phi}{\partial t} = g(|\nabla I|) \cdot |\nabla \phi| \cdot (k + v), \tag{1}$$

where

$$k = \mathrm{div}\left(\frac{\nabla \phi}{|\nabla \phi|}\right) = \frac{\phi_{xx}\phi_y^2 - 2\phi_x\phi_y\phi_{xy} + \phi_{yy}\phi_x^2}{(\phi_x^2 + \phi_y^2)^{3/2}}. \tag{2}$$

In Eq. (1), v is constant, and I is the intensity image. Finally, $g(x) = (1 + x^2)^{-1}$ is the edge stopping function. This model acts as an edge detector. That is, an initial curve around the object to be detected evolves in its normal direction with speed $g(|\nabla I|) \cdot (k + v)$ and stops on the boundary that has large gradient values (Caselles et al., 1993; Chan and Vese, 2001). However, an object boundary has a low gradient in natural scenes. In noisy images, a noisy region may have the higher gradient value than an object region. Therefore, this model fails to detect accurate object boundaries.

To solve this problem, Chan and Vese proposed the use of intensity information of regions instead of the gradient information to stop the curve (Chan and Vese, 2001). They assumed that an image I consists of two regions with approximately constant intensities, I^0 and I^1, and that the object to be detected is represented by the region with the value I^1. Then, the curve is described as follows:

$$\frac{\partial \phi}{\partial t} = \delta(\phi) \cdot [\mu k - v - \underbrace{\lambda_1 (I - c_1)^2}_{(A)} + \underbrace{\lambda_2 (I - c_2)^2}_{(B)}], \tag{3}$$

where k is curvature, and v, μ, λ_1, and λ_2 are constants. c_1 and c_2 are means of intensities of inside of object and outside of object, respectively. In Eq. (3), terms (A) and (B) are squares of intensity differences between the pixel (x, y) and the mean of the inside and outside regions, respectively. The curve moves in the normal direction with speed (B) − (A) until finding the desired boundary where (A) is equal to (B). In summary, this curve evolves until reaching the minimization of the sum of the intensity deviations. This model is useful to detect contours with weak boundaries or discontinuous boundaries (Chan and Vese, 2001). However, the application is restricted to a simple image that consists of two regions with approximately constant intensities.

This paper investigates the application of active contours to facial boundary detection problems. Facial boundary detection is formulated as an energy minimization problem leading to a solution via an active contour model in a Bayesian framework. Our contour model is represented and evolved using a level set method, which allows the automatic topological changes. Additionally, we use skin-color information to stop the curves, which is represented by a 2D-Gaussian model. The use of skin-color information endows the proposed method for robustness to noise as well as independence of the initial values.

The main advantages of the proposed method include robustness to noise, multiple views, and changing pose. Additionally, the proposed method can easily expand to face detection in video sequence, since it incorporates tracking mechanisms within its own architecture. These advantages are proven by experimental results that show the successful application of the proposed method to detect facial boundary in complex environments.

The rest of this paper is organized as follows. In Section 2, a skin-color model is described. Section 3 illustrates how to formulate a facial boundary detection problem as an energy minimization problem, and the minimization scheme is shown in Section 4. Experimental results are presented in Section 5. Finally, conclusions are presented in Section 6.

2. Skin-color model

Every person has different skin tones. Moreover, under various lighting conditions, the skin tones of a single person can be observed differently. This is because they differ much less in color than in brightness (Yang and Waibel, 1996). In the RGB space, color representation includes both color and brightness. Therefore, RGB is not necessarily the best color representation for detecting pixels with skin color. Brightness can be removed by dividing the three components of a color pixel (R, G, B) according to intensity. This space is known as chromatic color, where intensity is a normalized color vector with two components (r, g):