OPTI-GVF SNAKE MODEL FOR FACE SEGMENTATION FROM VIDEO SEQUENCES

P. K. Biswas[1] and Madhukar Pandit
Universität Kaiserslautern
Elektrotechnik, Energie- und Informationstechnik
Erwin-Schrödinger-Str. 12
67663 Kaiserslautern
Germany

Abstract

Face segmentation is one of the most important tasks in Model Based Video coding for Video Telephony and Video Conferencing applications. Since the face movement is usually different from the back ground movement, gradient of the optical flow field gives useful hints for the region of the image frame containing human face. In this work we have used a modified GVF-Snake (Gradient Vector Flow-Snake), where Gradient Vector Flow of the optical flow field is used as a component of the energy function to be minimized, to segment out the face from video sequence. GVF-Snake has a particular property that it can capture concave boundaries. Experiments with a number of test sequences give encouraging results.

Key Words

Video Segmentation, Optical Flow, GVF-Snake, Tracking.

1. INTRODUCTION

In this age of multimedia and computer networks, transmission of video over long distances has assumed great importance. Video is an integral part of multimedia, video conferencing, video telephony, video-on-demand, broadcast digital video etc. In the video telephony applications, the requirement is to transmit the video signal over the available Public Switched Telephone Network (PSTN). Typically the available channel capacity in PSTN is 64 Kbps. The two standard digital picture formats available for low bit rate video are CIF and QCIF. Transmission of CIF standard video at 10 frames per second requires a channel bandwidth of 8 Mbps, whereas that for QCIF standard video is 2 Mbps. This clearly demands high compression of video signal for video telephony applications.

The model based video coding scheme suggested for the video telephony application is an approach that provides high image quality despite low bit rate transmission. For this kind of application it is the face of the speaker which is of importance and not the back ground. The encoder analyses the face image and the decoder synthesizes it based on the analysis output and 3-D shape model of a face.

Face segmentation from video sequence is an essential component in model based coding to enable the encoder to properly analyze and the decoder to synthesize the human face. Segmentation of a region of interest in video is different from segmentation of a single image. Video segmentation based on intensity and motion was reported by Thompson [1]

[1] On sabbatical from Deptt. of E & ECE, Indian Institute of Technology, Kharagpur- 721 302, INDIA

in 1980. Segmentation is done by finding 4-connected regions that have similar gray-scale and optical flow values. Darrel and Pentland [2] proposed an iterative method to segment images into number of layers. Black [3] combined intensity and motion for segmentation of image sequences. Tekalp et. al. [4] proposed combining color and motion for segmentation purpose. Patras, Hendriks and Lagendijk [5] used watershed segmentation, motion estimation and labeling for video segmentation. Color, motion and spatial information have been used by Khan and Shah [6] for object based video segmentation.

In this paper we have presented an optical flow based GVF-Snake model for face segmentation from video sequences. There are two key difficulties with parametric snake algorithms. Firstly, the initial contour must, in general, be close to the true boundary. Secondly, they have difficulties progressing into boundary concavities. Both these problems are addressed in GVF-Snake model[9]. We have adopted Horn and Schunk [7] technique to determine optical flow. Both the magnitude and phase information of optical flow field are combined to construct velocity frame. Our basic assumption is that the face movement is different from the back ground movement, so the gradient of the velocity frame should indicate motion boundary. We have incorporated this concept in the GVF-Snake model which is a modification to Kass et. al. [8]. Gradient Vector Flow Field of the velocity frame has been added in the energy function that is to be minimized for the active contour to lock on to the object boundary.

The paper is organized as follows. In section 2 we present optical flow computation and velocity frame construction. Gradient Vector Flow is presented in section 3. OPTI-GVF Snake is discussed in section 4 followed by results and conclusions in sections 5 and 6 respectively.

2. OPTICAL FLOW AND VELOCITY FRAME

For calculation of Optical Flow, we have used the Horn and Schunck [7] method.

Let the image brightness at a point *(x,y)* in the image plane at time t be denoted by *E(x,y,t)*. Assuming that the brightness of a particular point is constant, we get

$$\frac{dE}{dt} = 0$$

Considering a patch of the brightness pattern that is displaced by a distance δx in the x-direction and δy in the y-direction in time δt and assuming brightness of the patch to remain constant

$$E(x,y,t)=E(x+\delta x, y+\delta y, t+\delta t)$$

$$\frac{\partial E}{\partial x}\cdot\frac{dx}{dt}+\frac{\partial E}{\partial y}\cdot\frac{dy}{dt}+\frac{\partial E}{\partial t}=0$$

Let $u=\frac{dx}{dt}$ and $v=\frac{dy}{dt}$ be the optical velocity components. Then,

$$E_x u + E_y v + E_t = 0$$

Where E_x, E_y, E_t are gradients w.r.t *x,y,t* respectively.

To compute flow velocity *(u,v)* the following smoothnes constraint is introduced.

One way to express the additional constraint is to limit the difference between the flow velocity (u,v) at a point and the average velocity ($\bar{u}, \bar{v}$) over a small neighborhood containing that point.

The problem is now to minimize the sum of the errors in the equation for the rate of change of image brightness,

$$\delta_b = E_x u + E_y v + E_t$$

and the estimate of departure from the smoothness in the velocity flow,

$$\delta_c = (\bar{u} - u)^2 + (\bar{v} - v)^2$$

Then the total error is,

$$\delta = \alpha\delta_c + \delta_b$$

where α is a weight factor proportional to noise in the measurement. Ideally δ_c, δ_b should be zero.

Solving for $\delta_{\min}$, we get

$$u = \frac{((\alpha^2 + E^2{}_y)\bar{u} - E_x E_y \bar{v} - E_x E_t)}{(\alpha^2 + E^2{}_x + E^2{}_y)}$$

$$v = \frac{((\alpha^2 + E^2{}_x)\bar{v} - E_x E_y \bar{u} - E_y E_t)}{(\alpha^2 + E^2{}_x + E^2{}_y)}$$

We can compute a new set of velocity estimates from the estimated derivatives and the average of the previous velocity estimates by,

$$u^{n+1} = u^n - E_x . \frac{[E^2{}_x . \overline{u^n} + E^2{}_y \overline{v^n} + E_t]}{[\alpha^2 + E^2{}_x + E^2{}_y]}$$

$$v^{n+1} = v^n - E_y . \frac{[E^2{}_x . \overline{u^n} + E^2{}_y \overline{v^n} + E_t]}{[\alpha^2 + E^2{}_x + E^2{}_y]}$$

The Optical Flow was calculated for all the frames of the video sequence. Our basic assumption is that even if there is background movement, the optical flow of foreground pixels is appreciably different from the optical flow of the background pixels. Hence to detect movement in the frame accurately, we used the scaled magnitude (scaled w.r.t to maximum magnitude ($r_{\max}$), where $r = \sqrt{u^2 + v^2}$) and scaled phase to construct a velocity frame F that gives us an idea of where the motion is taking place.

$$F = [w_1 . (\frac{r}{r_{\max}}) + w_2 . (\frac{\phi}{2\Pi})].255$$

where $r/r_{\max}$ is normalized magnitude of flow vector and $\Phi = \tan^{-1}(\frac{v}{u})$ is the phase. w_1 and w_2 control the weights of the magnitude and phase factors.

3. GRADIENT VECTOR FLOW

Gradient Vector Flow (GVF) method applied to snakes [9] is a unique method that deals with concavity problem. To solve this problem there must be an external force acting on the snake that will pull the snake to concave object boundary. Suppose f(x,y) is the edge map of an image I(x,y). Then the gradient of the edge map ∇f has vectors pointing towards the edges and these are normals to the edges on edge points. When this gradient force ∇f acts on the snake contour, it pulls the contour towards the concave boundary. However these gradient vectors have large magnitude only in the immediate vicinity of the boundary and are nearly zero at points away from the boundary. So the capture range of the snake will be very small. To increase the capture range the gradient map is extended to points away from the edges using a computational diffusion process. Gradient Vector Flow field is a vector field V(x,y)=[u(x,y),v(x,y)] that minimizes the energy functional as given below.

$$E = \iint (\mu(u_x^2 + u_y^2 + v_x^2 + v_y^2) + |\nabla f|^2 |V - \nabla f|^2). \; dxdy$$

When $|\nabla f|$ is small the energy is dominated by sum of the squares of the partial derivatives of the vector field, yielding a slowly varying field. On the other hand, when $|\nabla f|$ is large, the second term dominates and the energy function is minimized by setting $V = \nabla f$. The parameter μ is a regularization parameter governing the tradeoff between the first term and the second term.

Using calculus of variations, it can be shown that the GVF field can be found by solving the following Euler equations.

$$\mu\nabla^2 u - (u - f_x)(f_x^2 + f_y^2) = 0$$

$$\mu\nabla^2 v - (v - f_y)(f_x^2 + f_y^2) = 0$$

Detailed formulation and iterative solution can be found in [9].

4. OPTI-GVF SNAKE MODEL

For segmentation of area of interest we have used the snake model given by Kass et. al. [8] with a modification in the energy function that includes GVF of velocity frame. A snake is a set of control points represented by a vector, v(s)=(x(s),y(s)), having the arc length, s, as the parameter. Kass et. al. [8] defined an energy function of the contour and described a method for finding contours which correspond to local minima of the energy function. The energy function is written as

$$E = \int_0^1 E_{snake}(v(s))ds$$

$$= \int_0^1 \{E_{\text{int}}(v(s)) + E_{image}(v(s))\}ds$$

E_{int} represents the internal force of the contour due to bending or discontinuities and E_{image} is the image force. Thus the total energy of the snake can be represented as

$$E_{total} = \int (\alpha(s)E_{continuity} + \beta(s)E_{curvature} + \gamma(s)E_{image})ds$$

where $\alpha(s), \beta(s), \gamma(s)$ are the weight factors for the energy terms.

The continuity energy is given by

$$E_{continuity} = \sum | v_i - v_{i-1} |^2$$

Where v_i is a point on the contour called control point. The greater the distance between the neighboring control points, the greater is the continuity energy of the snake. Hence to minimize the continuity energy, the snake tends to bring all its control points as close to each other as possible.

To avoid control points from bunching we introduced an equidistance energy as given below.

$$E_{equi-dist} = \sum \| v_i - v_{i-1} | - | v_i - v_{i+1} \|^2$$

The equidistance energy maintains equal distance among the control points.

We have not used curvature energy in our formulation as this always tries to smooth the boundary of the contour and does not allow the snake to capture the concave boundaries.

The image energy E_{image} depends upon the brightness level at that pixel of the image. The greater the brightness of the pixel, the lesser is the image energy at that pixel (because of the negative weight $\gamma(s)$ factor used). E_{image} is computed on the edge image so that snake locks onto the target boundary.

This simple formulation of image energy E_{image} may lock the snake onto a background object which is not of our interest. To overcome this problem we have introduced an additional term in E_{total} computed from the Gradient Vector Flow of optical flow field in addition to gradient vector flow of the original image. Thus the modified energy function is

$$E = \int (\alpha(s)E_{cont} + \beta(s)E_{equi-dist} + \gamma(s)E_{gvf} + \xi(s)E_{opti-gvf})ds$$

$E_{opti-gvf}$ is computed from edge image of F. Again $\alpha, \beta, \gamma, \xi$ are used to balance the relative influence of the four energy terms.

The snake is initialized to a random shape large enough to engulf the entire area of interest. Then the snake is allowed to slither. For slithering energy is computed at all pixels in the 3x3 neighborhood of each control point. The control point is shifted to the pixel location with minimum energy.

5. RESULTS

The modified active contour model has been applied on a number of QCIF format (176x144) video sequences, namely Miss America, Akiyo, Grandma, containing head and shoulder image. The face segmentation result on all these sequences has been quite satisfactory. Example result with one of the sequence namely Grandma is being presented here.

Fig. 1(a)-(b) shows the 0^{th} and 1^{st} frames of the Akiyo sequence respectively. Fig. 1(c) shows the needle diagram of the optical flow field computed from these two frames and Fig. 1(d) shows the corresponding velocity frame F.

Fig. 1(a)

Fig. 1(b)

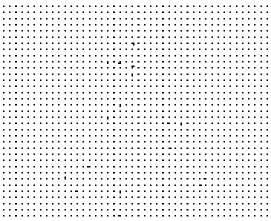

Fig. 1(c)

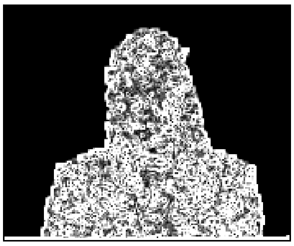

Fig. 1(d)

It may be noted that though the optical flow magnitude is not very prominent, as indicated in Fig. 1(c), but when we combine the magnitude and phase information of the optical flow field the resulting velocity frame F clearly indicates image region with motion (Fig. 1(d)).
Fig. 2(a) shows the 1^{st} to 3^{rd} frames of Grandma sequence. Fig. 2(b) and 2(c) show the gradient and

Fig. 2(a)

Fig. 2(b)

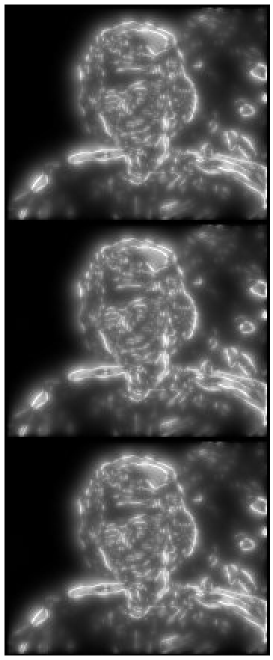

Fig. 2(c)

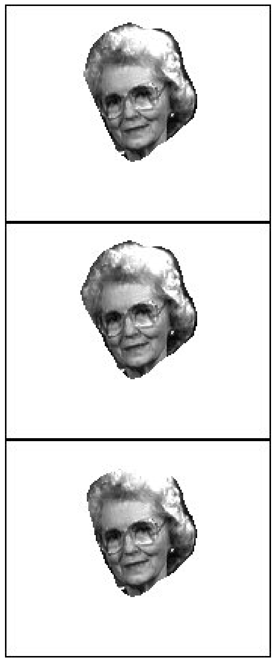

Fig. 2(d)

the GVF frames respectively. Segmented frames after 50 iterations are shown in Fig. 2(d). The parameter values in the energy function are $\alpha = 5, \beta = 12 \ \ \gamma = -3$ and $\xi = -6$.

For face segmentation in subsequent frames the following approach is taken. The locked snake contour on the n-th frame is inflated outwards by an amount more than the maximum optical flow magnitude to act as the initial contour for n+1st frame. This initial contour is then slithered following the same technique.
Experiments with other video sequences also give satisfactory results.

6. CONCLUSION

In this paper we have presented a modified GVF snake technique for face segmentation from video sequences containing head and shoulder image. Face segmentation is an important problem in model based video coding for video telephony and video conferencing applications. Our proposed method will give satisfactory result even when background is not stationary or the back ground contains other compact objects. The snake inflation technique enables face segmentation in subsequent frames.

The initial contour has to be selected very carefully. If the initial contour does not engulf the entire region of interest then the segmentation will fail. Further development is needed to take care of this problem. Such modification will also make this algorithm suitable for object segmentation meant for object based coding as in MPEG-4.

7. ACKNOWLEDGEMENT

The first author gratefully acknowledges the support provided by Alexander von Humboldt-Stiftung, Germany to carry out this research.

REFERENCES

[1] Willium B. Thompson, Combining motion and contrast for segmentation, *IEEE Trans. Pattern Anal. Machine Intelligence,* Nov. 1980, 543-549.

[2] Alex Pentland, A Darrel, Cooperative robust estimation using layers of support, *T.R 163,* MIT Media Lab, Vision and Modelling Group, Feb. 1991.

[3] Michael Black, Combining intensity and motion for incremental segmentation and tracking over long image sequences, *ECCV,* 1992.

[4] A. Murat Tekalp, Yucel Altunbasak, P Erhan Eren, Region based parametric motion segmentation using color information, *Journal of Graphical Models and Image Processing, 60*(1), 1998, 13-23.

[5] I. Patras, E. A. Hendriks, R. I. Lagendijk, Video segmentation by MAP labeling of Watershed Segments, *IEEE Trans. Pattern Anal. And Machine Intelligence, 23*(3), 2001, 326-332.

[6] Sohaib Khan, Mubarak Shah, Object based segmentation of video using color, motion and spatial information, *Proc. IEEE Intl. Conf. On Computer Vision and Pattern Recognition,* Dec. 2001.
http://www.cs.ucf.edu

[7] B. K. P. Horn, B. G. Schunk, Determining optical flow, *Arrificial Intelligence, 7,* 1981, 185-203.

[8] Michael Kass, Andrew Witkin, Demetri Terzopoulos, Snakes: Active contour models, *Proc. First Intl. Conf. On Computer Vision,* London, 1987, 259-269.

[9] Chenyang Xu, Jerry L. Prince, Snakes, Shapes and Gradient Vector Flow, *IEEE Trans. On Image Processing,* Vol. 7, No. 3, 1998, 359-369.