FAST MARCHING METHODS APPLIED TO FACE LOCATION IN VIDEOPHONE APPLICATIONS USING COLOUR INFORMATION

Prag Sharma and Richard Reilly
Department of Electrical and Electronic Engineering,
University College Dublin, Ireland.
prag@ee.ucd.ie, richard.reilly@ucd.ie

ABSTRACT

A new method is proposed to automatically segment out a person's face from a given sequence of images that consists of a head-and-shoulder view, using the Fast Marching Level Set approach. The method proposed here involves a fast, reliable and computationally efficient algorithm, which exploits the colour information in an image to segment the face region. The colour information is derived and used to classify pixels in the image into different levels depending on the value of the colour information for that pixel. Once this is achieved, a Priority-Labelled Fast Marching Level Set Algorithm is utilised to locate the face region. The performance of the Fast Marching face-segmentation algorithm is illustrated by some simulation results carried out on head-and-shoulder test images.

1. INTRODUCTION

The emergence of the MPEG-4 standard has brought the area of video segmentation and contour tracking to the forefront of research [1], [2]. Mobile technology and video conferencing type multimedia applications usually generate and display head-and-shoulder type images. In these images the face of the person is by far the most important part of the image as it is generally of the greatest interest to the viewers. Thus, by coding the facial region in a videophone sequence at a higher quality, the subjective quality of very low-bit-rate encoded videophone sequences can be improved [3], [4]. This is our main intuition behind locating and tracking the face region in a video sequence. Moreover, segmented human faces in video sequences can be used for face recognition, image enhancement, three-dimensional (3-D) human face model fitting and content-based representation to name just a few applications. Many of the approaches proposed to segment human faces involve the use of shape, motion, gradient, texture, colour and many forms of statistical analysis [5]. However, most of these approaches are computationally expensive and are suitable for off-line purposes. Our approach is aimed at a computationally efficient algorithm for on-line applications.

2. AIM

In this paper, attention is focussed on segmenting a person's face in a video sequence as the semantic object of interest using a Priority-Labelled Fast Marching approach to locate the skin-coloured pixels in the image. The aim will be to detect and then locate the face area as quickly as possible for real-time applications.

3. THE FAST MARCHING METHOD

The problem of locating and tracking an object of arbitrary shape is an old one and only recently have Level Sets been applied to images and video sequences [6]-[7]. The 'Level Set Methods' are numerical techniques for computing the position of propagating fronts [8]. Consider a closed curve moving in the plane, that is, let $\gamma(0)$ be a smooth, closed initial curve in Euclidean plane R^2, and let $\gamma(t)$ be the one-parameter family of curves generated by moving $\gamma(0)$ along its normal vector field with speed $F(k)$, a scalar function of the curvature k. The central idea in the level set approach of Osher and Sethian [8] is to represent this front $\gamma(t)$ as the zero level set $\{\Psi = 0\}$ of a higher dimensional function Ψ. In the general case, let $\gamma(0)$ be a closed, nonintersecting, $(N-1)$ dimensional hyper surface. Let $\Psi(x,t), x \in R^N$, be the scalar function such that $\Psi(x,0) = \pm d$, where $d(x)$ is the signed distance from x to the hyper surface $\gamma(0)$, where d is positive if x is outside γ. The goal is to produce an equation for the evolving function $\Psi(x,t)$, which contains the embedded motion of $\gamma(t)$ as the zero level set $(\Psi = 0)$. Letting $x(t)$ be the trajectory of a particle located on some level set $(\Psi = C)$, such that $\Psi(x(t),t) = C$, an evolution equation for Ψ can be generated for a given value of $\Psi(x, t = 0)$.

$$\Psi t + F|\nabla\Psi| = 0 \quad (1)$$

Equation (1) is in the form of a Hamilton-Jacobi type equation. This equation can be solved using the stable, entropy-satisfying finite difference schemes, borrowed from the literature on hyperbolic conservation laws and is termed the Level Set Equation.

A special case of such methods leads to the Fast Marching Level Set Method [9]. The Fast Marching Method (FMM) solves the general static Hamilton-Jacobi equation, which applies in the case of a convex, non-negative speed function $F > 0$. Imagine the two-dimensional case in which the interface is a propagating curve γ, and suppose we graph the evolving zero level set above the xy plane. That is let $T(x,y)$ be the time at which the curve crosses the point (x,y). The surface $T(x,y)$ then satisfies the equation

$$|\nabla T|F = 1 \quad (2)$$

0-7803-7304-9/02/$17.00 C2002 IEEE

This is a form of the well-known Eikonal equation, which converts the problem to a stationary formulation, because the front crosses each grid point only once. Equation (2) describes a basic algorithm that will be used to propagate a 2D front over an image.

3.1 Image Segmentation Using Priority-labelled FMM Approach

The FMM method propagates through the image depending on the value of the speed function F at each pixel in the image. For a good segmentation of the desired object, a speed function exists that is high in areas that do not belong to the face region but slows down near the edges of the face region and completely stops in areas where there is a high likelihood of that area being part of the face region. To achieve such behaviour, a priority level approach can be used.

3.2 Stopping the Level Set at the Right Priority Level

Ideally the Level Set should stop moving when it approaches the surface of the desired object. This can be achieved by reducing the speed function to zero at the object boundary i.e. at the right priority level. Equation (3) shows piece-wise linear stopping speed function $h(F(x,y))$ where $I(x,y)$ is each pixel in the image.

$$h(F(x,y)) = \begin{cases} 0 & \text{if } I(x,y) \in \text{Priority4} \\ 0.33 & \text{if } I(x,y) \in \text{Priority3} \\ 0.66 & \text{if } I(x,y) \in \text{Priority2} \\ 1 & \text{if } I(x,y) \in \text{Priority1} \end{cases} \quad (3)$$

However, most images do not have a clear distinction between the object of interest and the background. Due to the presence of varying lighting conditions, shadows or simply due to the incapability of the chosen image feature to identify the face region in its entirety, the Zero Level Set can over-segment the area of interest. To overcome this problem, the overall movement of the Fast Marching Method needs to be quantified.

To do this, the position of each pixel in the zero set after every 50 iterations was considered and the following thresholds applied:

- The percentage of zero level set pixels lying in priority level 4 were calculated and compared to a threshold value. If above this threshold value, the overall zero set was considered to be lying at sharp edges and the iterations were stopped.

This approach was found to be very successful in stopping the level set at the desired areas of interest.

4. SEGMENTATION USING COLOUR INFORMATION

Colour information is another low-level feature that has been used effectively in recent years to address the face-locating problem. Several recent publications that have reported this study include [10],[11]. A comprehensive survey on the use of colour information for detecting faces can also be found at [5]. These studies have shown that colour can prove to be a powerful feature in locating the face region in an image.

4.1 Choosing a Colour Space

An image can be represented in a number of different colour space models. In this paper, the YCbCr colour space was chosen for face segmentation after careful consideration. Firstly, the chrominance information can be used effectively for modelling human skin-colour since different skin complexions (white, yellow, brown, black) lie in a narrow range of the Cb and Cr values. This is due to the fact that the apparent difference in skin colour that viewers perceive is mainly due to the darkness or fairness of the skin which is characterized by the difference in the brightness of the colour, which is governed by the Y but not the Cb and Cr values. Since, the brightness is already separated from the chrominance values to a certain degree in YCbCr and not in some other colour spaces, such as RGB, the YCbCr colour space automatically stands out as the right choice. Secondly, the YCbCr format is typically used in video coding and (MPEG and JPEG) so the use of the same format will avoid the extra computation required in conversion.

4.2 Colour Information as Stopping Criterion

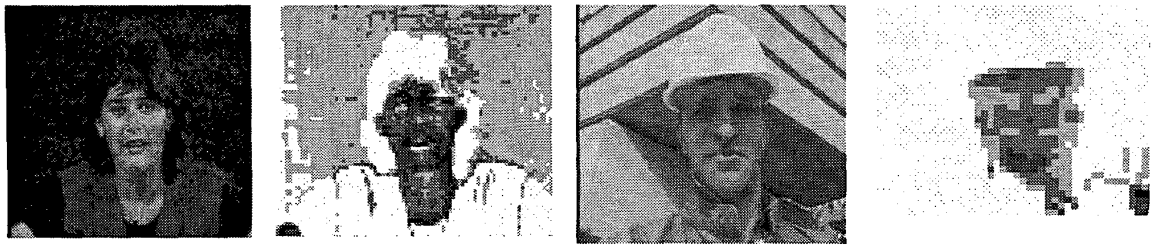
To use the colour information as a segmentation feature, each pixel of the input image should be classified into skin-colour and non-skin colour. To do so, a skin-colour reference map in the YCbCr colour space is needed. Previous research in this area [10] has identified the presence of skin coloured regions to lie in the range *Cbvalues* = [75 125] and *Crvalues* = [135 175] for Cb and Cr values respectively. This range of values was experimentally confirmed on some standard MPEG test sequences.

4.3 Combining Colour Feature with the Priority-labeled FMM Approach

Once the skin-coloured areas in an image are located, the FMM can be used to segment the face area in the image by assigning priority levels, based on where in the *Cbvalues* and *Crvalues* set, the skin-coloured pixel values lie. The resulting colour-map on which the FMM is applied is as shown in Figure 1.

4.4 Limitations of Colour Segmentation

Although colour can provide fast, effective segmentation on its own as a feature, some disadvantages also exist. For example, many objects commonly found in backgrounds, such as wooden object and furniture, are similar in colour to flesh tones, making the segmentation task more difficult. Also, lighting conditions particularly effect colour segmentation as dark shadows on the face as well as bright light reflecting off the subject's face can lead to the skin colour falling out of the predefined colour map thus making it more difficult to pinpoint the exact location of the face.



(a) (b) (a) (b)

Priority 4 (Strong Skin-coloured pixels) Priority 3 Priority 2 Priority 1 (Background)

Figure 1. (a) Original Image: Miss America and Foreman Sequence Respectively (b) Priority Levels assigned to the skin-coloured pixels and the background as shown above

Figure 2. Miss America after luminance regularization

One feature that can be used to distinguish between the background and the facial region where there isn't a clear distinction is to use the luminance, Y, component in the YCbCr colour space. It is found that in videophone sequences, the brightness is nonuniform throughout the facial region, while the background region tends to have a more even distribution of brightness [10]. Hence, based on this characteristic, background regions that have been previously detected due to its skin-colour appearance can be further eliminated. The standard deviation, $\sigma(x,y)$, can be calculated for the Y component of the YCbCr colour space on 4x4 blocks centred on each skin-pixel. If the standard deviation of the block is less than a value of 2, then the corresponding pixel is considered too uniform and so unlikely to be part of the facial region. The resulting colour-map of the Miss America sequence is able to separate the background as shown in Figure 2.

5. ALGORITHMIC PROPOSED APPROACH AND SIMULATION RESULTS

Our proposed algorithm is automatic using unsupervised segmentation, to process videophone image without the need of manual adjustment of any parameter. Moreover, the algorithm can be implemented in near real time, and its underlying assumptions are minimal.

Keeping in mind the need for real time implementation, further computational efficiency is achieved by sub-sampling. An original $n \times n$ image sub sampled by a factor k reduces the image size to $n/k \times n/k$, thus the area is reduced by k^2. This reduces the complexity also by a factor of k^2. The sub-sampled version can then be segmented and the result up-sampled to get the final image. From experimentation it has been observed that a down-sample factor of 4 gave good segmentation results. Thus, pseudo code for the proposed algorithm is as shown in Figure 3. This algorithm was applied to a wide range of standard MPEG test and live video sequences and was found to segment the face areas quickly and efficiently as seen in Figure 4.

A summary of the segmentation results (including the number of *false positives*, detection rate and average CPU time for processing) is shown in Figure 4 and some segmented images are shown in Figure 5. Note that these simulations were carried out on an 800MHz Pentium III processor and the algorithm was coded using Matlab. A face is correctly segmented if the eyes, nose and mouth are clearly visible in the segmented region otherwise it is considered a *false positive* (FP). Thus, a FP includes a partially segmented face region or a background region identified as the face. The detection rate is computed by the ratio of correctly detected faces to the total number of faces appearing in all frames of the sequence.

The simulation results show that colour can be used effectively to segment the face region in 'noisy' backgrounds. However, it does suffer from problems the presence of facial hair or bright light and dark shadows that make skin-colour fall out of the predefined range. Never the less, the colour information can be used successfully to segment faces. Moreover, significant improvement in the subjective quality of low-bit-rate video over mobile networks can be achieved as up to 78% of the image can be identified as the background and so can be transmitted at a lower resolution.

For *every image in the sequence*

Downsample the image

*Use [Cbvalues Crvalues] to extract the Skin-Coloured Pixels,**Sk**.*

*Seperate **Sk** into four Priority levels.*

If Sk**>thresh Use Luminance Regularization to extract skin pixels.**End

Use Priority-Labelled FMM to locate the face region.

Up-sample the final level set i.e. the outline contour of the face region.

Superimpose the result on the original image to get segmented image.

End

Figure 3. Proposed Algorithmic Approach

6. CONCLUDING REMARKS

In this paper, an algorithm for automatic segmentation of the face region in video sequences has been proposed. The above algorithmic approach was used successfully to segment sequences as can be seen in the simulation results. One of the main aims was to locate the face region in an image as quickly and accurately as possible. The time to segment each frame was found to depend mainly on the size of the image and the accuracy depended on the skin-colour type region in the image background. The percentage of frames over/under segmented was usually found to be low across a wide range of sequences. The algorithm is currently implemented in Matlab, but implementation of this algorithm in C or on a DSP chip is ongoing to provide real time segmentation of the video sequences.

Future work will involve the use of temporal and spatial correlation between frames in a video sequence to improve and speed up the segmentation process. The use of other features and a priori knowledge of the image to improve segmentation will also be considered

Video Clip	Total no. of Frames	Size in Pixels	No. of FP	Detection Rate (%)	Average Time/frame Matlab Simulations (in sec)
Miss America	150	288x360	10	96	6.4
Foreman	300	176x144	12	98	1.1
Paddy Live	265	240x320	9	98	5.0
Carphone	380	176x144	30	94	1.2
Grandma	870	176x144	94	91	1.1

Figure 4. Table of Contents

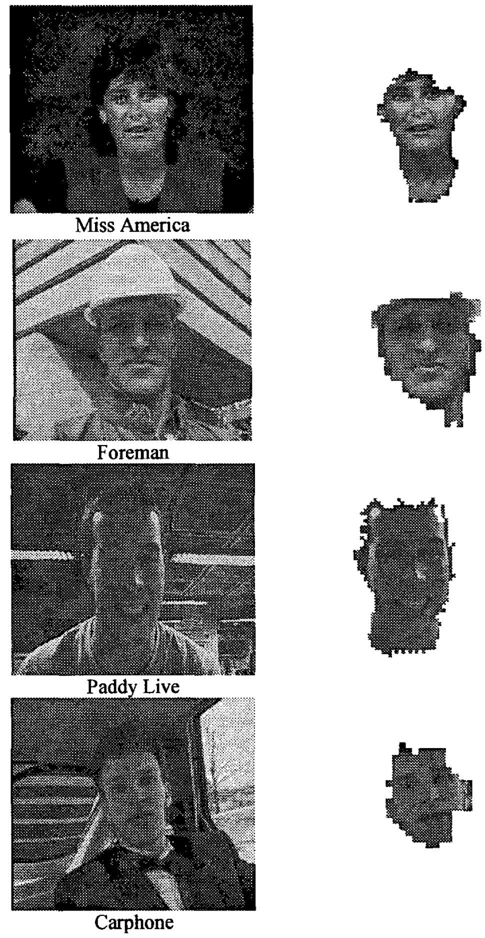


Figure 5. One frame from each sequence

7. REFERENCES

[1] F. Pereira, "MPEG-4: Why, What, How and When?", *Signal Processing: Image Communication. Vol. 15, 2000, p271-279.*

[2] L. Hanzo, P. Cherriman, E-L. Kuan, "Interactive Cellular and Cordless Video Telephony: State-Of-The-Art System Design Principles and Expected Performance", *Proc. of IEEE, Vol. 88, NO. 9, Sept. 2000, pp.1386-1413.*

[3] D. Chai, K. N. Ngan, "Foreground/background Video Coding Scheme", *Proc. IEEE Int. Symp. Circuits Syst., Hong Kong, June 1997, vol. II, pp. 1448-1451.*

[4] A. Eleftheriadis, A. Jacquin, "Model-assisted Coding of Video-teleconferencing Sequences at Low Bit Rates", *Proc. IEEE Int. Symp. Circuits Syst., London, U.K., June 1994, vol. 3, pp. 177-180.*

[5] Ming-Hsuan Yang, N. Ahuja, D. Kriegman, "A Survey on Face Detection Methods", *IEEE Transactions on pattern Analysis and Machine Intelligence, Vol. 24, NO. 1, Jan. 2002.*

[6] A. Mansouri, J. Konrad, "Multiple Motion Segmentation with Level Sets", *IEEE Transactions on Image Processing, Submitted in April 2000.*

[7] P. Kehoe, R. B. Reilly, "Applying Level Set Theory to Digital Video Segmentation using Edge Information", *Proc. of IEEE Workshop on Very Low Bit Rate Video, Greece, 2001*

[8] J.A. Sethian, "Level Set Methods and Fast Marching Methods", *Cambridge University Press, 1999.*

[9] J.A. Sethian, "A Fast Marching Level Set Method for Monotonically Advancing Fronts", *Proceedings of the National Academy of Sciences, 1995.*

[10] N. Herodotu, K. N. Plantaniotis, A. N. Venetsanopoulos, "Automatic Location and Tracking of the Facial Region in Color Video Sequences", *Signal Proc.: Image Communication 14 (1999) 359-388.*

[11] K. Sobottka, I. Pitas, "Extraction of Facial Regions and Features using Color and Shape Information", Tech. Review *Dept. of Informatics, University of Thessaloniki 540 06, Greece.*