Indian J.Sci.Res. 17(2): 306-309, 2018

ISSN: 0976-2876 (Print)
ISSN: 2250-0138(Online)

VIDEO SHOT BOUNDARY DETECTION BASED ON THE DETECTION OF HUMAN FACE REGION FEATURE

[1] ThummalaNagaveni, [2]KalaSamudramMaheshwaridevi, [3]Bommala Neeraja
[1,2,3] Department of Electronics and Communication Engineering, SriIndu College of Engineering and Technology, Hyderabad.

Abstract - A new method for detecting shot boundaries in video sequences using human face region detection is proposed. Skin detection is the process of finding skin-colored pixels and regions in an image or a video. This process is typically used as a pre-processing step to find regions that potentially have human faces. Human skin region detection is process of detecting skin region in the sequence of frames. Skin region detection is mainly used for the identification of the human face detection. This approach is very much suitable for finding shots TV News. The detection technique was tested on TV video sequences having different types of shots and significant object and camera motion inside the shots. It was favourably compared to other recently proposed shot cut detection techniques.

Keywords - Video Shot,SkinDetection,EdgeDetection,SobelOperator,Connectivity algorithm

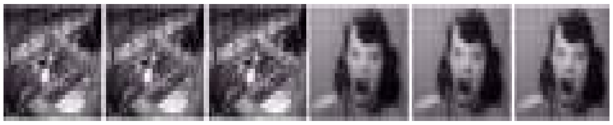
I. Introduction

Video shot detection refers to the detection of transitions between scenes in a digital video stream. It can provide disjoint contiguous video segments that can be utilized as basic units to be indexed, annotated and browsed. The detection of a shot cut involves detecting a significant change in visual content between two frames or gradual change within a number of frames. The development of video shot boundary detection techniques have the longest and richest history in the area of content based video analysis and retrieval. The importance of video shot detection comes from the necessity of these algorithms for almost all high level video processing applications.

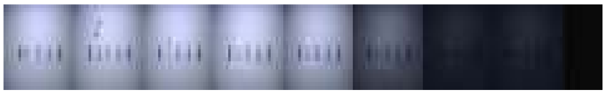
Video shot detection plays a fundamental role in video access/analysis. Among many shot detection schemes, color based schemes seem to be the most widely used ones which can be applied to videos of different domains. Segmenting video clips into continuous camera shots is the prerequisite step for many video processing and analysis applications.

A shot is defined as unbroken sequence of frames taken by one camera. Using motion picture terminology, shot change can belong to one of the following categories:

• Cut: An abrupt change between two consecutive frames where one frame belongs to the disappearing shot and the other belongs to the appearing shot.



• Fade: Either the intensity of disappearing shot changes from normal into black frame (fade out), or intensity of the black frame changes into appearing shot (fade in).



Fade in

Fade out

• Dissolve: In this case, few frames of disappearing shot overlap with few appearing frames of appearing shot. The intensity of disappearing shot decreases to zero (fade out) while intensity of appearing shot increases from zero (fade in).

• Wipe: Here, the appearing and disappearing shots coexist in different spatial regions of the intermediate video frames, and the region occupied by former grows until it gradually replaces the latter.

A. Existing system

The existing shot detection techniques can be classified into two categories: threshold based and machine learning based method. The former usually uses the frame differences for pixel, block-based or histogram

[1]Corresponding Author

comparisons and relies on the suitable threshold selection. However, it should be noted that threshold selection really is a hard problem and it usually depends on the test videos. The latter tries to overcome this drawback by machine learning.The proposed shot detection approach in this paper is based on the human skin detection to find the shot transition and non-shot transition.

B. Proposed approach:

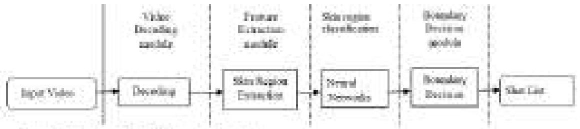


Figure 1. The framework of the proposed method.

The proposed approach is applied in the uncompressed domain of video and consists of three modules, including decoding, human skin region detection and boundary detection based on the human skin region detection, as shown in Fig 1. The input video is first decoded into video frames. Then, the color is to converted from RGB to HSV color model. HSV color frame is used for the skin region detection. Find the skin region is present in the consecutive frames. If the skin region is not present in the frame, then the frame is consider as the boundary frame.

II. Edge Detection

Edge detection refers to the process of identifying and locating sharp discontinuities in an image. The discontinuities are abrupt changes in pixel intensity which characterize boundaries of objects in a scene. Variables involved in the selection of an edge detection operator include:

Edge orientation: The geometry of the operator determines a characteristic direction in which it is most sensitive to edges. Operators can be optimized to look for horizontal, vertical, or diagonal edges .

Noise environment: Edge detection is difficult in noisy images, since both the noise and the edges contain high frequency content. Attempts to reduce the noise result in blurred and distorted edges. Operators used on noisy images are typically larger in scope, so they can average enough data to discount localized noisy pixels.

Edge structure: Not all edges involve a step change in intensity. Effects such as refraction or poor focus can result in objects with boundaries defined by a gradual change in intensity the operator needs to be chosen to be responsive to such a gradual change in those cases. Newer wavelet-based techniques actually characterize the nature of the transition for each edge in order to distinguish, for example, edges associated with hair from edges associated with a face.

H and S scales are partitioned into 100 levels and the color histogram is formed using H and S. In order to train for skin color, we download color images containing human faces from the Internet and extracted the skin regions in these images manually. Given an image, each pixel in the image is classified as skin or non-skin using color information.

Here sobel operator is used for the edge detection:

The operator consists of a pair of 3×3 convolution kernels as shown in Figure 2

-1	0	1
-2	0	2
-1	0	1

-1	-2	-1
0	0	0
1	2	1

Figure 2: Masks used by Sobel operator

These kernels are designed to respond maximally to edges running vertically and horizontally relative to the pixel grid, one kernel for each of the two perpendicular orientations.The kernels can be applied separately to the input image, to produce separate measurements of the gradient component in each orientation (call these *Gx* and *Gy*) [13]. These can then be combined together to find the absolute magnitude of the gradient at each point and the orientation of that gradient. The gradient magnitude is given by:

$$|G| = \sqrt{Gx^2 + Gy^2}$$

Typically, anapproximatemagnitudeiscomputed using

$$|G| = |Gx| + |Gy|$$

Whichismuchfastertocompute. The Sobel operator is slower to compute than the Roberts Cross operator, but its larger convolution kernel smooth's the input image to a greater extent and so makes the operator less sensitive to noise [6]. The operator also generally produces considerably higher output values for similar edges, compared with the Roberts Cross.

III. Skin Region Detection

The first step is to classify each pixel in the frame as a skin pixel or a non-skin pixel. The second step is to identify different skin regions in the skin detected frame by using connectivity analysis. The last step is to decide whether each of skin regions identified is a face or not. They are the height to width ratio of the skin region and the percentage of skin in the rectangle defined by the height and width.

A. Skin Pixel classification

Different color spaces used in skin detection, include HSV, normalized RGB, YCrCb, YIQ and CIELAB. According to Zarit et al., HSV gives the best performance for skin pixel

The histogram is normalized and if the height of the bin corresponding to the H and S values of a pixel exceeds a threshold called skinthreshold (obtained empirically),

E. Algorithm Steps:

Convert the input RGB image(rgb(i,j)) into HSV image (hsv(i,j))

Get the edge map image (edge(i,j)) from RGB image using Sobel operator.

For each pixel (i,j), get the corresponding H and S values.

If(colorhistogram(H,S)>skinthreshold) and (edge(i,j) <edgethreshold) then skin(i,j) = 1 i.e. (i,j) is a skin pixel else skin(i,j) = 0 i.e. (i,j) is a non-skin pixel

Find the different regions in the image by implementing connectivity analysis using 8- connected neighbourhood.

Find height, width, and centroid for each region and percentage of skin in each region.

For each region, if (height/width) or (width/height) is within the range (Goldenratio±tolerance) & (percentage of skin >percentagethreshold) then the region is a face, else it is not a face.

then that pixel is considered a skin pixel. Otherwise the pixel is considered a non-skin pixel. detection. In the HSV space, H stands for hue component, which describes the shade of the color, S stands for saturation component, which describes how pure hue(color) is while V stands for value component, which describes the brightness. The removal of V component takes care of varying lighting conditions. H varies from 0 to 1 on a circular scale i.e. the colors represented by H=0 and H=1are same. S varies from 0 to 1 representing 100 percent purity of the color.

B. Connectivity Analysis

Using the skin detected image, one knows whether a pixel is a skin pixel or not, but cannot say anything about whether a pixel belongs to a face or not. One cannot say anything about it at the pixel level. We need to go to a higher level and so we need to categorize the skin pixels into different groups so that they will represent something meaningful as a group, for example a face, a hand etc. Since we have to form meaningful groups of pixels, it makes sense to group pixels that are connected to each other geometrically. We group the skin pixels in the image based on a 8-connected neighbourhood i.e. if a skin pixel has got another skin pixel in any of its 8 neighbouring places, then both the pixels belong to the same region. At this stage, we have different regions and we have to classify each of these regions as a human face or not. This is done by finding the centroid, height and width of the region as well as the percentage of skin in the rectangular area defined by the above parameters. The centroid is found by the average of the coordinates of all the pixels in that region. For finding height

• The y-coordinate of the centroid is subtracted from the y-coordinates of all pixels in the region.

• Find the average of all the positive coordinates and negative y-coordinates separately.

• Add the absolute values of both the averages and multiply by 2. This gives the average height of the region.

Average width can be found similarly by using x-coordinates. Since the height to width ratio of human faces falls within a small range on the real axis, using this parameter along with percentage of skin in a region, the algorithm should be able to throw away most of non face skin regions.

C. Hardware Requirements

- Intel Pentium IV Processor,
- 1 GB RAM
- 20 GB HDD

D. Software Requirements:

- Operating System: Windows XP SP-3, Windows 7
- MATLAB

IV.Results

In our approach face or skin information from the two consecutive frames are calculated separately. If the skin region is not present in two frames then is consider as not shot boundaries. If one frame contains the face region and other frame doesn't contain the face region, then that frame is considered as shot change.

Below figure shows the output of the video shot detection based on the face detection. figure 3. (a) is the frame contain the face and figure 3. (b) is the does not contain the face, so then this consider as two different shot.

V. Conclusion

For uncompress video data, this paper proposes an innovative shot boundary detection method. A fast algorithm for face detection based on skin color, connectivity and edge information has been used. The algorithm is fast and can be used in real-time applications. The images on which the algorithm is tested are natural images taken under uncontrolled conditions and the algorithm does well on them. The algorithm locates faces but does not give the exact contour. Based on face information in the continues video frames are analyzed and shot boundary is detected. All of the experiments are conducted with

Visual Studio 6.0 (VC++) on Windows XP Platform. We use two parameter, recall and precision to evaluate the performance of the detection. The equation for recall and precision are the following :

$$Recall = \frac{N_c}{N_c + N_m} \times 100\ \%$$

$$Precision = \frac{N_c}{N_c + N_f} \times 100\ \%$$

Where, Nc is the number of correct detection;

Nm is the number of missed detection;

Nf is the number of false detection.

Figure 3. (a) Mega Construction video contain face, bounding box indicate the face region.

(b) Frame does not contain the face

References

[1]. A. Hanjalic, ”Shot-boundary detection: Unraveled and resolved” IEEE Transactions on Circuits and Systems for Video Technology, February 2002 vol.12, pp. 90–105.

[2] Raman Maini ,Dr.HimanshuAggarwal,” Study and Comparison of Various Edge Detection Techniques”, International Journal of Image Processing , Vol.3,Issue 1.

[3] S. Tsekeridou and I. Pitas. Content-based video parsing and indexing based on audio-visual interaction. IEEE Trans. on Circuits and Systems for Video Technology, 11(4):522–535, 2001.

[4] Smoliar, S.W., Zhang, H.-J.: Content-based video indexing and retrieval. IEEE Multimedia 1(2), 62–72 (1994)

[5] Boreczky J S, Rowe L A. Comparison of Video Shot Boundary Detection Techniques [A]. In SPIE Conf Storage & Retrieval for Image & Video Databases[C]. San Jose: SPIE, 1996, 170179.

[6] R. Lienhart. Reliable dissolve detection. In Proc. of SPIE Storage and Retrieval for Media Databases 2001, volume 4315, pages 219–230, January 2001.

[7] www.mathworks.in/products/matlab