A Shot Boundary Detection Method for News Video based Human Skin Region (Face) Detection

Punith Kumar M B
Research Scholar
Dept. of ECE, BGSIT
Mandya, India.

P. S. Puttaswamy, Ph.D
Professor
Dept. of EEE,
PESCE, Mandya
Mandya, India.

ABSTRACT
A New simple approach to detect, classify shot boundaries and store shot boundary frames in Video sequence using human skin region detection based approach is proposed. Human skin region detection is the process of detecting skin region in 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 in TV News so that we can classify anchor and non anchor frames to save the overall time which is required to watch overall news,

Keywords
Shot Detection, Boundary, Frame, Skin region, Histogram.

1. INTRODUCTION
Face detection of TV News can be adapted for the Video shot detection refers to the detection of transitions between scenes in a digital video stream, the face detection plays plenty of applications in different areas like security and where we can value for the times by this we can save the time, efficient communication, memory management, person tracking,

Today due to time and work commitment the people are not finding time to know what is happening in the society due to their busy work schedule so in order to know we have storing option with our communication device like TV but in order to know about it we can adapt video shot detection techniques. In case of TV news video we are not interested to watch the overall video whatever stored in our storing device the video will be consisted of anchors, reporter, advertise, etc so we need to identify those video with the help of frame analysis by face detection.

So when we reach the home by late night we cont find the time to watch the overall TV news whatever we have stored in our device due to the advance technology like Tata sky, dish TV sun direct, etc. this are all work basically on storing concept instead of writing the whole video we can use video shot detection for the summarization of anchors, non anchors, advertise and other related videos which are not required for us so that we can save the time not only with memory whatever for storing the whole news.

In a similar manner we can adapt this video shot detection even for games to watch the particular bating or bowling or audience or etc

Let us consider the TV news video in which we know that anchor will be there to take care of whole video presentation for the corresponding news, the news video will be consisting of different frames included by anchors shot, reporter shot and advertisement so our technique will find the frames for the corresponding video and store them in separate folder.

Generally speaking, 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 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. The rest of this paper is organized as follows. In section 2, we first introduce the framework. Then the skin region detection in section 3. In section 4 boundary detection is done based on the face detection in the frame. Conclusion are drawn in section 5.

2. FRAMEWORK
The proposed approach is applied in the uncompressed domain of video and consist of three modules, including decoding, human skin region detection and boundary detection as shown in the figure 1. The input video is first decoded into video frames than for the face detection there are lot many techniques here we using RGB to HSV color frames is used for the skin region detection. Find the skin region is present in the consecutive frames. Than the frame is considered as the boundary frame.

3. SKIN REGION DETECTION
The first step is to classify each pixel in the frame as a skin or non skin pixel. The second step is to identify different skin regions in the skin detected frames by using connectivity analysis. The last step is to decide whether each of skin region identified is a face or not. They are the height to width ratio of the skin region and the percentage of the skin in rectangle defined by the height and width.
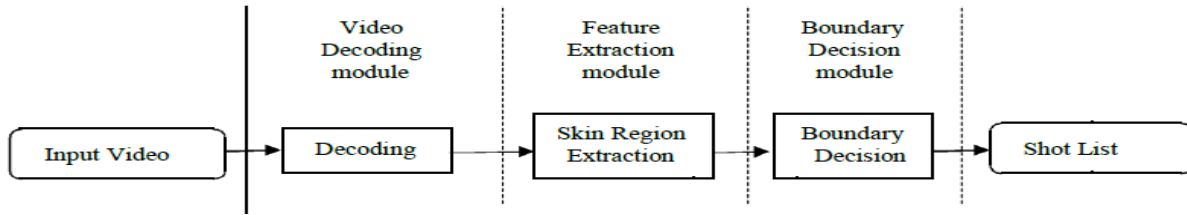
Figure 1: The framework of the proposed method

3.1 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 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 the hue(color) is while V stands for value component, which describes the, the removal of V component takes care of varying lighting conditions. H varies from 0 to 1 on a circular scale i.e. the colour represented by H=0 and H=1 are same. S varies from 0 to 1 representing 100 percent purity of a colour.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 colour information. The histogram is normalized and if the height corresponding to the H and S values of a pixels exceeds a threshold called skinthreshold (obtained empirically), than that pixel considered as a skin pixels otherwise the pixels is considered as non skin pixel. A skin detected frame is one in which only he skin pixels are shown.

Figure 2: Classification of pixels as skin or non-skin
(a) Original Image   (b) Skin detected Image

A general image and its skin detected image can be seen in the figure 2 (a) and (b) respectively. The figure 2 is taken from the internet.

3.2 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 neighborhood i.e. if a skin pixel has got another skin pixel in any of its 8 neighboring 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

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

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

3. 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.

Algorithm for finding face region:
Step 1. Convert the input RGB image (rgb(i,j)) into HSV image (hsv(i,j))

Step 2. Get the edge map image (edge(i,j)) from RGB image using Sobel operator.

Step 3. For each pixel (i,j), get the corresponding H and S values.

Step 4. 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

Step 5. Find the different regions in the image by implementing connectivity analysis using 8-connected neighborhood.

Step 6. Find height, width, and centroid for each region and percentage of skin in each region.

Step 7. For each region, if (height/width) or (width/height) is within the range (Golden ratio ± tolerance) and (percentage of skin > percentage threshold) then the region is a face, else it is not a face.

4. FACE DETECTION

In our approach the 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.

The figure shows bellow are the output of the video shot detection based on the face detection. Figure 3. (a) and (b) is the frame contain the face and figure 4. is the does not contain the face, so then this consider as two different shot.

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

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

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

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

Where **Nc** is the number of correct detection,
Nm is the number of missed detection,
Nf is the number of false detection.

5. CONCLUSIONS

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 images on which the algorithm is tested are natural images taken under uncontrolled conditions and the algorithm does well on them

Figure 4. Frame does not contain the face

The algorithm locates faces based on face information in the continues video frames are analyzed and shot boundary is detected. The algorithm is fast and can be used in real-time applications.

All of the experiments are conducted with the help of MATLAB 2010 b on windows Platform. The use of two parameters, such as recall and precision are used to evaluate the performance of the detection.

6. FUTURE ENHANCEMENT

As we are working with this project we planned for the one anchor frame detection future improvement can be done for the more than one anchor in the given news, along with moving line detection also can be detected using text detection algorithms

7. ACKNOWLEDGMENTS

8. 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] C. Cotsaces, N. Nikolaidis and L. Pitas,"Video Shot Detection and Condensed Representation: A Review", Signal Processing Magazine, Mar 2006, vol. 23, pp. 23–37.

[3] Y. Qi, A. Hauptmann, T. Liu, "Supervised Classification for Video Shot Segmentation", IEEE International Conference on Multimedia and Expo (ICME), 2003, vol.2, pp. 689–692.

[4] B. Gunsel, A.Ferman, A.M. Tekalp,"Temporal video segmentation using unsupervised clustering and semantic object tracking", Journal of Electronic Imaging 1998, pp. 592–604.

[5] T. Butz, J-P. Thiran, "Shot Boundary Derection with Mutual Information", International Conference on Image Processing (ISIP) 2001, vol. 3, pp.422–425.

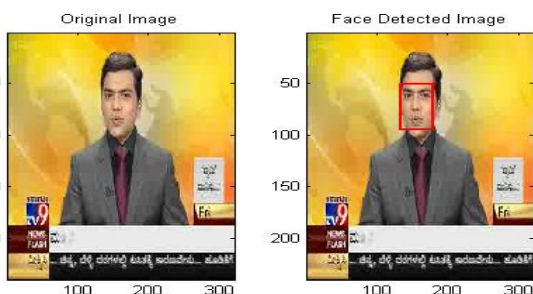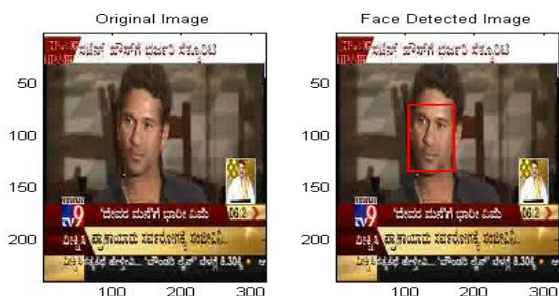[6]. "Shot Classification in Broadcast Soccer Video" by Xiaofeng Tong, Qingshan Liu, Hanqing Lu Electronic Letters on Computer Vision and Image Analysis 7(1):16-25, 2008.