An image calibration procedure for enhancing the performance of video-shot detection algorithm based on histogram analysis

A. Damoni\(^1\) and D. Mukherjee

Engineering Design Department, FESBE, London South Bank University, 103 Borough Road, London, U.K.

\(^1\) Currently with Faculty of Computer Science, Riinvest University, 10000 Pristine, Republic of Kosova

Abstract - The paper describes an image processing method in connection with the use of multimodal data analysis for detecting video-shot boundaries. The method has been effectively utilised in the design of a novel video-shot boundary detection system relying on the simultaneous implementation of luminance (Y) based histogram analysis and audio characterisation associated with the multimedia data.

The method is involved with two separate processes. First an image analysis is carried out on the primary colours, captured under the laboratory conditions. A detailed statistical analysis of the frames within a video sequence is then carried out via generating their histograms. This enables the most appropriate threshold difference for the histograms to be extracted towards identifying the video-shot boundaries with high accuracy. The method can serve as a routine calibration tool for both testing the validity a digital video measurement system and analysing its performance.

Keywords: Image analysis, Multimodal data analysis, Video indexing, Video shot detection, Histogram comparison

1 Introduction

The video-shot boundary detection system is involved with comparing the contents of consecutive frames in a video footage with a view to observing any significant difference between them. A video indexing algorithm monitors this difference via generating the histograms of the images [1], representing the distribution of the pixels within the frames in accordance with their intensities. The pixels having the same intensity are grouped together into the so-called bins.

A comparison of the histograms of successive frames monitors the change in the cumulative total of the pixels associated with the individual bins, the number of which depends on the data resolution used in the analysis. A significant variation in the characteristics of the histograms therefore demonstrates a substantial change in the image from frame-to-frame, thereby indicating the possibility of the existence of a shot-boundary. The main advantage of the method is that its performance is not influenced by any external manipulation of the image through rotation, scaling, transition as well as camera and object motions. It is more dependent on the frame’s overall content rather than the exact values of the individual pixels.

The efficiency of a shot-boundary detection method based on histogram comparison depends on the determination of an effective threshold for the difference in the histograms of consecutive frames. The use of a threshold value which is too high, leads to an increase in the number of missed detections M, while a low threshold value may lead to an increase in the number of false detections F. The threshold difference can be generally expressed [2] as:

\[
T = \mu + \alpha \sigma
\]

Where, the mean \(\mu\) and the standard deviation \(\sigma\) are obtained through considering frame-to-frame differences for the entire length of the video footage. The term \(\alpha\) is a constant, the value of which is known to vary between 5 and 6 [2], for luminance based algorithms and has a fixed value of 4 in the case of chrominance components [3]. In order to calculate the threshold difference (T) for identifying the shot boundaries, it is necessary to compute the histogram difference for all the successive frames that appear within the entire length of a video.

The performance of a shot detection system is primarily determined by the number of correctly identified shots (C), while taking into consideration the numbers of missed (M) and false (F) detections. They collectively give rise to two important parameters, namely the Precision P and Recall R which can be expressed as:

\[
P = \frac{C}{C + F}
\]

\[
R = \frac{C}{C + M}
\]
Ideally, both F and M should be zero, to give 100% accuracy of detection, meaning that all existing shot boundaries are correctly detected, without any shot being missed or falsely identified.

2 Experimental method

The calibration method proposed in this investigation is carried out in two stages. First the still images of the primary colours, captured under the laboratory conditions are analysed. The results of the analysis are then used to determine the most appropriate threshold difference for the histogram comparison. This provides a basis for optimising the P & R values to improve the shot detection accuracy.

The algorithm for the data capture and subsequent analysis was developed by using the MATLAB software package. Because of the large memory requirement by the MATLAB, the file size was kept limited to 10MB. The machine used for the experiments had a 2Ghz AthlonXP CPU, 512MB RAM and 40GB HDD. The coding for the development of multimedia indexing, which required video-shot detection was done in Object Oriented Programming (OOP) environment.

3 Results and analysis

The figures 1(a), 1(b) and 1(c) below show the histograms obtained for the three primary colours, Red (R), green (G) and blue (B) respectively. These graphs represent the distribution of pixels among 256 bins. This is the default value associated with the MATLAB software. All three graphs show that the distribution of the pixels are essentially localised within relatively narrow spreads of the bins. The peak intensity for R ($1.85 \times 10^5$), in figure 1 (a), for example, is seen to be significantly higher compared to that for G ($2.1 \times 10^4$) and B ($1.8 \times 10^4$), with similar behaviour being observed for G and B in figures 1 (b) and 1 (c) respectively. This is indicative of the presence of mainly a single hue, corresponding to the respective primary colour.

The intensities of all the peaks within a histogram collectively give the total number of pixels for the frame. This can be verified easily by re-plotting fig.1 (a), for example, with number of bins being reduced to say 64, as shown in 2. The bins 252 and 256 shown in figure 2 should thus be read here as 63 and 64 respectively. The accumulated total value of the pixels for these two bins is $4.1 \times 10^5$ ($0.8 \times 10^5 + 3.3 \times 10^5$). This is consistent with the expected count for the screen resolution of 720 x 576, which has been used in this study.

The generated histograms can therefore form the basis of comparing the content of two images (frames), as has been

Fig. 1 Histograms obtained for the still images of the primary colours: (a) Red, (b) Blue and (c) Green.

Fig. 2 The histogram of the primary colour R with 64-bins
which, as can be seen in the diagram, exceeds the
above is assumed to have a value between 5 and 6 [2].

For the three predetermined cuts to be detected in
this case, the value of T must lie within the range of 0.25
and 0.69, giving the corresponding α values to be 0.73 & 3.68
respectively. A detailed correlative study involving T and α
shows that a result of 100 % for both P and R can be achieved
for a T value of 0.5.

The results of this study also suggest that the P and R
values are dependent on the threshold difference T, as can be
seen in figure 5 for a video sample (Sample1). It is
clearly evident here that P generally increases with an increase
of the T value while the trend is opposite for the recall value.
Although there are no clear patterns for these changes, the
graphs provide a useful basis for the conditions of maximum
detection efficiency to be achieved. As can be observed in
the figure, a threshold value of 0.45 gives the combination of

![Histogram difference obtained for a video sample (Sample1), indicating the mean μ (grey line) and the standard deviation σ- (black line)](image)

![Precision & Recall values on the level of threshold](image)
the best possible values for both the P and R. This can therefore be regarded as the optimum threshold value, for the particular sample used (A1). This type of iteration process and corresponding adjustment of the threshold value are needed to be carried out for each video under investigation when the static threshold is used.

The precision and recall values obtained from figure 5 can be significantly improved by the application of a progressively varying threshold, as is indicated by the dark plot (T) in figure 6 below for the sample (A1). Here, the µ and σ values, represented by the dotted plots in light grey are shown in the lower part of the diagram. They are continuously monitored using a sliding window of length covering two frames. The threshold value thus becomes adaptive to the frame-to-frame histogram variations which are characterised by the plot in dark grey. As can be seen, several false cuts are avoided in this case that would have otherwise been detected by the frame-to-frame monitoring, represented by the dark-grey plot. The adaptive threshold also does not require the entire video to be previewed first for obtaining the µ and σ values, thereby saving in the processing time.

4 Conclusions

An image processing method has been proposed which can be effectively used for calibrating a general purpose histogram measurement system. The method enables the precision and recall values to be optimised with a view to determining the threshold differences for the measurement with good accuracy. Both the static and adaptive threshold values have been obtained with the later providing better accuracy and a basis for faster processing.

5 References

