Lighting Condition Instability Measure for Color-Based Video Information System

Xiaomeng Wu, Shunsuke Kamijo, and Masao Sakauchi

1 Introduction

In this paper, a novel approach, which automatically measures the instability of the lighting condition of the video, is proposed to provide selection measure of video for color-based video system. Foreground regions appearing in video frames that do not reflect the lighting condition are first detected and separated from the background images. An information-theoretic measure, termed entropy error rate, is used as a quantitative measure of the information distribution within an image. This is further extended to the video case and used to quantitatively represent the lighting condition of each video scene. The instability of the lighting condition of the video is thus measured by calculating the instability of features extracted from the extended measure. Furthermore, an experiment is generated to show the performance of the proposed approach on reflecting the instability. This work has high potential for use in practical applications to provide selection measure of video for color-based video system.

2 Foreground Region Separation

In this work, salient regions that might be the foreground objects are detected based on the saliency map and separated from the background image. Characters regions are extracted by face detection and the dress region estimation approach. The foreground region separation processing generates a more than 90% precision while a 40% recall due to the performance of the detectors. This is considered acceptable because our purpose is not to accurately detect the object or the character but to alleviate the effects of foreground regions on the lighting condition measure to the greatest possible extent.

3 Instability Measure

In this work, we use an information-theoretic measure as a quantitative measure of the information distribution within an image. In the intensity histogram of the keyframe extracted from the video, the minimum intensity value is denoted by I_{min} and the maximum intensity value is denoted by I_{max}. The mean intensity value I_{mean} can be computed with Equation (1).

$$I_{mean} = \sum_{K=I_{min}}^{I_{max}} K * P(K) \tag{1}$$

where K is the intensity value, and $P(K)$ is the probability of this intensity value occurring. Here, the average mean intensity value of the video is used to define the dark-light pixels of the keyframe. Given the mean intensity values of a set of N keyframes extracted from a video in time order $\{I_{mean1}, I_{mean2}, \cdots I_{meanN}\}$, the average mean intensity value of the set, i.e. the average mean intensity value of the video, can be computed with Equation (2).

$$E(I_{mean}) = \frac{1}{N}\sum I_{mean} \tag{2}$$

We use this value as the average lighting condition of the video to divide each keyframe into darker pixels, where the intensity values of pixels are below $E(I_{mean})$, and lighter pixels, where the intensity values of pixels are above $E(I_{mean})$.

Furthermore, a simple statistic S, called the singularity, can be introduced to estimate the relative position of the $E(I_{mean})$ within the intensity histogram.

$$S = 4\left(\frac{I_{max} - E(I_{mean})}{I_{max} - I_{min}}\right)\left(\frac{E(I_{mean}) - I_{min}}{I_{max} - I_{min}}\right) \tag{3}$$

Inside an image, information contained in the darker pixels H_D and in the lighter pixels H_B can be measured by the entropy in Equations (4) and (5), respectively.

$$H_D = \sum_{K=I_{min}}^{I_{mean}} -P(K)\log P(K) \tag{4}$$

$$H_B = \sum_{K=I_{mean}}^{I_{max}} -P(K)\log P(K) \tag{5}$$

Moreover, the average entropy of both sides, which measure the amount of information contained in only one unit of intensity level, can be individually calculated with Equations (6) and (7).

$$\overline{H}_D = \frac{H_D}{I_{mean} - I_{min} + 1} \tag{6}$$

$$\overline{H}_B = \frac{H_B}{I_{max} - I_{mean} + 1} \tag{7}$$

The asymmetry of the image information distribution between these two sides can thus be measured using the simple statistic in Equation (8), which is termed the entropy error rate (EER).

$$EER = \frac{\overline{H}_D - \overline{H}_B}{1 + a^S} \tag{8}$$

Furthermore, the lighting condition of each scene within the video is quantitatively represented with Equation (9). This is to ensure the measure of the lighting condition L being always positive. The parameters $a(a > 1)$ and $b(b > 1)$ are used to control the value range of the measure. $L > 1$ indicates a relatively dark image, while $L < 1$ indicates a relatively bright one. A high-quality image should be neither too dark nor too light, so its L value should be within an acceptable range (L should be close to 1).

$$L = b^{EER} \tag{9}$$

If the L values of the keyframes are far different from each other, it will be considered that the lighting condition of the video varies frequently in time order. On the other hand, L values with little variation indicate that the light condition of the video is stable and only vary within a narrow scope. Therefore, in this work, the instability of the lighting condition of the video is measured by simply calculating the instability of the features, i.e. the L values extracted from the extended measure. In this work, we use the long-variance index approach to take the time series of the L values of all the keyframes into account. Given the L values of a set of N keyframes extracted from a video in time order $\{L_1, L_2, \cdots L_N\}$, the instability of the lighting condition of the video is defined by Equation (10).

$$Instability = \left[\exp\left(V_{log}^{\frac{1}{2}}\right) - c\right] * d \tag{10}$$

where the parameters c and d are used to control the value range of the measure.

$$V_{log} = \frac{1}{N-1} \sum (\log L_{i+1} - \log L_i - M_{log})^2 \tag{11}$$

$$M_{log} = \frac{1}{N-1} \sum (\log L_{i+1} - \log L_i) \tag{12}$$

4 Experimental Results

To examine the wellness that the proposed measure characterizes the instability of the lighting condition, the relationship between the *Instability* and the performance of a color-based video object annotation system is statistically simulated. 36 data sets are randomly extracted from 20 videos, the size of which is 200 images/set. From these data sets, 50 object models are constructed and retrieved. The instability of each data set is quantitatively computed using the proposed approach described in section 3. The relationship between the *Instability* and the performance of the implemented system is illustrated in Figure 1.

From Figure 1, we can see that the approximation curve declines as the *Instability* values increases, which successfully reflects the relationship that the performance of color-based video system decreases as the lighting condition of the object video becomes instable.

The proposed approach also provides a selection measure of video for color-based video system. In Figure 1, the recall rate is fixed to 90% . In fact, approximation curves with the recall rate varying from 10% to 90% could all be generated from the experiment. The user can find videos that the system fits in well with or make a prediction on the performance of the system corresponding to the object video by only computing the *Instability* value with the use of the proposed approach and find the corresponding estimate values from these approximation curves. In this experiment, videos of real TV programs, the total length of which is about 147 hours, are recorded from NHK TV station. The *Instability* values of these program videos are computed, and the distribution is illustrated in Figure 1. For videos with relatively stable lighting conditions (*Instability* < 10), which account for 50% of the NHK videos, the implemented system shows a good performance with an average 74.4% precision and 90% recall.

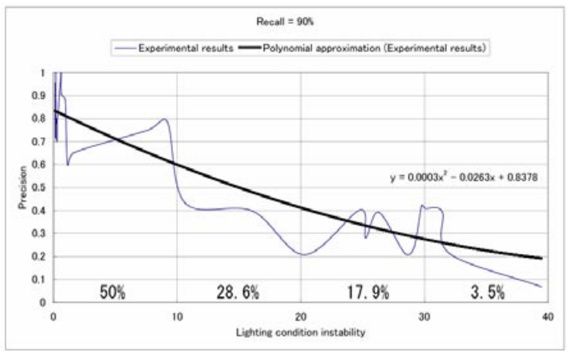

Figure 1: Relationship between *Instability* and performance of color-based video database system