# HORROR MOVIE SCENE RECOGNITION BASED ON EMOTIONAL PERCEPTION

*Jianchao Wang, Bing Li, Weiming Hu, Ou Wu*

National Laboratory of Pattern Recognition, Institute of Automation Chinese Academy of Sciences,
95 Zhongguancun East Road, 100190, BEIJING, CHINA
{ericustc, bjtulb}@gmail.com, {wmhu, wuou}@nlpr.ia.ac.cn

## ABSTRACT

The number of video clips available online is growing at a tremendous pace. Meanwhile, the video scenes of pornography, violence and horror permeate the whole Web. Horror videos, whose threat to children's health is no less than pornographic video, are sometimes neglected by existing Web filtering tools. Consequently, an effective horror video filtering tool is necessary for preventing children from accessing these horror videos. In this paper, by introducing color emotion and color harmony theories, we propose a horror video scene recognition algorithm. Firstly, the video scenes are decomposed into a set of shots. Then we extract the visual features, audio features and color emotion features of each shot. Finally, by combining the three features, the horror video scenes are recognized by the Support Vector Machine (SVM) classifier. According to the experimental results on diverse video scenes, the proposed scheme based on the emotional perception could deal effectively with the horror video scene recognition and promising results are achieved.

***Index Terms*—** Horror Movie Recognition, Affective Understanding, Color Emotion, Color Harmony

## 1. INTRODUCTION

With the proliferation of harmful Web content such as pornography, violence, and horror messages, which seriously affect human physical and psychological health, especially for children, effective content-filtering systems are essential. Most of existing Web filtering studies mainly focus on pornographic information filtering [1]. To the best of our knowledge, there is nearly no specific study on the horror information filtering although the horror information's threat to children's physical and psychological health is no less than pornographic. The studies of psychology and physiology found that too much horror information seriously affects children's physical and psychological health. The experiments in [2] indicate that 88.8% children ascribe their phobias to horror information acquisition. Considering the harm of horror information, an effective horror information filtering tool is necessary for preventing children from accessing horror content.

Horror movies, which are major component of horror messages, are films that strive to elicit the emotions of fear, horror and terror from viewers. How to bridge the gap between low-level perceptual features and high-level affective understanding is a challenge for horror video recognition. Although there is nearly no special research on horror video recognition, research on affective content analysis in films is an established research area. In a series of papers, Wang et al. [3] extract a number of effective audiovisual cues to help bridge the affective gap and introduce a holistic method of extracting affective information from the multifaceted audio stream. In [4], a computational framework for affective video content representation and modeling is proposed and the affective video content is mapped onto the 2-D emotion space. The emotion space is characterized by the dimensions of arousal (intensity of affect) and valence (type of affect), by using the models that link the arousal and valence dimensions to low-level features extracted from the video data. Rasheed et al. [5] present a framework for the classification of films into genres, based only on computable visual cues. Audio emotional events (AEE) such as laughing, horror sounds, are detected to locate corresponding video segments in the comedy and horror films in [6].

In the aforementioned papers, the features used for affective content analysis are all low-level. In order to bridge the affective gap, color emotion related features are extracted for horror video recognition. These color emotion related features are derived from psychological experiments. Consequently, they contain much high-level emotional perception. Combining low-level visual features and audio features with color emotion related features, we present a horror movie scene recognition method.

## 2. PROPOSED METHOD

The proposed approach consists of four main steps as illustrated in Fig.1. Audio stream and video frames are separated from each other firstly. Then audio stream is represented

Thanks to the NSFC (Grant No. 60825204, 60935002 and 60903147) for funding. The author Bing Li is also supported by China Postdoctoral Science Foundation and K.C.Wong Education Foundation.

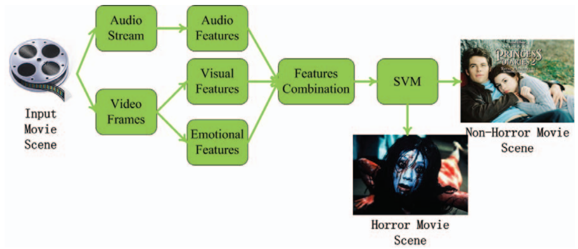


**Fig. 1**. Horror Movie Scene Recognition Scheme

by the audio features and video frames are characterized by the visual features and color emotion features. Four feature combinations are obtained from the three types of different features. Finally, based on the extracted features, the horror movie scene is recognized by the SVM classifier. The details of the components of the proposed approach are discussed in the subsection.

Video segmentation is a fundamental step in analyzing video content. Cernekova et al. [7] used mutual information (MI) to measure information transported from one frame to another. Abrupt transitions and fades between two shots lead to a low level of MI. This approach achieves an impressive performance on shot change detection, so we adopt this method to segment the movie scene into shots, and then the central frame of each shot is chosen as the key-frame. Besides video segmentation, feature extraction is the most important step for horror video recognition. In order to represent the horror movie scene, visual features ($VF$), audio features ($AF$) and color emotion features ($EF$) are extracted.

### 2.1. Visual Features

The average shot length, which is computed as the average number of frames of each shot in a movie scene, represents the tempo of the scene. To the viewer, rapid shot changes certainly convey the dynamic and breathtaking excitement far more effectively than a long duration shot [5]. Consequently, the average shot length is used as the first visual feature($VF_1$). Intuitively, the variance of color has a strong relation with movie genres. For instance, comedies tend to have a large variety of bright colors, whereas horror films often adopt only darker hues [5]. To represent the variance of color used in the movie scene, we employ the generalized variance of $Luv$ color space of the key frames in the movie scene. The covariance matrix of $L, u, v$ of each key frame is defined as:

$$\rho = \begin{bmatrix} \sigma_L^2 & \sigma_{Lu}^2 & \sigma_{Lv}^2 \\ \sigma_{Lu}^2 & \sigma_L^2 & \sigma_{uv}^2 \\ \sigma_{Lv}^2 & \sigma_{uv}^2 & \sigma_v^2 \end{bmatrix} \tag{1}$$

The generalized variance of each frame is obtained by finding the determinant of (1):$\Delta = \det(\rho)$. The color variance of the movie scene is characterized by the mean of the generalized variances of all the key frames,$VF_2 = \overline{\Delta}$.

Generations of film makers have exploited luminance to evoke emotions, using techniques that are well studied and documented in cinematography circles. Generally two major lighting techniques, low-key lighting and high-key lighting, are frequently employed. In the cinematographic perceptive, the sadness, fear, and surprise for sad, frightening, or suspense scenes are recreated by the use of dim lights, shadow play, and predominantly dark background [3][5]. Lighting key is determined by two factors: 1) the general level of light and 2) the proportion of shadow area. In order to detect the lighting key, two visual features are formatted in [5]. The first component, the general level of brightness, can be characterized by using the median,$Med$, of the $L$ value of the $Luv$ color space of the key frame. The proportion of pixels, $Pro$, whose lightness are below a certain shadow threshold $Th$ , is used as an indicator of the second component which is the proportion of shadow area. is experimentally determined to be 0.18. We can obtain the respective mean,$\overline{Med}$ and $\overline{Pro}$, of $Med$ and $Pro$ of all the key frames to represent the lighting key of the movie scene,$[VF_3, VF_4] = [\overline{Med}, \overline{Pro}]$.

The HSV color space emphasizes the visual perception of the variation in hue, saturation and intensity values of an image pixel, so we calculate the means and variances of the HSV color as the overall characteristics of the key frames,$[VF_5, VF_6, VF_7, VF_8, VF_9, VF_{10}] = [\overline{H}, \overline{S}, \overline{V}, \overline{\sigma_H^2}, \overline{\sigma_S^2}, \overline{\sigma_V^2}]$.

Texture is another important factor related with image emotion. Geusebroek et al [8] report a six-stimulus basis for stochastic texture perception. Fragmentation of the scene by a chaotic process causes the spatial scene statistics to conform to a Weibull-distribution, as shown in Eq(2).

$$wb(x) = \frac{\gamma}{\beta}\left(\frac{x}{\beta}\right)^{\gamma-1} e^{-\frac{1}{\gamma}\left(\frac{x}{\beta}\right)^{\gamma}} \tag{2}$$

The parameters of the distribution can completely characterize the spatial structure of the texture. The contrast of an image is indicated by the width of the distribution $\beta$, and the grain size is given by $\gamma$, which is the peakness of the distribution. Hence, a higher value for $\gamma$ indicates a smaller grain size (more fine textures), while a higher value for $\beta$ indicates more contrast. The Weibull texture feature of a key frame is also adopted in this paper,$[VF_{11}, VF_{12}] = [\overline{\beta}, \overline{\gamma}]$, where $\overline{\beta}$ and $\overline{\gamma}$ are the respective mean of $\beta$ and $\gamma$ of all the key frames in the movie scene.

### 2.2. Audio Features

It is well known that, in a sound film, movie editors usually use some specific sounds and music to highlight emotional atmosphere and promote dramatic effects. Although the emotional meaning of music is subjective and it depends on many factors including culture, it is also found that, within a given cultural context, there is an agreement among individuals as to the mood elicited by music [9].

The mel-frequency cepstrum has proven to be highly effective in automatic speech recognition and in modeling the subjective pitch and frequency content of audio signals. The mel-cepstral features can be illustrated by Mel-Frequency Cepstral Coefficients (MFCCs), which are computed from the Fast Fourier Transform (FFT) power coefficients. For audio signal, we extract a single-channel audio stream at 44.1 KHz and compute 12 MFCCs over 20ms frames. he means $Mean$ and variances $Var$ of both the 12 MFCCs of each frame and 12 MFCCS' first-order differential are adopted in this paper,$[AF_1, \cdots, AF_{24}, AF_{25}, \cdots, AF_{48}] = [Mean(1), \cdots, Mean(24), Var(1), \cdots, Var(24)]$.

For an audio signal $s(n)$, each frame is weighted with a hamming window $h(n)$, where $N$ is the number of samples of each frame. The spectral power($P$) of the signal $s(n)$ of the $kth$ audio frame is calculated as

$$P(k) = 10\log\left[\frac{1}{N}\left\|\sum_{0}^{N-1} s(n)h(n)\exp(-j2\pi\frac{nk}{N})\right\|^2\right] \quad (3)$$

The mean of spectral powers of the audio is calculated as one feature of the audio,$AF_{49} = \overline{P(k)}$.

The spectral centroid $sc$ of audio signal whose mean and variance are employed is a measure of spectral shape and higher centroid values correspond to "brighter" texture with more high frequencies. Time domain zero crossings rate $zcr$ provide a measure of the noisiness of the audio signal [10],$[AF_{50}, AF_{51}, AF_{52}] = [\overline{sc}, \overline{\sigma_{sc}^2}, zcr]$.

### 2.3. Color Emotion Features

Both visual features and audio features belong to low-level features that contain little high-level emotion perception. In this subsection, we extract some emotion features based color emotion and color harmony theories. Color emotion and color harmony are the high-level semantic concepts of images. Ou et al. [11] used psychophysical experiments to develop color emotion models for single colors and two-color combinations and investigated the relationship between color emotion and color preference. Color emotion model for single-colors are derived from psychophysical experiments, resulting in three color emotion factors: activity($A$), weight($W$)and heat($H$):

$$\begin{aligned}
A &= -2.1 + 0.06\left[(L^* - 50)^2 + (a^* - 3)^2 + \left(\tfrac{b^* - 17}{1.4}\right)^2\right]^{1/2} \\
W &= -1.8 + 0.04(100 - L^*) + 0.45\cos(h - 100°) \\
H &= -0.5 + 0.02(C^*)^{1.07}\cos(h - 50°)
\end{aligned} \quad (4)$$

where $(L^*, a^*, b^*)$ and $(L^*, C^*, h)$ are the color values in CIELAB and CIELCH color spaces respectively.

Color preference indicates whether a color/color combination is preferred by a group of viewers. Color preference (Pre) modeling [12] which is based on the three color-emotion factors in Eq (4) is defined as:

$$\begin{aligned}
A' &= -0.05 + 0.60\tanh(1.55A + 0.73) \\
Pre &= -0.01 + 0.84A' - 0.18W - 0.14H
\end{aligned} \quad (5)$$

Given a key frame $I$ of video sequence, we convert its RGB color to CIELAB and CIELCH color spaces; then the corresponding color preference value of every pixel $I(x, y)$ of $I$ can be calculated by means of Eq (4) and (5). The color preference histogram $CPH$ of key frames which has 20 bins can be obtained and is employed as one part of the color emotion features, $[EF_1, \cdots, EF_{20}] = [CPH(1), \cdots, CPH(20)]$.

In color research, a common definition of harmonious color combination combinations is "colors that are said to generate a pleasant effect when seen in neighboring areas". We utilize the quantitative two-color harmony model developed by Ou et al. [13] to derive color harmony scores. The model consists of three independent color harmony factors: chromatic effect ($H_C$), lightness effect ($H_L$), and hue effect ($H_H$). The factors are combined to form a two-color harmony model, where $CH$ defines the overall harmony score.

$$CH = H_C + H_L + H_H \quad (6)$$

Since the equations for the three factors are very complex, we do not provide the entire equation here, the details can be found in [12]. According to color harmony theory [12], we seek the block which is most disharmonious with the whole frame to represent the key frame. Firstly, the key frame of the video sequence is divided into $3 \times 3$ blocks, then we compute the color harmony scores between every block and the whole key frame image according to Eq (6), the three independent color harmony factors of the block whose harmony score between itself and the whole key frame is lowest are chosen as the other part of the color emotion features, $[EF_{21}, EF_{22}, EF_{23}] = [H_C, H_L, H_H]$.

## 3. EXPERIMENTS

In this section, the proposed method is evaluated on a large number of video clips, which are collected from diverse films.

### 3.1. Data Set

We download a large number of movies from the Internet. The film data collected from the Internet consist of 90 horror movies and 90 non-horror movies which are from different countries such as China, US, Japan, South Korea and Thailand etc. The genres of the non-horror movies are comedy, action, drama and cartoon. We get 365 horror movie scenes, which are no longer than two minutes, and 365 non-horror movie scenes, which all last one minute. The movie scenes are divided into subset A and subset B. A consists of 183 horror movie scenes and 183 non-horror movie scenes. B consists of 182 horror movie scenes and 182 non-horror movie

**Table 1**. Results on Horror Movie Scene Identification

| Features | $P(\%)$ | $R(\%)$ | $F_1(\%)$ |
|---|---|---|---|
| $VF$ | 75.20 | 76.44 | 75.82 |
| $AF$ | 78.83 | 77.53 | 78.18 |
| $EF$ | 71.87 | 70.68 | 71.27 |
| $VF + AF$ | 80.22 | 80.00 | 80.11 |
| $VF + EF$ | 72.40 | 72.60 | 72.50 |
| $AF + EF$ | 79.30 | 80.82 | 80.05 |
| $VF + AF + EF$ | 80.16 | 84.11 | 82.14 |

scenes. In order to remove the correlation, we place the movie scenes which derive from the same movie in the same subset.

### 3.2. Results

As described in previous sections, a movie scene is represented by visual features ($VF$, 12D), audio features ($AF$, 52D) and color emotion features ($EF$, 23D). Seven different combinations of the features can be got from the three types of features. We use A for training and B for testing and vice versa. Then the SVM classifier with a RBF kernel is employed to recognize the horror movie scene in all cases of feature combination. Table 1 shows the results of horror movie scenes identification, where Precision ($P$), Recall ($R$) and F-measure ($F_1$) are used to evaluate the recognition performances.

Experimental results in Table 1 show the best one among three types of features is the audio feature, which has the highest F-measure. Generally, the features have the complementary characteristics except visual and color emotion features. As can be seen, when associated with color emotion feature or visual feature, the preference of audio feature's recognition is improved. The best preference of recognition can be obtained when visual features and audio features are combined with color emotion features and the highest F-measure is 82.14%. The results show that the color emotion features which we propose improve the preference of the horror movie scene recognition. As the initial work, the experimental results are promising.

## 4. CONCLUSION

In this paper, we have proposed an effective approach to solve the problem of horror movie scene recognition, which is an initial stage for Web horror information filtering. Color emotion and color harmony theories are introduced into the horror video scene recognition and we extract color emotion features which are high-level features. Experimental results show that color emotion features improve the horror movie scene recognition effectively. In future work, more horror movie scenes will be collected. Meanwhile, we will be dedicated to improving the horror film scene recognition algorithm by better modeling of features space and bettering the classifiers.

## References


[1] W. Hu, O. Wu, Z. Chen, Z. Fu, and S. Maybank, "Recognition of pornographic web pages by classifying texts and images," *IEEE transactions on pattern analysis and machine intelligence*, vol. 29, no. 6, pp. 1019–1034, 2007.

[2] N.J. King, G. Eleonora, and T.H. Ollendick, "Etiology of childhood phobias: Current status of Rachmans three pathways theory," *Behaviour research and Therapy*, vol. 36, no. 3, pp. 297–309, 1998.

[3] H.L. Wang and L. Cheong, "Affective understanding in film," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 16, no. 6, pp. 689–704, 2006.

[4] A. Hanjalic and L.Q. Xu, "Affective video content representation and modeling," *IEEE Transactions on Multimedia*, vol. 7, no. 1, pp. 143–154, 2005.

[5] Z. Rasheed, Y. Sheikh, and M. Shah, "On the use of computable features for film classification," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 15, no. 1, pp. 52–64, 2005.

[6] M. Xu, L.T. Chia, and J. Jin, "Affective content analysis in comedy and horror videos by audio emotional event detection," in *IEEE International Conference on Multimedia and Expo*, 2005, pp. 622–625.

[7] Z. Cernekova, I. Pitas, and C. Nikou, "Information theory-based shot cut/fade detection and video summarization," *IEEE Transactions on circuits and systems for video technology*, vol. 16, no. 1, pp. 82–91, 2006.

[8] J.M. Geusebroek and A.W.M. Smeulders, "A six-stimulus theory for stochastic texture," *International Journal of Computer Vision*, vol. 62, no. 1, pp. 7–16, 2005.

[9] A.J. Cohen, "Music as a source of emotion in film," *Music and emotion: Theory and research, ed. PN Juslin & JA Sloboda*, pp. 249–272, 2001.

[10] G. Tzanetakis and P. Cook, "Musical genre classification of audio signals," *IEEE Transactions on speech and audio processing*, vol. 10, no. 5, pp. 293–302, 2002.

[11] L.C. Ou, M.R. Luo, A. Woodcock, and A. Wright, "A study of colour emotion and colour preference. Part I: Colour emotions for single colours," *Color Research & Application*, vol. 29, no. 3, pp. 232–240, 2004.

[12] L.C. Ou, M.R. Luo, A. Woodcock, and A. Wright, "A study of colour emotion and colour preference. Part III: Colour preference modeling," *Color Research & Application*, vol. 29, no. 5, pp. 381–389, 2004.

[13] L.C. Ou and M.R. Luo, "A colour harmony model for two-colour combinations," *Color Research & Application*, vol. 31, no. 3, pp. 191–204, 2006.