Recognition of Film Type Using HSV Features on Deep-Learning Neural Networks
Ching-Ta Lu, Jia-An Lin, Chia-Yi Chang, Chia-Hua Liu, Ling-Ling Wang, and Kun-Fu Tseng


View online: https://doi.org/10.11989/JEST.1674-862X.90904223

Articles you may be interested in


Follow JEST WeChat public account for more information
Recognition of Film Type Using HSV Features on Deep-Learning Neural Networks

Ching-Ta Lu* | Jia-An Lin | Chia-Yi Chang | Chia-Hua Liu | Ling-Ling Wang | Kun-Fu Tseng*

Abstract—The number of films is numerous and the film contents are complex over the Internet and multimedia sources. It is time consuming for a viewer to select a favorite film. This paper presents an automatic recognition system of film types. Initially, a film is firstly sampled as frame sequences. The color space, including hue, saturation, and brightness value (HSV), is analyzed for each sampled frame by computing the deviation and mean of HSV for each film. These features are utilized as inputs to a deep-learning neural network (DNN) for the recognition of film types. One hundred films are utilized to train and validate the model parameters of DNN. In the testing phase, a film is recognized as one of the five categories, including action, comedy, horror thriller, romance, and science fiction, by the trained DNN. The experimental results reveal that the film types can be effectively recognized by the proposed approach, enabling the viewer to select an interesting film accurately and quickly.

Index Terms—Deep-learning; film type recognition; hue, saturation, and brightness value (HSV) analysis; neural networks; video classification.

1. Introduction

A great number of films can be easily browsed over the Internet and data stores. The film contents are complex. It is difficult for a viewer to distinguish the type of films from the film names. Viewers find it impossible to locate quickly a film they want to watch. Thus the viewer has to watch trailers first or use the skip mode to play films for selecting a preferable film. However, the selection process takes a lot of time. Therefore, designing an effective method for a viewer to select an interesting film type automatically is important.

Andrej et al.[1] conducted an extensive evaluation of the convolutional neural network (CNN) using large-scale film classification. They attempted to extend the connectivity of CNN in the time domain to use local spatial-temporal properties. The best performance can reach 63.9% for a large video data set. Video skims or key-frames can create video abstracts[2], where the main affective contents of film sequences can be represented by key-
frames. Analyzing an original film sequence is time-consuming. It can be much shortened by only extracting the film clips named as film skimming. The multimedia contents are interpreted as the human perception by using the film summarization, enabling the content-based retrieval systems to be able to index content effectively for the reduction of the storage space. Consequently, the desired contents can be efficiently accessed.

The development of automatic video analysis is important for the reduction of the manual process in film summarization\cite{5,6}, where the film summarization methods can be categorized into high and low level summarization methods\cite{3,4}. Dimitrova et al.\cite{5} proposed an approach to the film type recognition based on text and face trajectories. The hidden Markov model was employed to classify video clips into appropriate classes. Avila et al.\cite{6} proposed an approach based on the extraction of color features for each film frame and a clustering method for the production of film summaries. At first, the K-means algorithm is employed to classify film frames into groups in sequential order. Adequately selecting one frame per cluster can generate the film summary. Naphide and Huang\cite{3} proposed using probabilistic multimedia objects for semantic video indexing. These multimedia objects map low level features to high level semantics of a film. Hence, a Bayesian multi-network was employed for video classification. Furini et al.\cite{4} proposed a summarization technique for the production of on-the-fly film storyboards. The approach generates a moving-and-still storyboard with a quick clustering method. Descriptive visual frames can be selected by using the hue, saturation, and brightness value (HSV) color distribution. The visual features are extracted frame-by-frame to denote the visual contents as input sequences. Hence detecting the frame groups with similar scenarios and selecting one representative film frame from each group were conducted. Color styles make people distinguish the essence of film events\cite{8,9}, such as the irritability, drama tension, and established atmosphere. In [10], the relationship between the emotional atmosphere and hue was established. Red can represent hatred, life, love, and nobleness. Green can denote peace life. Blue can represent nobleness and calm. Yellow denotes clear and happiness, etc. Color has a huge impact on emotions and is a subconscious element in films\cite{11}. In [6], a dreadful atmosphere can be created by a dark tone with either the red tone or blue one, and the low saturation color. A sad feeling can be created by the saturation and darkness. The joyful atmosphere only needs bright colors. Rasheed et al.\cite{7} proposed using a framework to classify videos into genres based on visual cues, including average shot length, color variance, motion contents, and lighting keys. A mean shift classification method was employed to classify films into four categories, including comedies, action, dramas, and horror videos. Convolutional neural networks have been widely used in pattern recognition, such as image recognition and classification\cite{12}, object detection\cite{13,14}, scene labeling\cite{15,16}, etc. Andrej et al.\cite{17} proposed using local spatial-temporal information on CNN for video classification. The improved performance was obtained. Recently, some state-of-the-art methods have been developed for film classification\cite{18,19}. Huang and Wang\cite{18} proposed a movie genre classification system which uses a self-adaptive harmony search to select important features, including visual and audio features, for film genres. These features were uploaded into a support-vector-machine (SVM) for the film type recognition. Huang et al.\cite{19} proposed using a text-based approach for the classification of video types. The titles, descriptions, lexicon, syntax, comments, and specific content were extracted as features. These features were uploaded into SVM, Naive Bayes, and C4.5 classifiers for the recognition of film types. Liu et al.\cite{20} proposed using the color space, including red, green, blue, yellow, and HSV for the analysis of film frames. The deviation and mean of HSV are computed and utilized as classification features for each film. A complex rule-based method was proposed for the recognition of film types.

In this study, we propose analyzing film colors for the recognition of film types. Initially, a film is sampled to obtain an image sequence. Feature parameters play an important role in the performance of a classifier. Here, the averages, standard deviations of the color containments in terms of HSV of a film, are computed and utilized as features for the recognition of film types. One hundred films, which were arbitrarily retrieved from the YouTube website and produced from many countries and regions, including Germany, America, France, etc. were utilized to evaluate the proposed film type recognizer. Because the film features vary significantly, selecting a robust...
recognizer is important for the identification of film types. An effective classifier can improve significantly the system performance in film type recognition. A testing film is recognized by using a deep-learning neural network (DNN), which recognizes the test films as one of the five major categories, including horror thrillers, literature-love (romance), science fiction, comedy, and action films.

The recall and precision rates and F-measure are utilized to assess the recognition performance. The experimental results reveal that the proposed approach can automatically recognize the film type. The major difference between the proposed method and the others is the usage of DNN with complex film features to recognize the film type, rather than using a rule-based method. Although some HSV features are discriminative, they can be employed to recognize film types effectively by using the proposed approach and rule-based methods. On the contrary, some HSV features are not discriminative. They cannot be effectively employed to recognize the film type by using the rule-based method. But the proposed DNN can still apply them for film type recognition. Accordingly, the major contribution of this study is that using DNN to recognize the film types by the HSV features enables viewers can quickly recognize the desired film type from the YouTube website.

The rest of the paper is organized as follows. Section 2 describes the proposed DNN-based film type recognition approach. Section 3 shows the experimental results. Conclusions are finally drawn in Section 4.

2. Proposed DNN-Based Film Type Recognition Approach

The block diagram of the proposed film type recognition system is shown in Fig. 1. Initially, a film is sampled to obtain image frames. In turn, the color analysis is performed for each frame. The standard deviation and averages of HSV are calculated as the features for recognizing the type of each film trailer. The film features are uploaded into DNN for training a film DNN, named as film-DNN, which is utilized for the recognition of film types. In order to validate the film-DNN, the features of testing films are analyzed, extracted, and uploaded into the trained film-DNN, where the feature extraction method is the same as the training phase. The type of a test film is then recognized. The film is recognized as one of five categories, including literature-love (romance) $C_r$, science fiction $C_s$, comedy $C_c$, action $C_a$, and horror thrillers $C_h$.

2.1. Recognition Features

A film is firstly sampled to obtain images sequentially. The color of each image is changed from the red, green, and blue (RGB) space to the HSV space. Hence the average values of the hue $\mu_H$ are computed by

$$\mu_H = \frac{1}{N} \sum_{i=0}^{N-1} H(i)$$

(1)

Fig. 1. Block diagram of the proposed recognition system for film type recognition.
where $N$ denotes the number of sampled frames for a film and $H(i)$ is the value of the hue at the $i$th sampled frame.

Fig. 2 demonstrates the distribution of $\mu_H$ for a film. It can be found that the mean values of the hue for horror thrillers and science fiction films are larger, whereas for other films they are smaller. Accordingly, the film types can be separated into two groups according to $\mu_H$. The first group is horror thrillers and science fiction films; the second group is comedy, literature-love, and action films.

The standard deviation of the hue $\sigma_H$ measures the dispersion of hues in a film relative to its mean $\mu_H$. It is an important feature for film recognition and can be computed by

$$\sigma_H = \sqrt{\frac{1}{N-1} \sum_{i=0}^{N-1} [(H(i) - \mu_H)^2]}.$$  \hspace{1cm} (2)

Fig. 3 demonstrates the distribution of $\sigma_H$. The values of $\sigma_H$ for science fiction and literature-love films are similar, so the film types are separated into two groups according to $\sigma_H$. The first group is the science fiction and literature-love films, whereas the second group contains horror thrillers, comedy, and action films.

The computation method for the mean value of color saturation $\mu_s$ is similar to (1). Fig. 4 demonstrates the mean values of color saturation $\mu_s$. It is clear that the action film is different from the others, so it can be classified independently. The others are regarded as the second group.

The ratio of color brightness $v_{ratio}$ can be employed as a feature for the recognition of film types, it can be computed by

$$v_{ratio} = \frac{1}{N_c} \sum_{i=0}^{N_c-1} v(F'(i))$$  \hspace{1cm} (3)

where $N_c$ denotes the number of the brightness levels. $v$ represents the brightness level which falls within a specified range, i.e., $\sqrt{v}$ $\leq v$ $\leq \sqrt{v'}$. $F'(i)$ represents a flag to indicate whether the brightness of the $i$th sampled frame falls within the specified range. $F'(i)$ can be expressed by
where the lower-bound \( v' \) and upper-bound \( v'' \) of the brightness levels are given as
\[
v' = \mu_v - k\sigma_v \\
v'' = \mu_v + k\sigma_v
\]
where \( \sigma_v \) and \( \mu_v \) represent the standard deviation and mean of the brightness. They can be computed according to the similar manner as given in (1) and (2). And \( k \) controls the dynamic range between \( v'' \) and \( v' \), it is empirically chosen equal to one.

In (3), the number of the brightness levels which fall within the specified range \( N_v \) can be computed by
\[
N_v = \sum_{i=0}^{N-1} F_v(i).
\]

Fig. 5 demonstrates the distribution of \( v_{\text{ratio}} \) for various films. Many \( v_{\text{ratio}} \) distributions are overlapped. Thus, the feature \( v_{\text{ratio}} \) is not discriminative for film type recognition. The mean value of lightness \( \mu_v \) can be computed according to (1). Fig. 6 demonstrates the distributions of the mean value of lightness \( \mu_v \). It can be found that literature-love and comedy films obtain similar distribution and the highest mean brightness, while the brightness distributions of the horror thrillers, science fiction, and action films are similar, thus, \( \mu_v \) is utilized as a robust feature for film type recognition. \( \mu_v \) is employed to classify films into two groups. The first group contains comedy and literature-love films, while the horror thrillers, science fiction, and action films belong to the second group. The computation method for the standard deviation of brightness \( \sigma_v \) is similar to (2). From Fig. 7, it can be found that the values of \( \sigma_v \) for literature-love films are significantly higher than the values for other films, hence, literature-love films can be successfully separated from other films by using \( \sigma_v \).

From Figs. 2 to 7, it is clear that each feature is discriminative to some types of films. The type of a film can be recognized by using a rule-based method, where the threshold of each feature should be carefully defined. Otherwise, the recognition accuracy would reduce significantly. Because a film consists of rich colors, defining a robust threshold for each color feature to distinguish the film type is very difficult. Accordingly, DNN is applied to map the rich color features and film types, enabling the film type to be efficiently identified.
Some additional features related to HSV also provide significant discriminability for the recognition of film types, including the average value of the hue ($\mu_H$), the standard deviation of the hue ($\sigma_H$), the standard deviation of color saturation ($\sigma_S$), the ratio of brightness ($v_{ratio}$), the standard deviation of brightness ($\sigma_v$), and the mean value of lightness ($\mu_L$). In addition, the ratio of color saturation ($S_{ratio}$) and the standard deviation of color saturation ($\sigma_S$) also provide slight discriminability among various types of films. Feeding these features into DNN can improve the recognition performance. Accordingly, the feature vector for the recognition of film types can be expressed by

$$\Psi = \begin{bmatrix} \mu_H & \sigma_H & \mu_S & \sigma_S & \mu_v & \sigma_v & S_{ratio} & v_{ratio} \end{bmatrix}.$$  (8)

The feature vectors given in (8) are utilized for the recognition of film types in the experiments.

### 2.2. Film Type Recognition

The mean and standard deviation of HSV given in (8) are computed as a feature vector for the recognition of film types. Each film provides a set of feature vectors. The feature vectors obtained by (8) are uploaded into DNN. At first, the number of layers in DNN is fixed to two. The number of neurons is increased to find the best recognition performance. The best result is obtained when the number of neurons is twenty. In turn, we use the same manner to find the number of layers, while the number of neurons is set to twenty for the input layer and each hidden layer. The best recognition result is achieved when the number of layers is five. Accordingly, we form a film-DNN with five layers shown in Fig. 8, while twenty neurons are utilized for the input layer and each hidden layer. The feature vectors of the training films obtained by (8) are uploaded into the film-DNN for training the weights and bias parameters. The soft-maximum activation function is utilized for each neuron.

The target of the film type is labeled manually. The supervised learning method is conducted for training the film-DNN. The target film type and the corresponding training features obtained by (8) are uploaded into the film-DNN sequentially. The mean-square-error is utilized as loss function, while the epochs and batch size are empirically chosen to be 10000 and 10, respectively. The extraction manner of the feature vectors of testing films is also the same as that of the training set. The features of testing films are uploaded into the trained film-DNN. The top score of the film-DNN outputs represents the recognized type.

### 3. Experimental Results

In the experiments, we employed one hundred films for training and testing the film-DNN. The films were
arbitrarily retrieved from the YouTube website and produced by many countries and regions, including Germany, America, France, etc. Each film belongs to one of the five categories, i.e., horror thrillers, science fiction, action, literature-love (romance), and comedy films. Every category includes twenty-five films, so the number of films is evenly balanced. In the training phase, we employed 75 films as the training set for training the film-DNN. The remainder of the 25 films is utilized as the validation set, which is used to evaluate the performance of the proposed film-DNN for the recognition of film types.

Initially, we employed 25 films for training and testing the film-DNN as a preliminary experiment. Fig. 9 demonstrates the relationship between the epochs and loss value for training the film-DNN. As the epochs number increases, the value of the loss decreases. We can find that the value of the loss function cannot be further reduced significantly if the number of epochs exceeds 4000. Therefore, the number of epochs is chosen to be 4000 in the experiments. Fig. 10 demonstrates the relationship between the number of epochs and the accuracy rate in training the film-DNN. The performance is further improved when the number of neurons increases for a layer. If the number of neurons is greater than twenty units in a layer, the performance cannot be further improved. Accordingly, the number of neurons is selected to be twenty for each layer in the experiments.

Fig. 11 demonstrates the scores of recognizing the film type by using the film-DNN. The best recognition scores in the categories of horror thrillers, action, science fiction, and literature-love (romance) films are much greater than those of the other film types. Employing the deviation and mean of HSV as feature parameters is suitable for the recognition of film types.

The recall rate, precision one, and F-measure are employed to assess the recognition performance of the proposed film-DNN. The precision rate of film type recognition is defined as

\[
\text{\( p\% = \frac{\text{The number of correctly recognized film types}}{\text{The number of recognized films}} \times 100\%. \) (9)}
\]

In (9), the score of precision rate \( p \) (in \%) is high when the number of correctly recognized film types increases. Conversely, the precision rate \( p\% \) gets a low value if the number of correctly recognized film types decreases.

The fraction of films, whose types are correctly recognized, can be reflected by the recall rate, it is defined by

\[
\text{\( r\% = \frac{\text{The number of correctly recognized film types}}{\text{The number of films}} \times 100\%. \) (10)}
\]
In (10), it can be found that the higher the number of correctly recognized film types, the larger the score of the recall rate $r\%$, which represents superior performance for the recognition of film types. The F-measure can be regarded as a measure to simultaneously consider the overall performance of a film type recognition system, we employ the F-measure for assessment, given as
\[
F\text{-measure} = \frac{2rp}{r+p} \times 100\%.
\] (11)

In (11), the score of the F-measure obtains a high score if the scores of recall rate $r\%$ in (10) and accuracy rate $p\%$ in (9) are both high. Conversely, the score of the F-measure decreases when the scores of the recall rate $r\%$ and accuracy rate $p\%$ are low. Accordingly, the F-measure is utilized to evaluate the overall performance of the proposed film-DNN for the recognition of film types. The performance in terms of the recall rate $r\%$, accuracy rate $p\%$, and F-measure are presented in Table 1.

In the performance of precision rate, the types of the horror thrillers, science fiction, and comedy films can be correctly recognized with precision rate equaling 100%. Most film types of literature-love and action films are also correctly recognized. The proposed film-DNN can successfully classify horror thrillers, action, science fiction, and literature-love films. The recall rate of these types approaches 100%. The overall performance in terms of the F-measure for the proposed film-DNN is greater than 90.91% for most film types, in particular for the types of horror thrillers and science fiction it reaches 100%. Consequently, the performance of the proposed film-DNN is acceptable.

By observing the performance of film type recognition presented in Table 1 and Fig. 11, the proposed approach can effectively recognize the film types. Therefore, using the standard derivation and mean of HSV, and the ratios of brightness and saturation as training and validation features given in (8) for the film-DNN one can well recognize the film types. Finally, all films were separated into a training set of 75 films and validation set of 25 films to test the performance of the film-DNN. Each type has twenty films as a training set and five films as a validation set. The number of films for each type is evenly balanced and distributed over training and testing sets. Because a film contains plentiful contents, the film could belong to two or more film types. Accordingly, we evaluated the performance of film type recognition by using the Top 1, Top 2, and Top 3 of the recognized results, where the Top 1 presents the most similarity category. The recall rates are 52%, 76%, and 100% for using Top 1, Top 2, and Top 3 criteria, respectively. These results mean that recognized results can completely fit the target film type by using the Top 3 criterion.

### Table 1: Comparison of recognition results for various film types in terms of precision rate, recall rate, and F-measure

<table>
<thead>
<tr>
<th>Measures</th>
<th>Film type</th>
<th>Score (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precision</td>
<td>Action</td>
<td>83.33</td>
</tr>
<tr>
<td>Recall</td>
<td>Comedy</td>
<td>100.00</td>
</tr>
<tr>
<td></td>
<td>Literature-love</td>
<td>83.33</td>
</tr>
<tr>
<td></td>
<td>Science fiction</td>
<td>100.00</td>
</tr>
<tr>
<td></td>
<td>Horror thrillers</td>
<td>100.00</td>
</tr>
<tr>
<td>Recall</td>
<td>Action</td>
<td>100.00</td>
</tr>
<tr>
<td>Recall</td>
<td>Comedy</td>
<td>60.00</td>
</tr>
<tr>
<td>Recall</td>
<td>Literature-love</td>
<td>100.00</td>
</tr>
<tr>
<td>Recall</td>
<td>Science fiction</td>
<td>100.00</td>
</tr>
<tr>
<td>Recall</td>
<td>Horror thriller</td>
<td>100.00</td>
</tr>
<tr>
<td>F-measure</td>
<td>Action</td>
<td>90.91</td>
</tr>
<tr>
<td>F-measure</td>
<td>Comedy</td>
<td>75.00</td>
</tr>
<tr>
<td>F-measure</td>
<td>Literature-love</td>
<td>90.91</td>
</tr>
<tr>
<td>F-measure</td>
<td>Science fiction</td>
<td>100.00</td>
</tr>
<tr>
<td>F-measure</td>
<td>Horror thriller</td>
<td>100.00</td>
</tr>
</tbody>
</table>

![Fig. 11. Recognition scores of film-DNN for various films.](image-url)
As shown in Fig. 11, the recognition accuracy of the comedy films is not as good as the other films. The error is caused by the romance films, which obtains the highest scores at the output of the film-DNN for the target film being a comedy. Due to the properties of the romance and comedy films being similar, the color styles of these two film types are also similar, because the HSV properties of films vary significantly. The color style also depends on the director of the film. It is hard to obtain a high accuracy rate for the recognition of film types, in particular for a great number of films. Accordingly, the proposed method of film type recognition can be regarded as an auxiliary tool for users to quickly obtain their interesting films.

Table 2 presents the performance comparison for the rule-based method\cite{lu2015recognition} and proposed film-DNN. These two methods utilized the same features as given in (8). The rule-based method creates an analysis tree for the film type recognition and requires to set appropriate thresholds for each feature according to the distribution characteristics of films. The performance is very sensitive to the selected thresholds, so the thresholds are difficult to be defined. On the contrary, the proposed film-DNN can learn the statistical properties of film colors. It is not needed to define the threshold of each feature. Thus, the proposed film-DNN is more effective and robust than the rule-based method. Although the recall rate of the rule-based method can reach 100%, the precision rate can reach only 83.33%, which is not as high as the recall rate. The performance of these two measures are not balanced. The overall quality in terms of F-measure is 90.71%. On the contrary, the precision and recall rates of the proposed film-DNN are 93.33% and 92%, respectively. The performance of these two measures are evenly balanced. The overall quality in terms of F-measure of the proposed film-DNN (91.36%) is better than the rule-based method (90.71%). Consequently, the proposed film-DNN is better and more robust than the rule-based method.

4. Conclusions

This paper presents using the deviation and mean of HSV as the features of film-DNN for the recognition of film types. The film-DNN can recognize different film types, including horror thriller, action, science fiction, comedy, and literature-love (romance) films. Experimental results reveal that the proposed film-DNN can recognize the type of a film effectively. In future work, we will try to find additional robust features for DNNs, enabling the performance of film type recognition to be further improved. Moreover, the number of films in the training and validation sets should be further increased. It will be beneficial to ensure the performance of film type recognition. Consequently, the proposed method can be utilized in an automatic recognizer of film types for the library and web-movie server, such as YouTube. The proposed system for film type recognition can help a viewer to select interesting films from rich databases efficiently.

References

Ching-Ta Lu received both B.S. and M.S. degrees in electronic engineering from National Taiwan University of Science and Technology, Taipei in 1991 and 1995, respectively; and his Ph. D. degree in electrical engineering from National Tsing Hua University, Hsinchu in 2006. He has been a full professor with the Department of Information Communication, Asia University, Taichung since Feb. 2008. He has published more than 40 journal papers and 70 conference papers. His current research interests include speech enhancement, image denoising, speech coding, and speech signal processing.
Jia-An Lin received both B.F.A. and M.F.A. degrees from Chaoyang University of Technology, Taichung in 2002 and Tainan National University of the Arts, Tainan in 2007, respectively. Currently, he is an associate professor with the Department of Digital Media Design, Asia University. He has been the Chairman of this department since 2016. And he is pursuing the Ph.D. degree in art creation and theory with Tainan National University of the Arts. His research interests include photographic and video recording practice, digital media industry, documentary production, advertising film production, and postcolonial film studies.

Chia-Yi Chang is pursuing his B.S. degree with the Department of Information Communication, Asia University. His research interests include artificial intelligence and application software programming.

Chia-Hua Liu received her B.S. and M.S. degrees from the Department of Information Communication, Asia University. Currently, she is a research assistant with the Department of Information Communication, Asia University. Her research interest includes video content analysis.

Ling-Ling Wang received the B.S., M.S., and Ph.D. degrees in computer science and information engineering from National Chiao Tung University, Hsinchu in 1984, 1986, and 1990, respectively. From 1986 to 1987, she was an associate engineer with the System Software Department, ERSO, ITRI, Hsinchu. From 1991 to 1997, she was an associate professor in computer science with National Tsing Hua University. Currently, she is a professor with the Department of Information Communication, Asia University. Her major research is in image processing and artificial intelligence.

Kun-Fu Tseng received his Ph.D. degree in electronic engineering from NDU, Taoyuan in 1998. He then joined the Faculty of the Multimedia and Game Science, Asia-Pacific Institute of Creativity, Miaoli in 2000. Currently, he is with the School of Electronics and Communication Engineering, Quanzhou University of Information Engineering, Quanzhou. His research areas focus on the high frequency electrical property analysis and simulations on 3-dimensional models, and signal processing.