A Novel Horror Scene Detection Scheme on Revised Multiple Instance Learning Model

Bin Wu[1], Xinghao Jiang[1,2,*], Tanfeng Sun[1,2], Shanfeng Zhang[1], Xiqing Chu[1], Chuxiong Shen[1], and Jingwen Fan[1]

[1] School of Information Security Engineering, Shanghai Jiao Tong University
[2] Shanghai Information Security Management and Technology Research Key Lab
{benleader,xhjiang,tfsun,may3feng,xqchu,bear0811,DENISEFAN}
@sjtu.edu.cn

Abstract. Horror scene detection is a research problem that has much practical use. The supervised method requires the training data to be labeled manually, which can be tedious and onerous. In this paper, a more challenging setting of the problems without complete information on data labels is investigated. In particular, as the horror scene is characterized by multiple features, this problem is formulated as a special multiple instance learning (MIL) problem – Multiple Grouped Instance Learning (MGIL), which requires partial labeled training. To solve the MGIL problem, a learning method is proposed – Multiple Distance- Expectation Maximization Diversity Density (MD-EMDD).Additionally, a survey is conducted to collect people's opinions based on the definition of horror scenes. Combined with the survey results, Labeled with Ranking – MD – EMDD is proposed and demonstrated better results when compared to the traditional MIL algorithm and close to performance achieved by supervised method.

Keywords: Horror Scene Detection; Multi-Instance Learning; Machine Learning.

1 Introduction

With the evermore rapid development of the Internet and the prevalence use of video capture devices, it is now much more convenient for people to upload their videos or movies. However, some of them may contain horror content, and parents are concerned about the negative influence of such scenes on children. Therefore, an effective automatic filtration algorithm for horror scenes is needed.

There has been some related work for scene detection. W.J. Gillespie et al. used RBF network to classify videos [1]. Hana, Ricardo O applied the conventional MLP (Multiple Layer Perceptron) Neural Network and SVM (Support Vector Machine) to classify crime scenes [2]. Simon Moncriefi et al. study on the affect computing in film through sound energy dynamics [3]. Xinghao Jiang et al. adapted Second-Prediction strategy to video pattern recognition [4]. Zhiwei Gu et al. [5] proposed a Multi-Layer Multi-Instance Learning method for video concept detection.

* Corresponding author.

K.-T. Lee et al. (Eds.): MMM 2011, Part II, LNCS 6524, pp. 377–388, 2011.

However, there are some deficiencies of previous work. First, in [1][2][3][4], the supervised method was adapted. In order to achieve high accuracy, the supervised method requires every instance to be labeled. Therefore, the labeling work asks for *more input of effort* before the training process. Moreover, since general horror scenes are not distributed evenly in videos, the supervised method - *detecting the video as an instance* -does not work effectively in such circumstance. Second, in [3], the authors mainly discussed the sound effect of horror scenes and did not take visual features into consideration. Also, the dataset contains only 4 films. Third, in [5], the proposed method were to detect video concepts and not directly applicable in horror detection (will be demonstrated in part 2). Fourth, the *relationship* among multiple features in horror scenes has seldom been discussed.

Therefore, in order to make the labeling work more efficient, this paper seeks to investigate a more challenging setting of the problems, in which the information of data label is incomplete – Multiple Instance Learning (MIL). What's more, the relationship among horror scenes' features is discussed. Since different features probably differ in their contribution to characterize horrible atmosphere, we devise the MIL and propose- *Multiple Grouped Instance Learning* problem (MGIL).Then, a new learning method – *Multiple Distance – EMDD* (MD-EMDD) is proposed based on EM-DD [6] to tackle the MGIL problem, especially the horror scene detection problem. Besides, a survey is conducted to collect the surveyees' scores on all kinds of video. The MD-EMDD is then combined with the scores ranking and *Labeled with Ranking– MD – EMDD* is proposed.

The rest of this paper is organized as follow: In section 2, the horror scene detection as MIL problem is discussed. Then, in section 3, 2 learning methods are proposed based on EM-DD for horror scene detection problem: Multiple Distance – EMDD (MD-EMDD) and Labeled with Ranking– MD – EMDD. Finally, the proposed method is applied to the horror scene detection problem. The experiments in section 4 show that the performance of our method achieves a better result compared to previous MIL methods and close to performance achieved by supervised method.

2 Horror Scene Detection as MIL Problem

MIL is a set of algorithms designed to solve the multiple instance problems, where instances are packaged in a whole bag. A bag is defined as positive if there is *at least* one positive instance in it, and negative if *all* instances in it are negative.

MIL was first proposed by Dietterich et al. [6] in order to solve the drug activity prediction problem. More recent studies have extended MIL from drug problem to other areas. For example, Zhiwei Gu et al. [5] proposed a Multi-Layer Multi-Instance Learning method for video concept detection. Rouhollah Rahmani et al. [7] presented a localized CBIR system using MIL.

Many horror videos consist mainly of non-horror scenes, while the rest are very horror-intensive and may have profound negative influence to children. Figure 1.gives an example. It is a typical video that the effect only takes place in the last 2-3 seconds. This video has been selected in our survey (refer to part 5.2) and rated first by the surveyed subjects. Thus, the goal to detect horror videos is to try and find a part that has the highest matching possibility to the definition of horror. Then, the video can be labeled according to the probability.

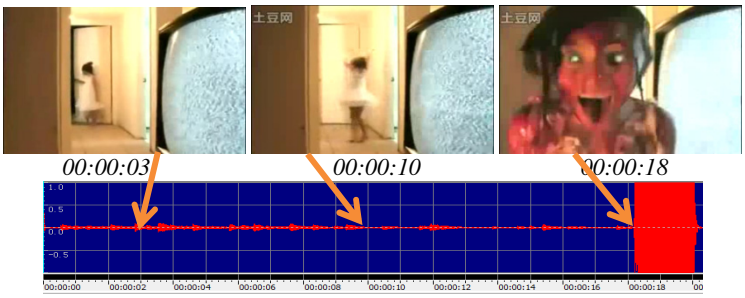

Fig. 1. The three screen shots are from a typical horror video. Form the beginning to *00:00:17,* there is a lovely girl dancing gracefully in the middle of the screen (as the first two screen shots show). However, at *00:00:18,* a horrible figure suddenly jumps into the screen, with a sudden scream. Below the screen shots is the corresponding audio track wave. Three orange arrow-heads indicate the corresponding positions of the three screen shots in the track wave.

In practice, videos are segmented into scenes. Now suppose a video have been segmented into scenes. The probability of a video to contain horror contents can be expressed as:

$$max\ P(L = 1|s_i, h),\ \text{i}\epsilon(1, \text{K}) \tag{1}$$

Where $S = \{s_1, \ldots, s_K\}$ represents the K scenes of the video, $\text{L}\epsilon(0, 1)$ is a binary label indicating whether the video is horror or not. is the hypothesis which consists of a set of model parameters to be determined. The above is the analysis of the "video - scene" structure of a horror video. Next a single scene will be discussed.

(a) (b) (c)

Fig. 2. The three screen shots describe some typical features of horror scenes. (a) is a horror figure that suddenly appears in the screen, which leads to intensive motion level; (b) is a screen shot with the dominant color of blood-red, which is one of the commonly used colors in horror scenes; (c) is the audio track wave of a scene, where the suddenly increased part is caused by a woman's scream.

There is no authoritative definition for horror scenes. However, some points of view are widely accepted: 1) The sound track of a horror scene often has a direct say on the impact of the visual component of the film, such as screaming, sudden change of volume and the change in sound energy intensity over time. 2) Color is also special. In horror scenes, the colors are usually dark green, blood-red, etc. 3) Motion intensity is another feature. Directors usually create horror effects by a set of shots with high motion intensity. 4) Horror figure is also important in horror scenes. For example, Regan MacNeil (The Exorcist), Chucky (Child's Play), Jigsaw (Saw) etc.

Figure 2.gives 3 examples of the typical features of horror scenes.

As is shown, each feature helps characterize a horror scene. Therefore, labeling horror scenes is a process to train and predict the features. With the features extracted, the horror scene labeling can be formulated as a machine learning problem: to estimate the probability of a given scene to rated horror.

Suppose$X_i = \{ x_{i1}, \dots, x_{iN} \}$ is a set of features that extracted from the i^{th} scene. The probability that a feature matches the horror genre can be expressed as (suppose it is the j^{th} feature):

$$P\big(L = 1|x_{ij}, h\big) \tag{2}$$

Suppose there exists a mapping $F = \{f_1, \dots, f_N\}$, and define an operation "+" which satisfies:

$$P(L = 1|X_i, h) \; = \; f_1(P(L = 1|x_{i1}, h)) \; + \; \dots \; + \; f_N(P(L = 1|x_{iN}, h)) \tag{3}$$

As long as an F and "+"can be found to make the prediction error decrease, the F and "+" can be considered as reasonable. Therefore, the cores of the above formulation are: 1) The learning of . 2) To find a suitable F and "+".

As for the first point, the supervised method is most commonly adapted, which requires each scene needs to be labeled as accurate as possible instead of labeling the whole piece of video. The process costs much effort and time. In addition, since labeling scenes is more ambiguous than labeling videos (a video can be labeled positive if it contains horror scene), labeling to scenes by users' own judgments may also introduce some ambiguity. Therefore, if the videos can be labeled directly, it may make the detection technique more practical and reduce the ambiguity introduced by labeling instances.

Suppose video V consists of K scenes. $L = \{ l_1, \dots, l_K \}$ represents the K scenes' labels. According to the analysis above, V's label can be expressed as:

$$L_V \; = \; l_1 \cup \dots \cup l_k \tag{4}$$

Intuitively, it matches the multiple instance learning (MIL) problem: a set of scenes is grouped into a bag – video, and each scene is an instance. Therefore, the horror scene detection problem can be formulated into the labeling bag problem in MIL.

Since MIL can be applied to solve the "video-scene" level problem, now consider the possibility to solve the "scene-feature" level problem, as stated in [4]. Thus, let's consider:

$$P(L = 1|X_i, h) \; = \; f_1(P(L = 1|x_{i1}, h)) \cup \dots \cup f_N(P(L = 1|x_{iN}, h)) \tag{5}$$

However, as stated above, horror scenes are characterized by multiple features. Therefore, applying "$\cup$" may not achieve a good result (as demonstrated it in part 5).

One solution is to concatenate features as a new one. In this case, there is only one feature in (2) -X_i. However, this method ignores the possibility that each feature may contribute differently to the horror effect.

According to the discussion above and (4)'s idea, the F and "+" can be expressed as:

$$P(L = 1|X_i, h) = \sum_{l=1}^{N} \mu_l * P(L = 1|x_{il}, h), \sum_{l=1}^{N} \mu_l = 1 \tag{6}$$

In the discussion above, a three-level-model "video-scene-feature" has been set up. However, in practice, a scene consists of features and no single entity exists as "scene". Therefore, the horror scene detection model should be "video-feature", which can be expressed as:

$$max\ P(L = 1|s_i, h) = max\ \sum_{l=1}^{N} \mu_l * P(L = 1|x_{il}, h) \tag{7}$$

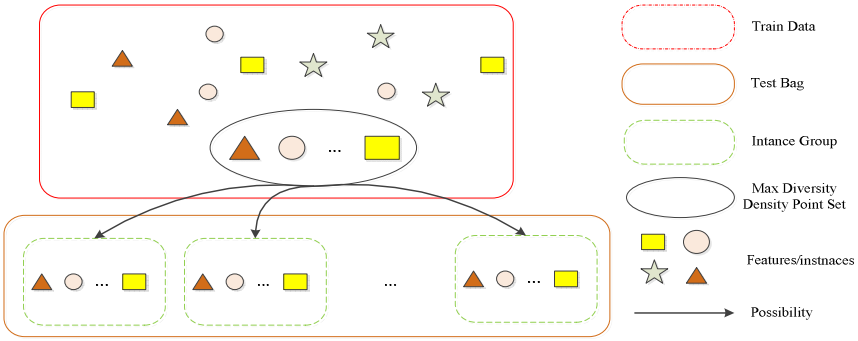

Fig. 3. The formulation of horror scene detection as an MIL problem

Apparently, if traditional MIL is adapted, then a video will be detected when a type of feature matches the horror definition, which may render a big error to ignore other instances in the same group. Here, an instance in the video is a type of feature. A scene consists of *a group of instances*. Figure 3.shows the model of *"bag-group-instance"*.

Thus, we believe that the proposed problem is a special MIL problem: There are a number of *instance groups* in a bag. A bag will be labeled positive only if at least *one positive group* exists, and its *instances as a group* matches the horror definition. Therefore, this special MIL problem is coined as *Multiple Grouped Instance Learning* (MGIL). In view of this, a method is proposed based on EMDD: *Multiple Distance – EMDD*. Specially, when$\mu_1 = \mu_2 = \cdots = \mu_N = 1/N$, the MD-EMDD is termed *non-weighted* MD-EMDD. Furthermore, the scores collected on the survey and integrated into the training data can be combined with the score information with the MD-EMDD method and propose *Labeled with Ranking – MD – EMDD* (LR-MD-EMDD).

3 Learning Method

3.1 Multiple Distance – EMDD (MD-EMDD)

EM-DD [8] is a multiple-instance (MI) learning technique. It combines EM with the Diverse Density (DD) [9] algorithm. It is relatively insensitive to the number of relevant attributes in the data set and its running time does not change much when bags size gets larger. EM-DD is used in a single feature space for traditional MIL problem, and it is intuitive that EM-DD cannot be directly used for MGIL problem. Therefore, a special method based on EM-DD: Multiple Distance – EMDD (MD-EMDD) is proposed.

In MD-EMDD, the maxDD point of each feature space is first calculated by EM-DD. Then, since each instance group consists of different feature values corresponding to different feature space, the Euclidean Distance cannot be computed as [8] does. In order to find the threshold, three steps should be followed:

1. Suppose there are N feature spaces, and the importance of each features is the same. For the i^{th} instance group and j^{th} feature, the Euclidean Distance from j^{th} feature in the instance group to the j^{th} feature space's maxDD point is computed:

$$d_{ij} = ED\left(x_{ij} maxDD_j\right) \tag{8}$$

   Here, ED is the function of normalized Euclidean Distance computing. Then define the distance from i^{th} instance group to the maxDD point as:

$$D_i = \sum_{j=1}^{N} \mu_j * ED\left(x_{ij},\ \ maxDD_j\right), \sum_{l=1}^{N} \mu_l = 1 \tag{9}$$

2. Under this $\mu = \{\mu_1, \dots, \mu_N\}$, a corresponding threshold can be obtained by 10-fold cross validation, which minimize the average error (number of wrongly predicted labels)among training bags labels $L = \{l_1, \dots, l_K\}$ and the predicted training bags labels $P = \{p_1, \dots, p_K\}$:

$$t = \arg min_{t \in D}\ avg\ error(L, P) \tag{10}$$

$$p_k = \begin{cases} 1, & D_k < t \\ 0, & otherwise \end{cases}, k \in \mathrm{K} \tag{11}$$

   This threshold is used to predict the training bags' labels and obtain an error, denote$e$as the error:

$$e = error(L, P) = \sum_{m=1}^{K} L_m \oplus P_m \tag{12}$$

3. Optimize $\mu = \{\mu_1, \dots, \mu_N\}$, and back to step 2. When average e gets to the minimum, the corresponding μ is the result needed. Name the μ as μ' and the corresponding threshold as t'.

To predict a bag, as is shown in step1, the Euclidean Distance should be computed from each feature in each instance group to every feature space's maxDD point. Then, as step3, a modified Euclidean Distance is obtained. Suppose there are M instance groups in a test bag:

$$D_t = min_{i \in M} \sum_{j=1}^{N} \mu'_j * ED\left(x_{ij} maxDD_j\right) \tag{13}$$

Then, compare it with the threshold found in step3 in order to predict the test bag's label:

$$p_{T\mathrm{est}} = \begin{cases} 1, & D_{Test} < t' \\ 0, & otherwise \end{cases} \tag{14}$$

3.2 Labeled with Ranking– MD – EMDD (LR-MD-EMDD)

Specially, if the bags with more information like scores of horror can be described, the distances and the scores ranking can be connected. First, for the K videos, $R = (r_1, \dots, r_K)$ is the ranking sequence of the videos' scores, where r_i denotes the rank of the i^{th} video. Second, for all the videos, apply the step 1 and step 2 of MD-EMDD. Denote the distance set obtained from eq.(9) as $\boldsymbol{D} = \{D_1, \dots, D_K\}$. Then, according to $\boldsymbol{D}$, a new ranking sequence $R' = (r_1', \dots, r_K')$ is obtained.Third, denote q as the number of the videos whose ranking differ in $R$from R'.Then, e is redefined as:

$$e = avg\ error(R, R') = \frac{\sum_{m=1}^{K}(r_m - r_m')^2}{q} \tag{15}$$

Fourth, apply step 3 of MD-EMDD and get optimized μ. Since bags are labeled with not only binary "is/not" labels but also ranking of scores, the method is called as *Labeled with Ranking – MD – EMDD* (LR-MD-EMDD).

4 Experiments

In this section, the training and testing dataset, and the experiment set-up for the horror scene detection problem including how features are selected are described. Then, the performance of the proposed learning methods is presented and compared with other previous methods.

4.1 Dataset

In order to generate positive bags, the comments on the horror movies and "Top 50 horror movie" ranked by famous movie website [10] were taken into consideration. Up to 50 classic horror movies like "The Exorcist", "Saw", "Kill Bill", etc. are selected. The 50 movies are segmented into 1200 pieces of video. In fact, MGIL only consider the most possible instance group in a bag, as long a video (bag) contains a number of scenes (instance groups), it will be detected. Therefore, the detection result will not change much as a video's length increases. Furthermore, in order to mine human's sense about horror, a survey on horror video is conducted.

First, 300 videos are briefly selected, all of which are labeled "horror" or "not horror" by the surveyed subjects. Moreover, 50 out of the 300 videos are chosen to be scored by audiences for their "horror" levels. Specifically, due to the practicability of the survey, the 50 videos are divided into 10 groups and each group consists of 5 videos (Please see 3.2 and 4.2). The subjects will be randomly given a group of videos to score and they are required to score all the videos in the group, which guarantees that the same subject will score all videos in a group. Only if this prerequisite is satisfied could the videos be compared with each other in the same group. The survey has lasted for about 4 months and has received 6920 pieces of valid questionnaires.

As for the negative bags, 300 videos of multiple topics are chosen. For example, talk show, news broadcasting, sports, games, humor show, music TV and so on. 40

out of the 300 are selected as training data, while the other is test data. The total 600 videos are automatically segmented into 5825 scenes (instance groups).

4.2 Experiment Set-Up

First, the EM-DD method is used to test the video descriptors' accuracy. 19 visual and audio MPEG-7 descriptors were adopted, as well as 1 motion level descriptor, which is designed to examine the contents of consecutive images. For the 20 descriptors, data is trained and tested under each feature space. Please note that all descriptors use the same training and testing data. Table 1 summarizes the result.

Table 1. Accuracy of each descriptor using EM-DD (including 19 MPEG-7 descriptors and 1 motion level descriptor). The adopted descriptors are stressed in bold type.

Descriptor	Accuracy	Descriptor	Accuracy
Dominant Color	0.540	Audio Spectrum Distribution	0.483
Color Layout	0.575	**Audio Spectrum Spread**	0.598
Color Structure	0.737	Background Noise Level	0.510
Scalable Color	0.471	Band Width	0.532
Homogeneous Texture	0.534	Dc Offset	0.575
Edge Histogram	0.579	Harmonic Spectral Centroid	0.537
Audio Fundamental Frequency	0.602	Harmonic Spectral Deviation	0.571
Audio Harmonicity	0.542	Harmonic Spectral Spread	0.552
Audio Signature	0.591	Harmonic Spectral Variation	0.526
Audio Spectrum Centroid	0.591	**Motion Level**	0.591

As is shown in Table 1, 5 descriptors are adopted in the following experiments. For comparison purpose, the performances of 5 algorithms are evaluated: non-weighted MD-EMDD,MD-EMDD and LR-MD-EMDD are proposed for MGIL problem; EM-DD is widely used traditional MIL algorithms; support vector machine (SVM) is a supervised learning algorithm. The first 3 algorithms work only with the bag labels while the SVM work in supervised setting with all instances labeled. The training dataset for SVM is selected from the videos' scenes and labeled by us.

For EM-DD, non-weighted MD-EMDD,MD-EMDD and LR-MD-EMDD are ran with 20 randomly chosen starting points. In order to combine the hypotheses resulted from the multiple runs, the average of the hypotheses is computed.

For SVM, the dataset is labeled to scenes and 2751 instances are selected based on the same dataset as the previous 4 methods. Each feature's accuracy is separately tested and the top 5 are selected. Then, the 5 features' weights are set according to their accuracies'. Finally, the 5 features are fused with their weights. The SVM

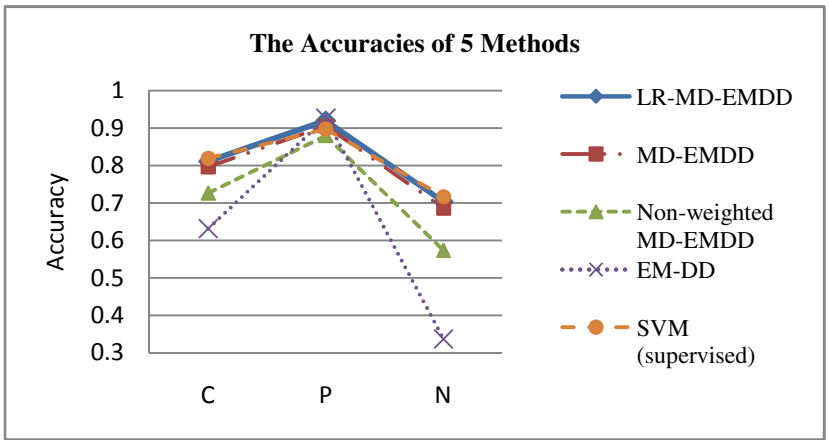

Fig. 4. Comparison of accuracies of horror scene detection achieved by 5 methods. C: Accuracy of correctly detected bags(MIL)/scenes(SVM); P: Accuracy of positive bags(MIL)/scenes(SVM); N: Accuracy of negative bags(MIL)/scenes(SVM).

Table 2. Detail information of accuracies achieved by 5 methods

Algorithm	N_c	$N_{P\to N}$	$N_{N\to P}$	Accuracy
LR-MD-EMDD	487	24	89	0.812
MD-EMDD	478	28	94	0.797
Non-weighted MD-EMDD	436	36	128	0.727
EM-DD	379	22	199	0.632
SVM (supervised)	2256	132	363	0.820

N_c : Number of correctly detected bags(MIL)/scenes(SVM); $N_{P\to N}$:Number of positive bags(MIL)/scenes(SVM) detected as negative; $N_{N\to P}$: Number of negative bags(MIL)/scenes(SVM) detected as positive.

method is implemented based on libsvm[1]. We use RBF as the kernel function, and select the best parameters c and γfromcross validation.

4.3 Results

Figure 4 and Table 2 summarize the comparison of the accuracies achieved by the 5 methods. First, non-weighted MD-EMDD outperforms the traditional EM-DD by 9.5%. Second, MD-EMDD outperforms the non-weighted MD-EMDD with 7%, which indicates the importance to exploit the MGIL problem. Third, the LR-MD-EMDD's performance is 1.5% better than MD-EMDD, which indicates the significance of combining score information with MD-EMDD. The accuracy of LR-MD-EMDD is found to be close to the SVM with a margin of 0.8%.

[1] Chih-Chung Chang, Chih-Jen Lin. LIBSVM: a library for supportvector machine, http://www.csie.ntu.edu.tw/~cjlin/libsvm/

Furthermore, the 4 MIL methods make little difference on the number of positive bags detected as negative, but differ much on the number of negative bags detected as positive. It is intuitive that in EM-DD, if any feature were predicted as positive, the video will be labeled positive. Therefore, it only achieves 63.2%. Instead, the 3 proposed methods give restriction to each feature and take all features into consideration and perform much better on the detection of negative scenes. What's more, the more accurate the importance of features is mined, the better the performance is.

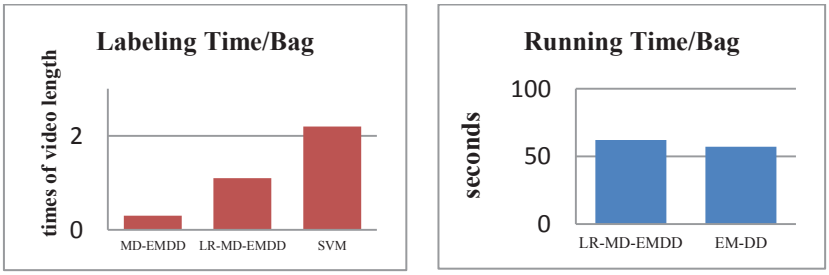

Fig. 5. Comparison of labeling and running time of horror scene detection

Figure 5 compares the training data labeling time of MD-EMDD, LR-MD-EMDD and SVM. Thirty videos for a total of 150 minutes are tested. The SVM requires 334 minutes while MD-EMDD requires a total of 29 minutes and saves 91% labeling time, LR-MD-EMDD requires 115 minutes and save 50% labeling time. To explain the results: the MD-EMDD needs to find whether a video contains a positive scene. Therefore, the users do not need to sort through the whole video. For LR-MD-EMDD, the users only need to go through the video. However, for the SVM, the users need to first watch the video and then delete the negative parts. Please note that the time does not include the video segmentation process.

Also, as EM-DD performances well on its running time, the EM-DD and LR-MD-EMDD are compared, where LR-MD-EMDD costs 62 seconds per bag and is only 5seconds more than EM-DD. The reason is that EM-DD costs mostly in LR-MD-EMDD and different features can be trained and tested simultaneously.

Table 3. μ of five adopted descriptors obtained by MD-EMDD

Descriptor	μ
Color Structure	0.234
Audio Spectrum Spread	0.212
Audio Fundamental Frequency	0.150
Audio Signature	0.103
Motion Level	0.301

Table 4. μ of five adopted descriptors obtained by LR-MD-EMDD

Descriptor	μ
Color Structure	0.251
Audio Spectrum Spread	0.211
Audio Fundamental Frequency	0.131
Audio Signature	0.235
Motion Level	0.172

Finally, Table 4.and Table 5.give the μ of five adopted descriptors obtained by MD-EMDD and LR-MD-EMDD. As is shown, the visual features account for about 43% of the μ, which indicates the significance of the introduction of visual features.

5 Conclusions

In this paper, horror scene detection problem is investigated with incomplete information on training data labels. Specifically, a formulation of the problem as a Multiple Grouped Instance Learning problem is presented. A discriminative learning method is proposed to solve the MGIL problems. Also, a method to effectively make use of "scores" of bags is demonstrated, which offers a new method for ambiguous concept detection. The newly devised method is demonstrated to be more effective and superior compared to the traditional MIL algorithms and close to performance achieved by supervised method.

As for the future work, we plan to: 1) Discuss the possibility of the co-effect of different features. 2) Video is special for its multiple layer structure. It may be valuable to introduce the multiple layer structure and combine it with our method. 3) High level feature can be extracted to achieve higher accuracy. 4) Optimize our method by combining LR-MD-EMDD with SVM to solve the MGIL problems.

Acknowledgements

Project supported by The National Natural Science Foundation of China (No.60802057, No.61071153), Shanghai Rising-Star Program (10QA1403700) and Shanghai College Students Innovation Project (IAP3027).

References

1. Gillespie, W.J., Nguyen, D.T.: Video Classification Using aTree-Based RBF Network. In: IEEE International Conference onImage Processing, vol. 3, pp. 465–468 (2005)
2. Hana, R.O.A., Freitas, C.O.A., Oliveira, L.S., Bortolozzi, F.: Crime scene classification. In: Proceedings of the ACM Symposium on Applied Computing, pp. 419–423 (2008)
3. Moncriefi, S., Venkatesh, S.: Horror Film Genre Typing And Scene Labeling Via Audio Analysis. In: ICME 2003, pp. 193–197 (2003)
4. Jiang, X., Sun, T., Chen, B.: Automatic Video Pattern RecognitionBased on Combination of MPEG-7 Descriptors andSecond-Prediction Strategy. In: Second International Symposium on Electronic Commerce and Security, pp. 199–202 (2009)
5. Gu, Z., Mei, T., Hua, X.-S.: Multi-Layer Multi-Instance Learning for Video Concept Detection. IEEE Transactions on Multimdeia 10(8), 1605–1616 (2008)
6. Dietterich, T.G., Lathrop, R.H., Lozano-Perez, T.: Solving The Multiple Instance Problem With Axis-parallel Rectangles. Artificial Intelligence 89(1-2), 31–71 (1997)

7. Rahmani, R., Goldman, S.A., Zhang, H., Cholleti, S.R., Fritts, J.E.: Localized Content Based Image Retrieval. IEEE Transactions On Pattern Analysis And Machine Intelligence, Special Issue, 1–10 (November 2008)
8. Zhang, Q., Goldman, S.A.: EM-DD: An Improved Multiple-instance Learning Technique. In: NIPS, vol. 14, pp. 1073–1080 (2002)
9. Maron, O., Lozano-Pérez, T.: A Framework for Multiple-instance Learning. In: Advances in Neural Information Processing Systems, vol. 10, pp. 570–576. MIT Press, Cambridge (1998)
10. Top Rated "Horror" Titles (2010), http://www.imdb.com/chart/horror