Multi-layer Objectionable Video Classification System Using Local-Global Information

Seungwan Han, Chiyoon Jeong, and Taekyong Nam
Electronics and Telecommunications Research Institute
161 Gaejong-dong, Yuseong-gu, Daejeon, 305-350
Republic of Korea

Abstract: In this paper, we propose multi-layer objectionable video classification system using local information and global information simultaneously. We also analyze the additional information of video files through the Internet for use in the objectionable video classification from a statistical point of view. The proposed system consists of 3 analyzers and uses MPEG-7 visual descriptors as features for content-based analyzer. To gain the local information of a video, we extract 200 representative frames from video file by uniform sampling and classify every frame using MPEG-7 visual descriptors and a SVM. To avoid the misclassification by a few frames, results of frame classification is used as the global information. Experiment results show that the proposed system has an excellent performance in classifying a video as the objectionable or as the unobjectionable.

Key-Words: Objectionable video, video classification, support vector machine, content-based analyzer

1 Introduction

As the Internet had grown up, a variety of new applications appeared in cyberspace: E-Commerce, Web Marketing, On-line Banking, Messenger Service, VoD(Video on Demand), etc. It enables us to overcome the spatial-temporal limitation. Using it, we can easily gain the information everywhere and everywhen and also watch Internet broadcasting at our own convenience according to individual tastes. In addition, the Internet accelerated the spread of information dramatically. In spite of many advantages of the Internet, the abuse of the Internet produces serious social problems such as the misuse of information, libel, and the dissemination of the objectionable information. Among other things, the mindless dissemination of the sexually objectionable information through the Internet has a bad influence on the children. For this reason, we must cope with this problem with interest urgently.

There have been some researches on preventing the dissemination of the objectionable information through the Internet. Most of them are based on matching the objectionable keyword or blocking the black list of URL and only a few are based on analyzing the text or image content of the information. Recently, as the speed of the Internet is gradually fast, the data type interchanged through the Internet is shifting from the text or image to the multimedia data including videos. When considering such a situation, it is very important to take effective measures for preventing the dissemination of the objectionable multimedia data including videos in the diverse aspects. In this paper, we propose and implement as a technical measure the objectionable video classification system which can classify and block the objectionable videos efficiently.

Until now, the studies of the video classification have mainly concentrated on video genre categorization but there are few studies of classifying videos according to the degree of the sexual objectionableness[1,2,3,4,5,6]. Recently the authors proposed the system that classifies a video as the objectionable or as the unobjectionable[7]. The proposed system consisted of two steps and used machine learning algorithm for classifying videos. In the first step, the system filtered the objectionable video by using the text information which is extracted from the header. In the second step, the system classified a video that passes through the first step by using a SVM(Support Vector Machine) classifier that has 13 features of a video including the skin color information as the input. Although the system had a pretty high performance, there were some shortcomings in that it did not use the local information of a video as the features and depended excessively on the skin color information. The skin color information is a very important clue to decide the sexually objectionable video but the excessive

dependency on the skin color information can cause the misclassification.

In this paper, we propose a novel system that classifies a video as the objectionable or as the unobjectionable using the local information extracted from each frame and the global information derived from a set of frames. The proposed system is named MOVICS(Multi-layer Objectionable Video Classification System). MOVICS gets rid of the dependency on the skin color information by using MPEG-7 color descriptors to get the local information from every frame and avoids the misclassification by a few frames as using a set of frames that are extracted from a video to decide if it is objectionable or not.

The remainder of this paper is organized as follows. In section 2, the design concept of MOVICS is described. Section 3 describes the two-tiered content-based analyzer that is at the core of MOVICS. In section 4, experimental results are given and in section 5, we conclude this work.

2 Design Concept

MOVICS has 3 components as depicted in Figure 1: SA(Statistics-based Analyzer), CA(Content-based Analyzer), and UA(Unified Analyzer).

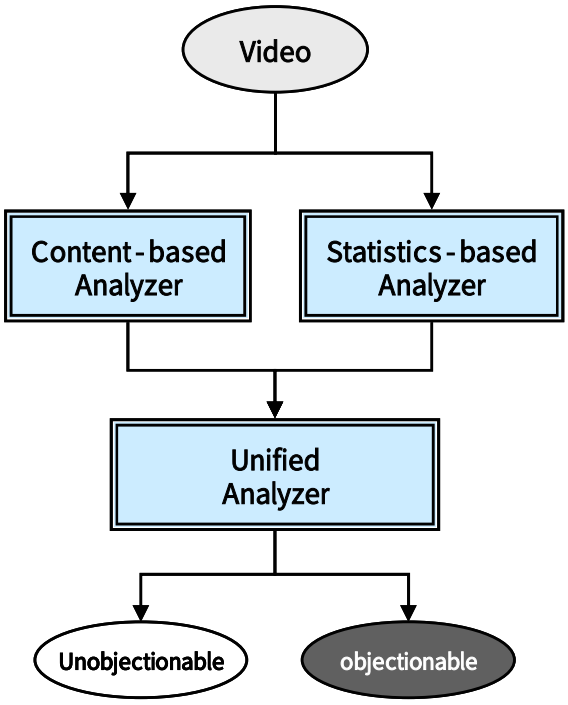

Fig. 1 Multi-layer Objectionable Video Classification System

Using the additional information of a video, SA analyzes the objectionableness. CA scores a video according to the objectionableness of the content. Finally, UA determines whether a video is objectionable or not with the results of SA and CA.

2.1 Statistics-based Analyzer

Using statistics about the additional information of videos that are disseminated through the Internet, SA analyzes the objectionableness of a video.

Generally speaking, the information in the Internet and on the computer has the filename that is used for distinguishing one from the other. The filename of information reflects the content of information in many cases, if not always. In particular, the filename of a video includes a keyword representative of the content generally. This can be used for deciding the objectionableness of a video. However, to evade a filtering system that uses a keyword matching technique, a modified filename tends to be used. Also a keyword matching technique may cause a false-positive problem.

In addition to filename, file size, creator, resolution, aspect ratio, video container, audio codec, etc are used as the additional information for classifying a video as the objectionable or as the unobjectionable. The followings show analysis results of the currently disseminating videos from a statistical point of view.

- **Filename**: a video including a sexual keyword in the filename, if not all, is objectionable mainly.
- **File size**: the file size of an objectionable video is mainly small.
- **Creator**: an objectionable video includes generally the creator information to advertise the porno site whereas movie or drama in the Internet does not.
- **Resolution**: an objectionable video is the low resolution video with a strong probability.
- **Aspect ratio**: an objectionable video has an aspect ratio of 1.33:1 commonly whereas movie or drama in the Internet has an aspect ratio of 1.78:1 chiefly.
- **Video container**: while most of movies and dramas in the Internet uses AVI container including DivX or XviD video codec, the adult broadcasting by Internet real-time streaming service uses ASF or WMV container.
- **Audio codec**: most of movies in the Internet are compressed by AC3(Dolby Digital) or

DTS(Digital Theater Systems) codec and most of objectionable videos are compressed by MP3(MPEG-1 Layer3) or WMV codec.

The above-mentioned additional information can easily be changed by a creator or a distributor to evade a filtering system. As a result, the above-stated information must be used as extra means.

2.2 Content-based Analyzer

To analyze the content of a video is the most accurate method to determine whether a video is objectionable or not. CA analyzes simultaneously the local information, i.e., a frame and the global information, i.e., a set of frames. As illustrated in Figure 2, CA consists of two layers.

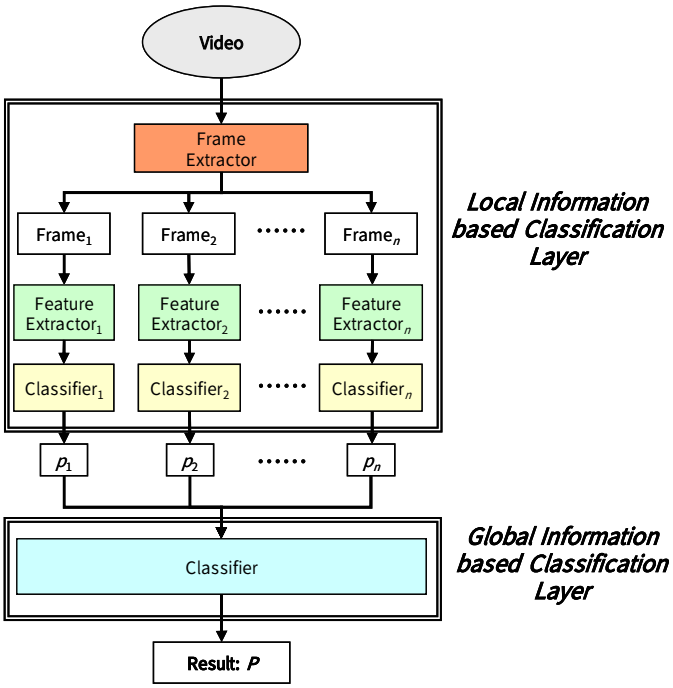

Fig. 2 Content-based Analyzer

LIC-Layer(Local Information-based Classification Layer) extracts frames from a video and computes a objectionable probability in every frame. Because a frame is the smallest unit of a video, results from analyzing a frame individually do not represent the objectionableness of a video entirely. To solve such a problem, GIC-Layer(Global Information-based Classification Layer) computes a objectionable probability of a video using results from LIC-Layer as input. CA will be described further in section 3.

2.3 Unified Analyzer

Combining the result of SA and the result of CA, UA determines whether a video is objectionable or not. According to the policy, UA can give the weight to the result of SA and the result of CA differently. In this paper, to focus on the content analysis, we give priority to the result of CA and use the result of SA as extra means.

3 Two-tiered Content-based Analyzer

3.1 LIC-Layer

LIC-Layer is composed of two steps. One step extracts frames from a video. Other step computes an objectionable probability using a prior-learned classifier.

3.1.1 Frame Extractor

In general, according to a genre and a theme, a video has a distinctive plot and story. Until now, video segmentation and shot boundary detection are mainly used for extracting meaningfully significant frames from a video[5,6]. These methods have to analyze all frames in video to get the shot and extract meaningful frames. As a result, these methods require the amount of time that is directly proportional to the running time of a video.

These methods are very inefficient in the operational aspect because we need representative frames but not meaningfully significant frames to determine whether a video is objectionable or not. The method which extracts frames by uniform sampling or random sampling is more efficient that the above two methods. In this paper, we extract 200 representative frames from a video by uniform sampling.

3.1.2 Feature Extractor

There are several methods that decide whether a frame is objectionable or not. In this paper, we use the objectionable image classification technique because a frame is a still image. To efficiently classify an image, good feature extraction from an image is required.

There are two feature extraction methods: skin color-based feature extraction and MPEG-7 visual descriptor-based feature extraction. In this paper, we use two MPEG-7 visual descriptors, CLS(Color Layout Descriptor) and CSD(Color Structure

Descriptor) that showed the most high performance in the objection image classification.

Feature vector $f(F_t)$ that is used in this paper is defined in (1). $f(F_t)$ consists of 268 feature values: 12 feature values by CLD and 256 feature values by CSD.

$$f(F_t) = CLD(F_t) + CSD(F_t) = (x_1, x_2, \cdots, x_{268}) \quad (1)$$

3.1.3 Classifier

There are many techniques that are used for classifying the information automatically. In this paper, we use a SVM that is extremely efficient and robust among machine learning based classification method. A SVM is based on the Structural Risk Minimization principle from computational learning theory[8,9]. A SVM maps the input data into a predetermined very high dimensional space via a kernel function and finds the hyper-plane that maximizes the margin between the two classes. Because a SVM is a kind of supervised learning method, SVM must be learned with the prior-known knowledge before being used for the classification.

For the prior learning, we extracted frames from various videos, gave every frame a mark(objectionable or unobjectionable), and made SVM learned with marked frames.

This paper uses SVM with probability output as a result of frame classification for the probability concept[10]. The result of SVM is used as the input of GIC-Layer.

Expression (2) presents mathematically an objectionable probability, a result of SVM for a frame *F*.

$$p(F_t) = \Pr obSVM(f(F_t)),\ 0 \le p(F_t) \le 1 \quad (2)$$

3.2 GIC-Layer

GIC-Layer determines the objectionableness of a video entirely but not that of a frame. GIC-Layer uses outputs of LIC-Layer, which are the results of classifying 200 representative frames, as the global information of a video. GIC-Layer can use various methods for classification. In this paper, we use SVM with probability output as in LIC-Layer. SVM of GIC-Layer must also be learned like that of LIC-Layer. We learned SVM used in GIC-Layer with sample videos that the result of classification is previously known.

Expression (3) presents mathematically an objectionable probability, a result of SVM for a video *V*.

$$P(V) = \Pr obSVM(p_1(F_1), p_2(F_2), \cdots, p_{200}(F_{200})),\ 0 \le P(V) \le 1 \quad (3)$$

$P(V)$, which is the result of GIC-Layer, is passed to UA in order to finally determine whether a video is objectionable or not.

4 Experimental results

We gathered various video files distributing through Peer to Peer (P2P) networks. Unobjectionable video files contain the various categories such as movie, animation, TV show, documentary, and sports. Every video file has its own resolution, frame rate, and length.

We extracted 30,000 unobjectionable frames and 30,000 objectionable frames from videos and gained features used in a SVM by tow MPEG-7 visual descriptors, CSD and CLD, from frames and then made a SVM classifier of LIC-Layer learned by using these features. As shown in Table 1, the accuracy of a SVM classifier of LIC-Layer is about 89.7%.

Table 1. Classification result of LIC-Layer SVM

	Classified as objectionable	Classified as unobjectionable
Objectionable Frames(30,000)	26,756	3,234
Unobjectionable Frames(30,000)	2,937	27,063

Table 2. Dataset for objectionable video classification

	Training samples	Test samples
Objectionable video(177)	142	35
Unobjectionable video(177)	142	35

We gathered 177 objectionable video files and 177 unobjectionable video files. For a SVM used in GIC-Layer, we divided video files into two groups: training group and test group. Training group amounts to 80% and test group amounts to 20% as shown in Table 2.

Classification result of objectionable video is shown in Table 3. The result shows that the proposed system performs the classification of objectionable videos with accuracy of about 97.14%.

Table 3. Classification result of MOVICS

	Classified as objectionable	Classified as unobjectionable
Objectionable videos (35)	34	1
Unobjectionable videos (35)	1	34

5 Conclusion

In this paper, we proposed and implemented MOVICS, a multi-layer objectionable video classification system that classifies a video as the objectionable or as the unobjectionable by analyzing the additional information and the content of a video. Also MOVICS used simultaneously the local information extracted from each frame and the global information derived from a set of frames to analyze the content of a video. Experiment results showed that the proposed system had a high accuracy and a low false-positive rate in classifying a video as the objectionable or the unobjectionable.

We expect that this system will help to protect children from the objectionable videos that are recklessly disseminated through the Internet.

References:

[1] M.J. Roach, J.S.D. Mason, and M. Pawlewski, Video genre classification using dynamics. *Proc. IEEE ICCASP*, Salt Lake City, Utah., May, 2001

[2] B-T. Truong, S. Venkatesh, and C. Dorai, Automatic Genre Identification for Content-Based Video Categorization. *Proc. International Conference on Pattern Recognition*, 2000

[3] Y. Chen and E. K. Wong, A Knowledge-Based Approach to Video Content Classification. *Proceedings of SPIE* Vol. 4315: Storage and Retrieval for Media Databases, 2001

[4] S. Fisher, R. Lienhart, and W. Effelsberg, Automatic Recognition of Film Genres. *In the 3rd ACM Int. Multimedia Conference and Exhibition*, 1995

[5] I. Koprinska and S. Carrato, Temporal video segmentation: A survey. *Signal Processing: Image Communication*, 16, 2000

[6] Y. Yuan, Q.B. Song, and J.Y. Shen, Automatic video classification using decision tree method. *Proc. 1st International Conf. on Machine Learning and Cybernetics*, Beijing, 2002

[7] C. Y. Jeong, S. W. Han, and T. Y. Nam, Automatic Objectionable Video Classification System. *Proc. 9th International Conference on Internet and Multimedia Systems, and Applications*, Hawaii, 2005, to appear

[8] VN. Vapnik, The Nature of Statistical Learning Theory, Springer, 1995

[9] T. Joachims, *Making large-Scale SVM Learning Practical. Advances in Kernel Methods - Support Vector Learning* , MIT-Press, 1999

[10] T.-F. Wu, C.-J. Lin, and R. C. Weng, Probability Estimates for Multi-class Classification by Pairwise Coupling, *Journal of Machine Learning Research*, 2004. A short version appears in NIPS 2003