Objectionable Video Filtering Using Hierarchical SVM Classifier

[1]Po-Wei Lin [1]Yi-Leh Wu [2]Cheng-Yuan Tang
[1]Department of Computer Science and Information Engineering ,
National Taiwan University of Science and Technology, Taiwan
[2]Department of Information Management, Huafan University, Taiwan
ywu@csie.ntust.edu.tw

ABSTRACT

As the P2P software prevails on the internet, people contact with the objectionable information* more often than before. Because the objectionable information is not suitable for the minors, how to block or filter the objectionable information has became a critical issue. One of the major objectionable information is pornographic videos. Many studies have been researched on filtering objectionable images, but few studies have been investigated on filtering objectionable videos.

In this paper, we propose a high accuracy objectionable video classifying system. We extract frames from videos to classify objectionable videos with a two-tier SVM classifier. In the first tier, we adopt the traditional image classifier to classify video frames. In the second tier, we propose methods to analyze the classification results from the image classifier in the first tier and generate features to classify videos with a second tier SVM classifier. We show that even if the image classifier in the first tier is far from perfect the proposed two-tier classifier can still produce satisfactory result in classifying videos. Finally, our experiment results suggest that the proposed methods are promising and applicable in real world situations.

Keywords:objectionable video classification, content-based video analysis, web information filtering, support vector machine

1. INTRODUCTION

As the P2P software prevails on the internet, people can obtain anything on the internet. Naturally, some objectionable information would be found easily. Because the objectionable information is not suitable for the minors, how to block or filter the objectionable information has became a critical issue.

One of the major objectionable information is pornographic videos. Although many studies have been done on filtering objectionable images, and many studies have been done on automatic content-based video classification [9] [10] [11], few studies have been investigated on objectionable video classification [3] [4] [8]. The purpose of this paper is to classify objectionable videos for filtering objectionable information on the internet.

Leveraging on the good research results in filtering objectionable images, we employ the OpenCV library to extract the video frames from videos for analysis. Some researchers also adopt the same way to analyze a video, but the critical problem is how to analyze these frames. The method of frame extraction also affects the classification results. In [3], Lee et al. extract a frame every 60 seconds and Wang et al. [4] extract frames in motion analysis. Because one may extract not enough frames or too many frames, we decide to extract one per 100 frames to get a tractable number to reduce the analysis time while maintaining high accuracy in objectionable video classification. Almost all previous works [3][4][8] focused on the image filtering part by analyzing the content of a image. But we focus on analyzing new features from the image classifier results (the first tier) for the video classifier (the second tier). And the two-tier framework will enhance the classification of the proposed filtering system. The major contributions of this work are as follows:

- We propose efficient methods to generate features from the image filter.
- We propose a two-tier objectionable video classifier with high accuracy using the Support Vector Machines (SVMs)[2].

2. THE CLASSIFICATION PROCESS

Figure 1 illustrates the proposed classification process with 4 stages as follows:

1. Preprocess: This stage will extract frames from a video.
2. First tier image classification: This stage employs the SVM image classifier to calculate the classification scores of the input frames and then send the scores to next stage.
3. Second tier (feature generator): This stage will analyze all frames' scores, and use the proposed methods to calculate the feature values for the video SVM classifier.
4. Output: In this stage we apply the video SVM classifier, to obtain the overall results on deciding whether the input video is an objectionable video or not.

We employ the image filter from [1] as our base image classifier to classify an input image into benign or objectionable. The positive classification result value indicates that the input image is objectionable, and the larger positive value suggests that the input image is more likeable to be objectionable. On the contrary, the negative value means a picture is benign, and the smaller negative value represents a picture is more benign. The range of the positive value is [0, 2], and the range of the negative value is [-2, 0]. If the input image is considered not an objectionable picture with very high confidence, the classification value will be -10.

We manually label each image frame from all videos as objectionable or benign and then we compare the image filter classification result. We find that the overall average accuracy of the image filter is 79.02% with 63.05% accuracy in classifying

* Throughout this paper, unless otherwise noted, the objectionable information refers to pornographic information.

objectionable images and 94.98% in classifying benign images. This result shows that the image filter in filtering benign samples is excellent and better than in filtering objectionable samples.

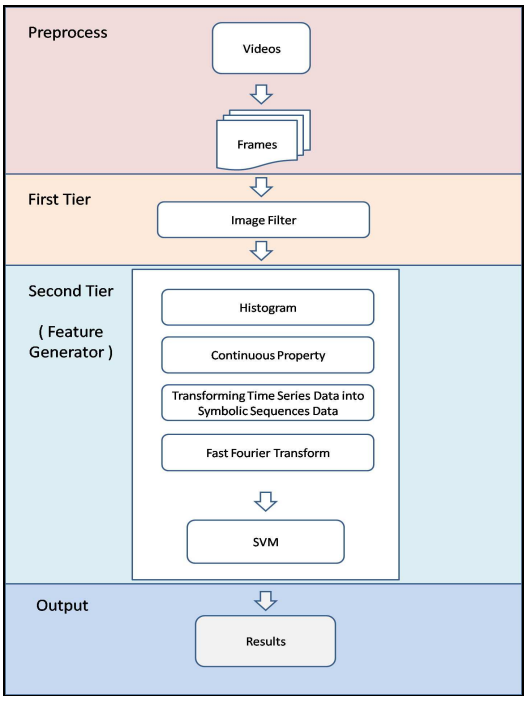

Figure 1. The classification process.

3. SECOND TIER – VIDEO FEATURE GENERATION

We now discuss our feature generation method and explain every feature in details. Because our video data can be consider as time series data [6] [7] [12], we plot the original data value into time series data to observe whether there are features in the data. Most time series patterns can be described in terms of two basic classes of components: trend and seasonality. But in our observation, our time series data does not have obvious features in these two classes. And because finding patterns in time series data is not efficient, we only use the original time series data figure to inspire our ideas.

3.1 Histogram

The simplest feature generation method is to calculate each sample's percentage of frames in different score range. We partition the score range into 7 different ranges: [0, 0.5], [0.5, 1], [1, 2], [0, -1], [-1, -2], [0, 2], [0,-2], and -10. Because the scores distribute in the positive value range [0, 1] is not even, we cut the range [0, 1] into two smaller ranges [0, 0.5] and [0.5, 1]. The other ranges distribute evenly, so we partition them into [1, 2], [0, -1], [-1, -2]. The range [0, 2] represents all the positive value range, and the range [0,-2] represents all the negative value range. The frame counts in score = -10 is the most important feature to identify benign video image, so we retain it to be a standalone feature. In Figure 2 and 3, we take two videos as an example. Figure 2 is an example of objectionable videos, and Figure 3 is an example of the benign videos in time series data with lines.

3.2 Continuous Property

This feature is according to some continuous properties in videos. The actions in a video are almost continuous. If the input video is objectionable, there will be many adjacent frames are all objectionable. For example, a series frames score F = {1, 0.2, -0.3, 0.2, 0.4, 1.1}. F has 5 adjacent pairs: {1, 0.2}, {0.2, -0.3}, {-0.3, 0.2}, {0.2, 0.4}, {0.4, 1.1}. Because there are 3 pairs whose score are all positive value, we can suspect F is an objectionable video. In the same way, if F has over 50% pairs whose score are all negative value, we can suspect F is a benign video. The detail verification will discuss in section 6.2.2.

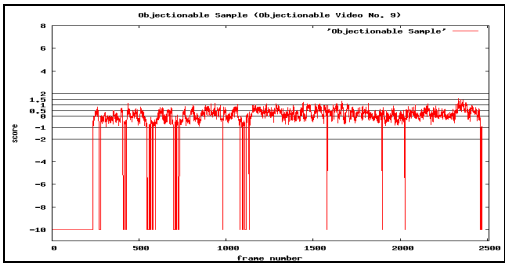

Figure 2. Time series data figure (objectionable sample).

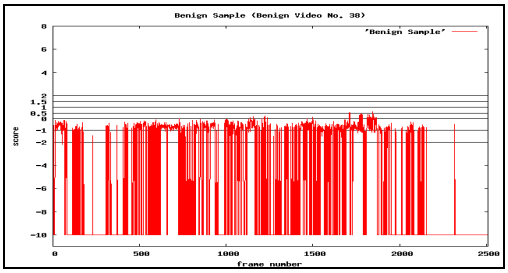

Figure 3. Time series data figure (benign sample).

3.3 Transforming Time Series Data into Symbolic Sequence Data

In [6], to make time series data easy to understand, Ou-Yang et al. propose a new approach to transform time series data into symbolic sequence data. First, given a time series data $\{X_i\}$, $i = 1, 2, 3, ..., n,$ its mean value is called P_0. Then, compare each X_i with P_0, if $X_i > P_0$, let $Y_i = 1$, otherwise $Y_i = 0$. The sequence $\{Y_i\}$ is called the first order symbolic sequence. Next, calculate the sequence $\{Y_i\}$ mean value P_1, then compare each Y_i with P_1, if $Y_i > P_1$, let $Z_i = 1$, otherwise $Z_i = 0$. The sequence $\{Z_i\}$ is called the second order symbolic sequence. In the same way, we can obtain other higher order symbolic sequences. The advantage of this method is that we can transform original sequence into higher order symbolic sequence with only a little information loss; i.e., we can use the new symbolic sequence and the mean value to reconstruct the original sequence with a little error. Because the special score = -10, which may make the general mean value much smaller, we adjust the score -10 to -2 to alleviate this problem.

3.4 Fast Fourier Transform

The fast Fourier transform (FFT) denotes a rapid and efficient algorithm to compute the discrete Fourier transform (DFT). DFT is

important and widely use in digital signal system nowadays. The DFT can transform a function in the time domain into the frequency domain. According to our data is in the time domain, so we try to transform them into the frequency domain to observe whether there are some different features. In this paper, we use the most common FFT algorithm proposed by Cooley-Turkey [4] to transform our data into frequency sequence data.

4. EXPERIMENTS

When evaluating the proposed objectionable video classifier, we have to answer the following three questions:

1. Are the video features proposed applicable?
2. What combinations of the proposed features are more suitable for video classification?
3. How accurate is the proposed video classification method?

We answer these above questions through the following empirical studies.

We collect 150 videos to be our sample data. Among the 150 samples, half of them (75 videos) are benign, and others are objectionable. Most videos' length are about 1 hour in length while some videos are only 10 ~ 15 minutes in length. For every video, we extract one frame from video per 100 frames. Certainly, extracting more frames may result in higher classification accuracy but may take much longer classification time for longer videos. In our case we decide to only extract one frame per 100 frames from videos.

4.1 Experiment - Feature Verification

In the following experiments, we always use two forms of figure to represent the difference between the benign sample and the objectionable sample. We first cut one sample into frames, and then we test every frame's score to some conditions (e.g., score = -10) to get the PF value of this condition. PF is defined in Equation (7).

$$\text{PF(condition)} = \frac{\text{\# of frames satisfy the condition}}{\text{\# of all frames}} * 100\% \quad (7)$$

We include these percentage values as our video feature, now we use them to verify these condition functions.

In the following discussion, the vertical axis represents the sample count in the benign and the objectionable class, and the horizontal axis represents the PF value range: e.g., in Figure 4, the number of objectionable samples that satisfy the condition (score = -10) with 0% ≦ PF(score = -10) ≦ 10 % is 39. Figure 5 show the same statistic as in Figure 4 but in cumulated sample count distribution: e.g., Figure 5 shows that there are 65 out of 75 benign samples with PF (score = -10) < 30%. In other words, 65 of 75 benign video samples have more than 30% of all the video frames that are classified with score equals to -10 by the first-tier image classifier.

4.1.1 Histogram Verification

Figure 4 and Figure 5 shows the result with condition score = -10. In Figure 4, we discover that there are 39 of 75 (52%) objectionable samples whose PF values are between 0%~10%. In Figure 5, there are only 6 of 75 (8%) objectionable samples whose PF(score = -10) values are more than 50%, but there are 47 of 75 (62%) benign samples whose PF(score = -10) values are more than 50%. From these two figures, we can conclude that PF(score = -10) values in benign samples are always higher than objectionable samples, and we can conclude the condition score = -10 is very powerful to identify benign samples.

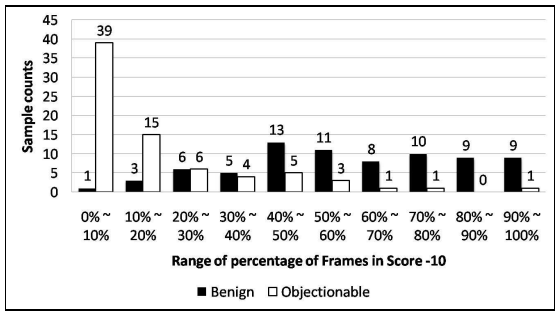

Figure 4. Histogram - the sample count distribution with condition score = -10.

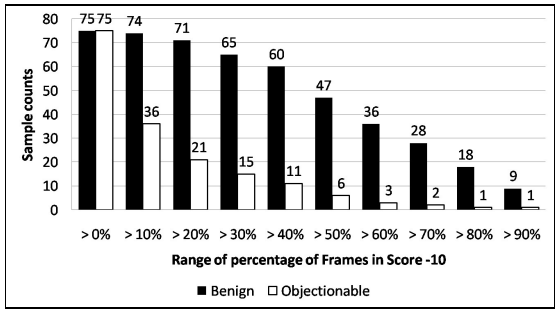

Figure 5. Histogram - the cumulated sample count distribution with condition score = -10.

Figure 6 shows the cumulated sample count distribution with condition score > 0. Figure 6 shows that benign samples' PF values are all less than 30%, but there are 50 out of 75 (67%) objectionable samples' PF values are more than 30%, so this condition of score > 0 can easily differentiate the objectionable and the benign classes. In conclusion, through histogram verifications, we can identify the set of powerful and useful features.

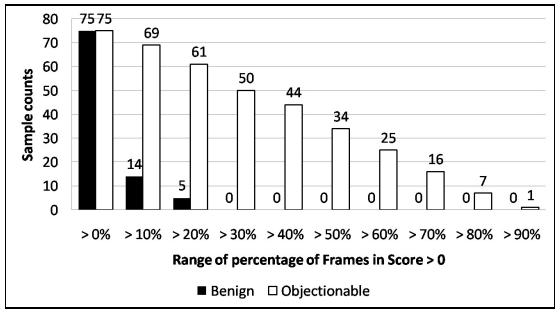

Figure 6. Histogram - the cumulated sample count distribution with condition score > 0.

4.1.2 Continuous Property Verification

Figure 7 shows the statistic result of negative score pairs condition; i.e. PF(the scores of adjacent video frames are both negative). Figure 8 shows the statistic result of PF(the scores of adjacent video frames are both positive). In Figure 7, there are 55 of 75 (73%) benign samples' PF values are between 90% ~ 100%, and in

Figure 8, there are 69 of 75 (92%) benign samples' PF values are between 0% ~ 10%. From the above results we show that the benign samples follow the continuous property as described in Section 5.2. If this video is benign, the number of negative frame pairs will be higher and the number of positive frame pairs will be low. However, we also observe that the objectionable samples do not follow this continuous property; i.e., the PF values of the objectionable samples distribute much more evenly. We conclude that the proposed continuous property can only benefit to classify benign samples more accurately.

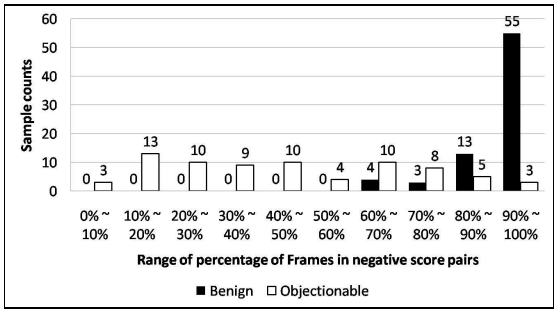

Figure 7. Continuous property - The sample count distribution of negative score pairs.

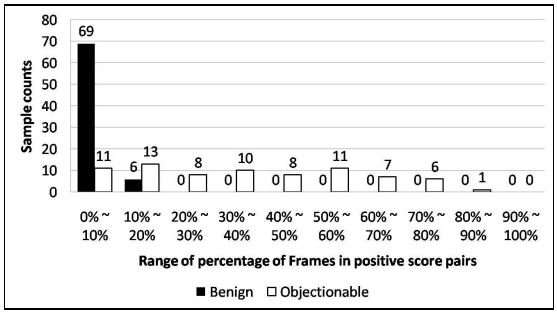

Figure 8. Continuous property – the sample count distribution of the positive score pairs.

4.1.3 Transforming Time Series Data into Symbolic Sequences Data Verification

We employ the transformation in Section 5.3 to transform the original time series data into symbolic sequence data.

Figure 9 shows the cumulated sample count distribution in the first order symbol data form with condition score = 1. Figure 10 shows result of the same video sample but in the second order symbol data with score = 1. A higher PF (score = 1) represents the higher possibility of being an objectionable sample.

In Figure 9, 68 out of 75 objectionable samples' PF(score = 1) values are higher than 50%, but only 28 out of 75 benign samples' PF value are. The result of the first order transform shows the great potential to classify the two classes of images. However, the second order transform does not seem to have the same effect. In Figure 10, we observe that the number of benign samples is higher than the number of objectionable samples in high PF values. This abnormal result is caused by some information loss during the higher order transformation process, so we decide to employ only the first and second order transformation to transform our original time series data.

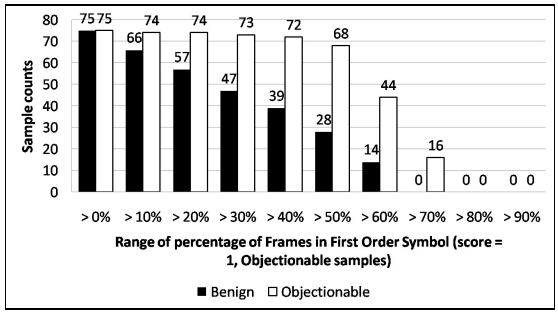

Figure 9. Transforming the time series data into symbolic sequences data - the cumulated sample count distribution in first order symbol with condition score = 1.

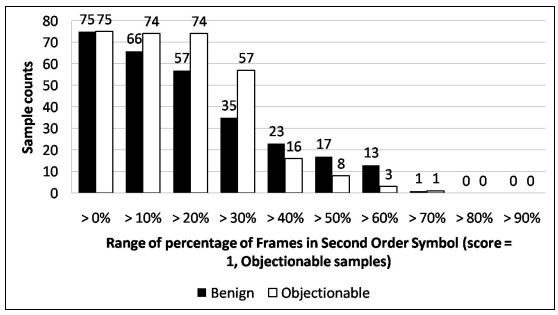

Figure 10. Transforming time series data into symbolic sequences data - the cumulated sample count distribution in second order symbol with condition score = 1.

4.1.4 Fast Fourier Transform Verification

After using the FFT to transform our raw data, we perform statistic analysis on each sample's FFT coefficients (power) that are greater than the pre-defined thresholds.

In Figure 11, both the FFT positive coefficients (power) values of the benign and the objectionable samples descend rapidly with the objectionable samples descending slightly faster. There are 38 out of 75 benign samples' coefficients (power) are over 800, but only 7 of 75 objectionable samples' coefficients (power) are over 800. We can use the difference as features to classify the benign and the objectionable samples.

In Figure 12, we show that both the FFT negative coefficients (power) values of the benign and the objectionable samples descend less rapidly as in Figure 11. But the objectionable samples descend slower than the benign samples as in Figure 11.

From Figure 11 and 12 we observe that if the FFT coefficients (power) of the powers in an input sample are of extreme positive or extreme negative values, the probability of input sample is an objectionable sample will be low. From this conclusion, we can employ the FFT coefficients in out feature set to obtain higher classification accuracy.

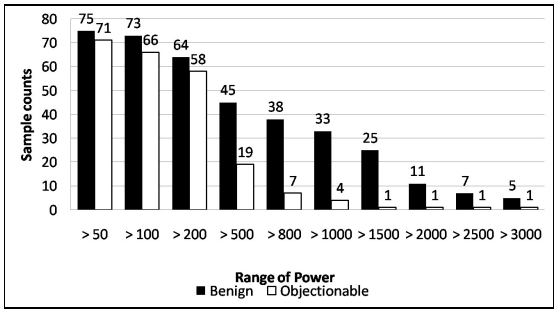

Figure 11. Fast Fourier transform –the sample count distribution in positive power.

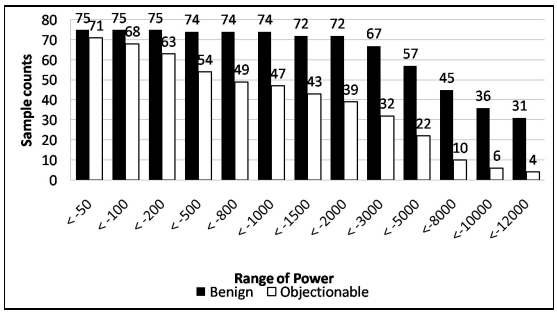

Figure 12. Fast Fourier transform –the sample count distribution in negative power.

For the following experiments, we establish two methods to calculate the FFT features (weighing).

First method – Use ratio

1) Calculate the ratio of objectionable sample to benign sample in every range of positive power; the ratio is to be a weight.
2) Calculate the sum of all ranges' ratio value, and calculate 1/sum to get the 1 ratio unit.
3) Each range's ratio multiplies by 1 ratio unit, and we get each range's score.
4) Repeat above steps in negative power.

Second Method – Use the count difference

1) Calculate the difference in counts between objectionable sample and benign sample in every range of positive power; the difference is to be a weight.
2) The other steps are same to the first method.

After above feature verifications, we answer the question 1 in the beginning of this section by showing the proposed features are applicable and can differentiate between the benign and the objectionable samples, and we can use them to generate features for the proposed classifier.

4.2 Experiment – Accuracy Verification

We now conduct experiments with different combinations of the features discussed earlier. Table 1 shows the 21 combinations where the CP is the abbreviation of the Continuous Property method; the TSISS is the abbreviation of the Transforming Time Series Data into Symbolic Sequences Data method; the FFTR is the abbreviation of the Fast Fourier Transform use ratio; the FFTD is the abbreviation of the Fast Fourier Transform use difference in counts. We adopt the 3-fold cross-validation process for this experiment with 10 rounds; every round has 3 results (3-fold). We then calculate the average accuracy of every round. The results show the accuracy of each round with different combinations.

Table 1. Method combinations.

No.	Method	No.	Method
1	Histogram	**12**	CP & FFTD
2	CP	**13**	TSISS & FFTR
3	TSISS	**14**	TSISS & FFTD
4	FFTR	**15**	Histogram & CP & TSISS
5	FFTD	**16**	Histogram & CP & FFTR
6	Histogram & CP	**17**	Histogram & CP & FFTD
7	Histogram & TSISS	**18**	CP & TSISS & FFTR
8	Histogram & FFTR	**19**	CP & TSISS & FFTD
9	Histogram & FFTD	**20**	Histogram & CP & TSISS & FFTR
10	CP & TSISS	**21**	Histogram & CP & TSISS & FFTD
11	CP & FFTR		

Table 2 summarizes the results from all method combination experiments. We list the rank no. 1, no. 2, no.3 and the worst in each round and the average accuracy of all rounds.

The results of combination no. 20 (Histogram & CP & TSISS & FFTR) in most rounds rank the highest and also the average of all rounds, we conclude that the combination no. 20 is the optimum combination in our system. On average, the combination no. 20 produces accuracy of 93.41%. Figure 13 shows the ROC curve in Round 6 - fold 1 with combination no. 20 that produces the highest accuracy (96.27%) of all experiments.

From above experimental results, we answer the questions 2 and 3 in the beginning of this section. The best combination is combination no. 20 (Histogram & CP & TSISS & FFTR), and the accuracy of our classifier is high (best average accuracy is 93.41%). The experiment results suggest that the proposed two-tier video classifier can classify objectionable videos effectively.

Table 2. The abridged experiment results.

Round No.	Rank			
	1	**2**	**3**	**The worst**
1	No. 20 (93.78 %)	No. 1 (93.76 %)	No. 6 (93.72 %)	No. 3 (82.37 %)
2	No. 8 (94.67 %)	No. 15 (93.79 %)	No. 16 (93.73 %)	No. 3 (81.94 %)

3	No. 9 (93.49 %)	No. 12 (93.26 %)	No. 21 (93.22 %)	No. 3 (78.71 %)
4	No. 6 (94.17 %)	No. 15 (93.94 %)	No. 1 (93.52 %)	No. 5 (80.65 %)
5	No. 20 (93.5 %)	No. 17 (93.38 %)	No. 9 (93.28 %)	No. 3 (84.25 %)
6	No. 20 (94.03 %)	No. 7 (93.85 %)	No. 9 (93.68 %)	No. 3 (81.48 %)
7	No. 20 (93.5 %)	No. 1 (93.1 %)	No. 6 (92.94 %)	No. 4 (80.75 %)
8	No. 6 (93.15 %)	No. 8 (93.12 %)	No. 9 (93.04 %)	No. 3 (81.43 %)
9	No. 8 (94.13 %)	No. 7 (94.02 %)	No. 20 (93.83 %)	No. 3 (81.05 %)
10	No. 20 (94.17 %)	No. 6 (94.01 %)	No. 7 (93.9 %)	No. 3 (83.01 %)
Average	No. 20 (93.41 %)	No. 2 (93.33 %)	No. 7 (93.26 %)	No. 3 (81.92 %)

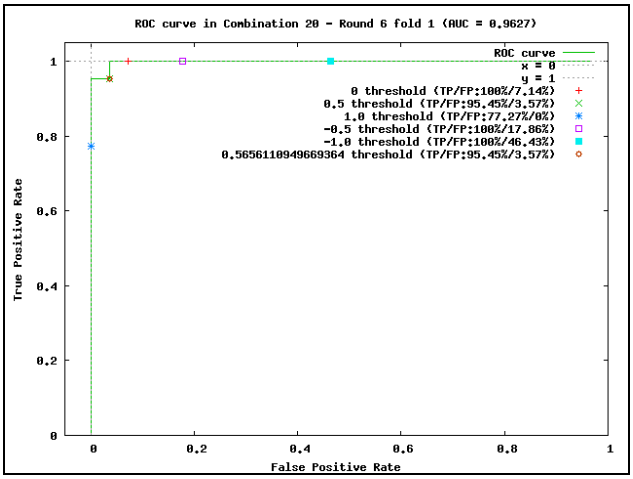

Figure 13. ROC curve in Combination no. 20 – Round 6 fold 1.

5. CONCLUSION AND FUTURE WORK

We propose a hierarchical SVM classifier with two tiers. The first tier is the image filter with the first SVM, and the second tier is the proposed methods analyzing the image filter's results and classifying with the second SVM. From feature verifications, we verify the proposed four sets of features are applicable. Through different method combinations, we find the optimal feature combination for the proposed system to classify the objectionable and the benign videos. Our experiments show that the proposed method has high accuracy on classifying the objectionable and the benign videos with highest accuracy of 96.27%. The experimental results suggest that even with less accurate image classifiers in the first tier; the proposed two-tier classifier can still classify objectionable videos with high accuracy.

Our future work will in two directions. First, we will research on high accuracy and real-time classification. Second, we will explore on using the audio features in objectionable video classification to further improve the effectiveness of the proposed system.

Acknowledments

This work was partially supported by the iCAST project sponsored by the National Science Council, Taiwan, under the Grant No. NSC97-2745-P-001-001 and NSC97-2221-E-011-090.

REFERENCES

[1] Yi-Leh Wu, Edward Y. Chang, Kwang-Ting Cheng, Cheng-Wei Chang, Chen-Cha Hsu, Wei-Cheng Lai, and Ching-Tung Wu. MORF: A Distributed Multimodal Information Filtering System. In *Proceedings of the third IEEE Pacific-Rim Conference on Multimedia (PCM 2002)*, Pages 279-286, Taiwan, December 2002.

[2] Chih-Chung Chang and Chih-Jen Lin. LIBSVM: A library for support vector machines. 2001. Software is available at http://www.csie.ntu.edu.tw/~cjlin/libsvm/

[3] Hogyun Lee, Seungmin Lee, and Taekyong Nam. Implementation of High Performance Objectionable Video Classification System. In *Proceedings of the 8th International Conference on Advanced Communication Technology (ICACT 2006)*, vol. 2, pages 959-962, Korea, February 2006.

[4] Qian Wang, Wei-Ming Hu, Tie-Niu Tan. Detecting Objectionable Videos. *ACTA AUTOMATICA SINICA*, vol.31, no.2, pages 280-286, Beijing, March 2005.

[5] W. T. Cochran, J.W. Cooley, D. L. Favin, H. D. Helms, R.Kaenel,W.W. Lang, G. C. Maling, D. E. Nelson, C. M. Rader, and P. D.Welsh. What is the fast Fourier transform? *IEEE Trans. on Audio Electroacoustics*, vol. 15, no. 2, pages 45-55, June 1967.

[6] Kai Ou-Yang, Wenyan Jia, Pin Zhou, and Xin Meng. A new approach to transforming time series into symbolic sequences. In *Proceedings of the 1st Joint BMES/EMBS Conference*, vol. 2, pages 974, Atlanta, GA, USA, October 1999.

[7] Fu-Lai Chung, Tak-Chung Fu, Vincent Ng, and Robert W. P. Luk. An evolutionary approach to pattern-based time series segmentation. *IEEE Trans. on Evolutionary Computation*, vol.8, no. 5, pages 471-489, October 2004.

[8] Seungwan Han, Chiyoon Jeong, Taekyong Nam. Multi-layer objectionable video classification system using local-global information. In *Proceedings of the 9th WSEAS International Conference on Computers*, Athens, Greece, July 2005.

[9] S. Vakkalanka, C. Krishna Mohan, R. Kumaraswamy, and B. Yegnanarayana. Combining multiple evidence for video classification. *In Proceedings of the International Conference on Intelligent Sensing and Information Processing*, Pages 187-192, January 2005.

[10] N. Watcharapinchai, S. Aramvith, S. Siddhichai, and S. Marukatat. A discriminant approach to sports video classification. In *Proceedings of International Symposium on Communications and Information Technologies (ISCIT '07)*, pages 557-561, October 2007.

[11] Yu-Fei Ma, Hong-Jiang Zhang. Motion pattern based video classification using support vector machines. In *Proceedings of the* 2002 IEEE *International Symposium on* Circuits and Systems (ISCAS 2002), vol. 2, pages II-69 - II-72, May 2002.