PornProbe: an LDA-SVM based Pornography Detection System

Sheng Tang [1,2], Jintao Li[1], Yongdong Zhang[1], Cheng Xie[1], Ming Li[1], Yizhi Liu[1], Xiufeng Hua[1]
[1]Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China, 100190
{ts, jtli, zhyd, xiecheng, mli, liuyizhi, huaxiufeng}@ict.ac.cn

Yan-Tao Zheng[2], Jinhui Tang[2], Tat-Seng Chua[2]
[2]School of Computing, National University of Singapore, Singapore, 117417
{yantaozheng, tangjh, chuats}@comp.nus.edu.sg

ABSTRACT

We present PornProbe, a pornography detection system that detects pornographic contents in videos. To build such a detection system, we leverage a large scale training data set with 65,827 positive training image samples out of a total of 420,615 training samples, and a novel detection scheme based on hierarchical LDA-SVM. The system combines the unsupervised clustering in Latent Dirichlet Allocation (LDA) and supervised learning in Support Vector Machine, so as to achieve both high precision and recall while ensuring efficiency in both training and testing. This demonstration shows how the system detects the pornographic scenes in restricted artistic (RA) movies.

Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval;
H.2.4 [**Database Management**]: Systems—multimedia databases

General Terms: Algorithms, Experimentation, Performance

Keywords: Latent Dirichlet Allocation, SVM, Pornography Detection

1. Introduction

With the advancement of Internet, the proliferation of images and videos has inevitably increased the chance of individuals encountering adult-oriented contents such as pornographic images and videos. The pornographic materials contain subjects that induce sexual excitements. As the pornographic materials raise many social, moral and religious issues, automatic detection of pornographic contents in images and videos is highly demanded to facilitate possible regulation and censorship.

In this demonstration, we present PornProbe, a pornography detection system that can detect pornographic video shots and images. The design of PornProbe aims to accomplish two goals. First, we want to achieve high precision and recall in detection. This is because as compared to the total amount of images and videos on the Internet, the number of pornographic images and videos is relatively small. Second, the system must be highly efficient, in order to scale up to handle huge amount of videos and images on the Internet. In this research, we tackle the above goals from the aspects of database and machine learning. Since the pornographic images vary considerably, it is necessary to establish a large-scale image database to include all possible variations (such as significant variations in races, lighting conditions, textures, viewing angles, and complicated body structures, etc.) of such images as much as possible. Here, we set up a large-scale database with more than 10^5 images. To the best of our knowledge, this is the largest pornographic image dataset, as most existing systems usually use less than 10^3 pornographic images for training. However, training support vector machines on such a large data sets is very time-consuming and it is often a bottleneck. Therefore, it is important to develop fast algorithms for training SVM to learn the pornography detection rules both efficiently and effectively from large-scale training databases.

2. Large-scale training data set

Table 1. Statistics of our training data set

	Images	Video Keyframes	Total
Positive	21,699	44,128	65,827
Negative	51,680	303,108	354,788

Table 1 presents the statistics of our large-scale training image database (including key frames extracted from videos) for pornography detection. Altogether, we collected 420,615 training image samples from a wide variety of sources. For videos, we collected 1,108 pornographic videos from offline VCD sources; 20,000 short pornographic video clips from online media streams by the skin-based detection method [1] from Dec 2007 to Dec 2008; and about 65,000 non-pornographic videos from YouTube, Tudou, YouKu and other websites. For images, we utilized the non-pornographic images from Corel database while downloaded pornographic ones from Pinkworld.

3. LDA-SVM Model

Our proposed solution is motivated by the insight from psychophysical studies that humans can perform coarse categorization of visual objects quite easily and quickly, followed by successively finer but slower discrimination [2]. Specifically, as shown in Figure 1, we propose a hierarchical LDA-SVM model which can scale up to large data sets through a combination of unsupervised clustering and supervised learning. First, we use the generative Latent Dirichlet Allocation (LDA) [3] to model the relationship between images to mine the hidden structure of images. We then perform coarse categorization by clustering large-scale training data set into small topic sets according to the

This work was supported by National Nature Science Foundation of China (60873165), National Basic Research Program of China (973 Program, 2007CB311105) and Co-building Program of Beijing Municipal Education.

MM'09, October 19–24, 2009, Beijing, China.
ACM 978-1-60558-608-3/09/10.

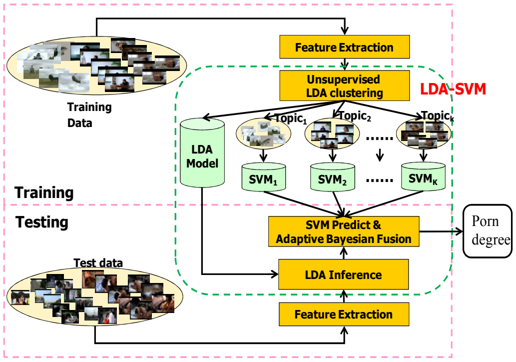

Fig.1 The framework of PornProbe

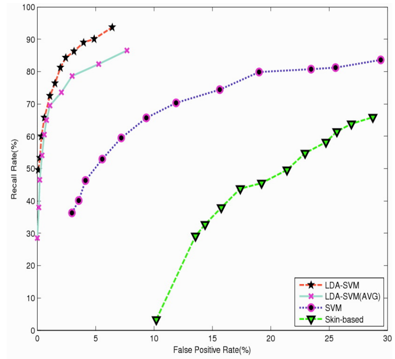

Fig.2 ROC Curves

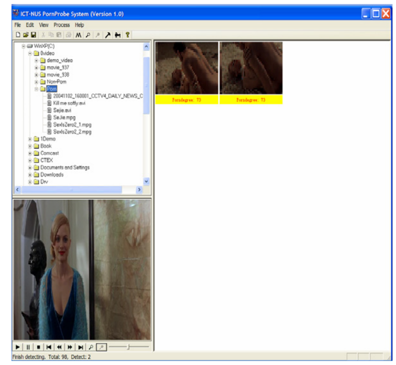

Fig.3 The interface of PornProbe

principle of maximum membership probability determined by the topic-simplex representation vector inferred by the LDA model. Second, we perform successive finer discrimination by training each clustered topic set to generate multiple smaller SVM models. These topic-based SVM models are generally more effective since their optimal separating hyperplanes may be much simpler to discriminate the data and have better generalization performance in the small homogenous topic sets rather than in the large overall data set. Finally, we propose an adaptive Bayesian approach to fuse membership probability with the probability predicted by the corresponding SVM models over all the topic clusters. For a given test sample, we adaptively select only the most probable clusters' SVM models for prediction. Furthermore, as the number of support vectors (SVs) is greatly reduced by training on the smaller topic set, testing efficiency can be considerably improved. This makes it practical for online detection in spite of large training data set.

The separate training of multiple SVM models on each topic set gives rise to the parallelism of SVMs. The distributed training can drastically lower the computational complexity from $O(n^2)$ to $O(n^2/k^2)$. The overall computational complexity for training all the k SVM models is $O(n^2/k)$, where n is the total number of training samples and k is the topic number. The optimal topic number k is not necessary since the first clustering is coarse categorization. We only need to partition the training data set broadly. Therefore, k can be roughly determined by the ratio of the total number of training samples n to the desired average size of topic set.

In our hierarchical framework for combining unsupervised clustering and supervised learning, we use the global color histogram for coarse categorization at the top layer due to its relatively lower dimension and faster extraction [4]; and the prior fusion (concatenation) of the color moment and edge histogram for finer discrimination at the bottom layer to further remove false detection caused by many existing skin-based methods. Although there is no special emphasis on detecting skin, skin detection is actually included in the latent semantic analysis of the color histogram and training of SVM models with color moment. Since we train multiple models on the large-scale training data set that includes nearly all possible variations, it is more stable and robust as compared to skin-based methods and single SVM methods.

For comparison, we evaluate the following 3 systems: (a) skin-based method [1]; (b) single SVM (where we randomly selected 120,000 training samples from our training data set instead of the whole set due to the impractical amount of training computation); and (c) our proposed LDA-SVM (*k=40*) method. Figure 2 shows the ROC Curves of testing the 3 methods on 1695 pornographic and 5415 non-pornographic samples. The Figure clearly indicates that the LDA-SVM is much more effective than other two methods. To test the effectiveness of our adaptive Bayesian fusion method, we compare it against the average fusion method, where its ROC curve is also shown in Fig.2 (the cyan one). The training time, testing time and the numbers of samples and SVs of the single SVM method and LDA-SVM are shown in Table 2, which demonstrates the high training and testing efficiency of the proposed method.

Table 2. Training time and testing time of the SVM methods

	Training samples	Number of SVs	Training time	Testing time for 320×240
SVM	120, 000	24,112	72 hours	667 ms
LDA-SVM	420,615	1842 per topic	6 hours	49 ms

4. Demonstration

In this demonstration, we focus on the effectiveness and efficiency of PornProbe, in pornography detection. Figure 3 shows the interface of PornProbe. The interface composes of 3 panes: a folder browsing pane (left-top) used for selecting the video to be detected; a play-back pane (left-bottom) used for playing back the detected shots or the whole video; and a result browsing pane (right) used for displaying the detected pornographic key frames. After the input video is specified, the pornography detection process will be started to examine the key frames of input video. Once the pornographic key frame is detected, it will be blurred and displayed in the result pane. The pornographic content in a key frame is measured by its porn degree, which is shown at the bottom of each key frame. Our demonstration shows that our system achieves both high precision and recall while ensuring efficiency in testing.

5. References

[1] H. Zheng; "Maximum entropy modeling for skin detection: with an application to Internet filtering"; Ph.D. Thesis, Univeristé des Sciences et Technologies de Lille, France, 2004.

[2] S. Thorpe, D. Fize, and C. Marlot; "Speed of processing in the human visual system"; Nature, 381:520-522, June 1996.

[3] Blei,D.M., Ng, A.Y., and Jordan, M.J; "Latent Dirichlet allocation". Journal of Machine Learning Research, 3, 2003, 993-1022.

[4] S. Tang, J.-T. Li, M. Li, C. Xie, Y.-Z. Liu, K. Tao, S.-X. Xu; "TRECVID 2008 High-Level Feature Extraction By MCG-ICT-CAS"; Proc. TRECVID Workshop, Gaithesburg, USA, Nov 2008.