TIME-CONSTRAINT BOOST FOR TV COMMERCIALS DETECTION

Tie-Yan Liu[1], *Tao Qin*[2] *and Hong-Jiang Zhang*[1]

Microsoft Research Asia, 49 Zhichun Road, Haidian District, Beijing 100080, P. R. China[1]
Dept. Electronic Engineering, Tsinghua University, Beijing, 100084, P. R. China[2]
{t-tyliu, hjzhang}@microsoft.com

ABSTRACT

Commercials detection is very important for TV broadcast analysis. However, independent classification of video shots is very difficult because a considerable portion of individual commercial shots look like program very much. In this paper, the authors proposed a novel way to tackle this problem: to treat successive video shots dependently and improve the final classification performance by considering their temporal coherence. Following this idea, the authors discussed how to apply the majority-based windowing and minority-based merging techniques to the training and test process of statistical classifiers. As a result, a new algorithm named Time-Constraint Boost is proposed. Simulation results show that this algorithm can improve both the training and generalization performance and lead to a promising commercials detection accuracy.

1. INTRODUCTION

Commercials detection plays an important role in various video content analysis applications for TV broadcast. It provides high-level program segmentation so that other algorithms can be applied directly on the true program material. However, it is a great challenge to have robust commercials detection methodology for various content formats and broadcast styles all over the world.

In the literature, many methods were proposed for detecting commercials. In the earlier years, people developed methods [1-3] to detect and recognize known commercials. For example, in [1], the energy envelope of the logged commercial audio was used as signature. With a pattern matching technique it was compared to the energy envelope of the broadcast sound to find known commercials. Comparatively, it is more difficult to detect unknown commercials. Researchers usually used the information of black frames and the change of activity for this purpose. In [4], based on the distribution of black frames and the rate of scene changes, simple heuristic rules were applied to figure out commercials. In [2], black frames and scene change rate were used as fast pre-selector, then edge change ratio and motion vector length were used to determine the final commercials presence. In [5], the authors utilized the black frame positions, frame luminance, letter box and key frame distances as visual features. Then genetic algorithm was applied to optimize the performance by locating the best parameter set.

However, most aforementioned algorithms showed only partially promising results. There are several reasons for this, one of which is that they took use of the black frame information as a dominant feature. Although for the TV broadcasts under their investigation (e.g. USA and Europe), black frames are really used as separators between programs and commercials, this rule does not hold for many other regions such as Asia.

In order to provide a more robust solution, one should use some shot-based general features rather than black frame distribution. However, it was observed that the scenario in today's TV broadcast becomes so complicated: although most programs do not look like typical commercials very much, a considerable portion of commercial shots can not be distinguished from programs in sense of any general visual and audio features [6]. That is to say, if we use independent feature vectors to represent video shots, any classification methods will encounter problems because those program-like commercial shots are very difficult or even impossible to classify.

To tackle this problem, we take advantage over the following two observations. First, the programs and commercials are both locally continuous and last for at least several or even tens of shots (called a block). Second, although not all commercials shots are easy to classify, within each commercial block there really exist several, if not many, non program-like commercial shots. We denote these two observations by "temporal coherence". They provide us a clue to improve the commercials detector by no longer treating video shots independently, but instead taking a sequence of shots into consideration at the same time.

In Section 2 we discuss how to integrate temporal coherence into the training and test process of a statistical classifier. Specifically, a new algorithm named Time-Constraint Boost is proposed for TV commercials detection. Simulation results are listed in Section 3 and some concluding remarks are given in the last section.

0-7803-8554-3/04/$20.00 ©2004 IEEE.

2. TIME-CONSTRAINT BOOST

TV commercials detection is a typical binary classification problem. As we know, most existing statistical classifiers assume the input samples to be independent. However, as mentioned in section 1, the successive video shots are not really independent: they have temporal coherence. It is easy to understand that this coherent information cannot be employed by the traditional classifiers. If we can take use of this, we may have chance to correctly classify those shots which can not be correctly classified before.

In this paper, we clearly propose the viewpoint that the classification accuracy of commercials detection can be improved by taking temporal coherence into consideration. With this idea, many approaches could be used, such as Hidden Markov Model (HMM) and Dynamic Bayesian Network (DBN) to take advantage of the temporal coherence. However, such generative models may feel difficulty when the data itself is complex and with no strict grammar constraints. Instead, in this paper, we focus on how to add temporal coherence to a famous discriminative classifier – AdaBoost [7].

In order to utilize temporal coherence, one choice is to apply majority-based smoothing window to the classification results of an input shot sequence, so as to filter out the strange shots whose temporal neighbors are correctly classified. However, although in some cases this technique helps, it would not work as well as expected when several successive shots are miss-classified. In this case, an alternative way is to use minority-based merging technique. It corrects all the miss-detections between two detected commercial shots which are close enough to each other. However, it will mistake some correctly-classified program shots as well due to false positives (See Fig.1). In fact, both techniques have their conditions of effectiveness. For the former one, a requirement is that the overall accuracy of the classifier should have been high enough. And for the latter one, the condition is that the false positive rate should be very low.

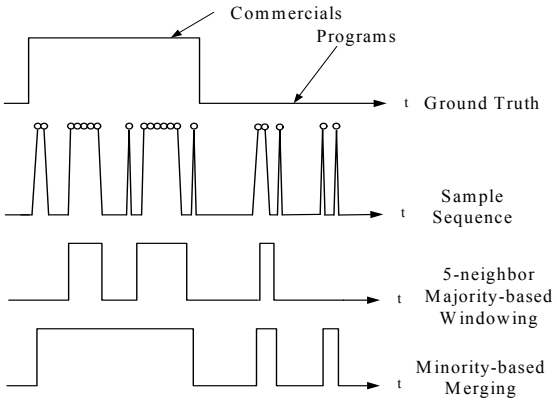

Fig.1 Effects of temporal coherent techniques

As we know, AdaBoost can be decomposed into a chain of weak learners. For each, AdaBoost selects a feature dimension and corresponding threshold that lead to the lowest error rate. However, this "lowest error rate" does not mean a low false-positive rate. In order to make the merging technique applicable, we train each weak learner in a different way as shown in Fig.2. Our second observation listed in Section 1 tells us that even with a low false-positive rate those typical commercials can still guarantee a certain recall. So we set the low false-positive rate (e.g. >90%) as a compulsory constraint. With this, we enumerate all feature dimensions and thresholds to find the one with highest recall rate. After that it becomes okay to adopt the minority-based merging technique. Because the minority-based merging technique can increase the recall rate, as a total result, the final classification accuracy of the composed strong classifier is possibly improved. For the next step, the majority-based windowing technique becomes applicable as well.

The above process is shown in Fig.3. We name this by Time-Constraint Boost (TC-Boost).

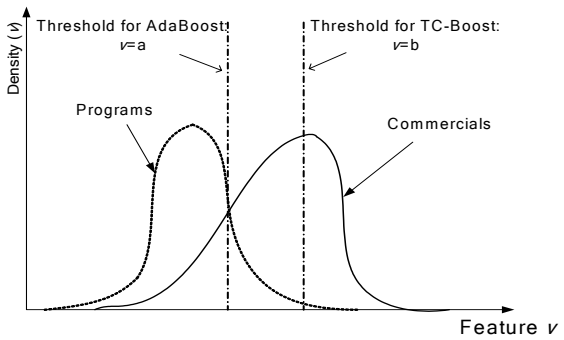

Fig.2 Threshold selection for weak learners in TC-Boost

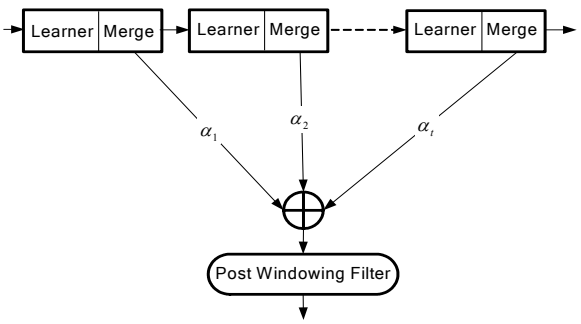

Fig.3 Framework of TC-Boost algorithm

TC-Boost's training process is described as follows.

1) Segment a video sequence into shots and extract a feature vector $\boldsymbol{x}_i$ for each shot.
2) Feed N successive labeled video shots into TC-Boost: $(\boldsymbol{x}_1, y_1)$, …, $(\boldsymbol{x}_N, y_N)$, where $\boldsymbol{x}_i \in \boldsymbol{X}, \boldsymbol{y}_i \in \boldsymbol{Y} = \{-1,+1\}$.
3) Initialize $D_1(i)=1/N$.
4) for $t = 1, \ldots, T$.

a) Train the t-th weak learner using the strategy in Fig.2, and get weak hypothesis g_t: $\boldsymbol{X} \rightarrow \{-1, +1\}$.

b) Apply the minority-based merging technique to the sequence $\{g_t(\boldsymbol{x}_i)\}$ to get a refined weak hypothesis h_t: $\boldsymbol{X} \rightarrow \{-1,+1\}$, with error rate

$$\varepsilon_t = \sum_{h_t(\boldsymbol{x}_i) \neq y_i} D_t(i) \cdot \text{ Let } \alpha_t = \frac{1}{2}\ln\left(\frac{1-\varepsilon_t}{\varepsilon_t}\right).$$

c) Update $D_{t+1}(i) = \frac{D_t(i)}{Z_t} \times \begin{cases} e^{-\alpha_t}, & if \quad h_t(\boldsymbol{x}_i) = y_i \\ e^{\alpha_t}, & if \quad h_t(\boldsymbol{x}_i) \neq y_i \end{cases}$

$$= \frac{D_t(i)\exp(-\alpha_t y_i h_t(\boldsymbol{x}_i))}{Z_t}.$$

where Z_t is a normalization factor.

5) Apply majority-based windowing technique to the sequence $\left\{H(\boldsymbol{x}_i) \mid H(\boldsymbol{x}_i) = sign\left(\sum_{t=1}^{T} \alpha_t h_t(\boldsymbol{x}_i)\right)\right\}$.

6) Filter the sequence once again using the constraint of minimal commercial and program lengths (see the first observation mentioned in section 1). Output the refined results denoted by $\{H^*(\boldsymbol{x}_i)\}$.

7) Write $\{\alpha_t\}$ and $\{g_t(.)\}$ into the training model.

For the test process, we first load the training model and then classify a sequence of video shots just similar to the training process. However, it is noted that the refined weak learners and composed strong classifier in the test process may be different from $h_t(.)$ and $H^*(.)$ due to the change of temporal contexts.

It is easy to understand that by using temporal coherence, the training performance can be improved. Besides this, we have even more benefits for generalization. As we know, for most statistical classifiers, the generalization capability could be reflected by the margin distributions of the training samples. Because AdaBoost focuses on the hard samples, its training process will iteratively decrease the margin in case that hard samples are difficult or even impossible to classify. While because TC-Boost deal with some of such hard samples using temporal coherence, they will not be iteratively emphasized in the new distributions. As a result, the margins will be larger and the generalization capability will be better.

At the end of this section, we should point out that although we discuss this problem based on AdaBoost, similar ideas can be adopted to other discriminative classifiers. For example, we can reduce the false-positive rate of SVM [8] by removing the program-like commercial shots from the training set, or by some other methods. After that the minority-based merging technique can be used, and then the majority-based windowing technique can also be applied. In fact, there is enough space for us to design various algorithms to utilize temporal coherence.

3. SIMULATION RESULTS

In the experiments, we collected 40-hour video sequences from TV broadcasts in both USA and Asia. Half of them were used as training data and the second half as test data. For each video clip, we did shot boundary detection [9] and extracted several basic features to represent the activity of each shot. These features include frame difference, edge change ratio [2], shot frequency, audio energy and so on. After that we used linear combination to wrap them into 156 dimensions.

In the implementation of TC-Boost, we use the following experimental settings. The windowing technique is based on a 5-neighbour window, and the maximal temporal distance between two positives that will be merged together is 6 seconds. The requirements of the final filtering in 6) are that commercial blocks should not be shorter than 30 seconds while the program blocks should be longer than 60 seconds.

Because there are no black frames in the Asia TV streams used, no algorithms referred in Section 1 can work on this dataset. As a result, we only take AdaBoost as a reference algorithm.

Firstly, we compared the training process of TC-Boost with AdaBoost. The recall and precision values in each iteration were listed in Fig.4. We can see from the results that TC-Boost has better training performance than AdaBoost.

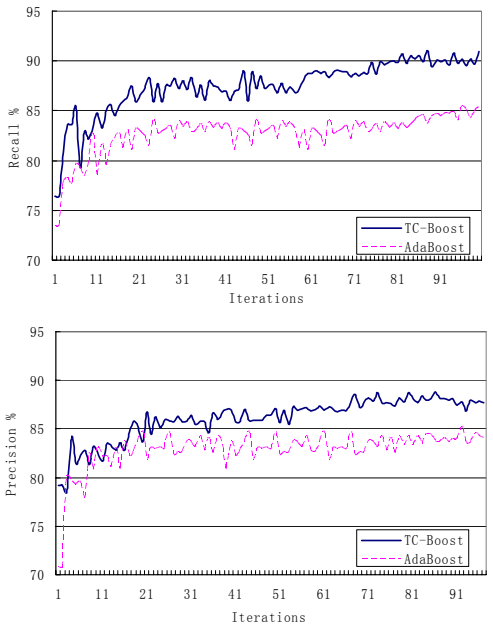

Fig.4 Training performance of TC-Boost and AdaBoost

Secondly, we investigated the generalization capability of TC-Boost. For this purpose, we first calculated the margin distribution graph [10] of both TC-

Boost and AdaBoost. This graph represents the proportion of the training samples whose margins are less than a given value $z \in (-1,+1)$. From Fig.5, we found that TC-Boost has larger margins than AdaBoost, which indicates a better generalization capability.

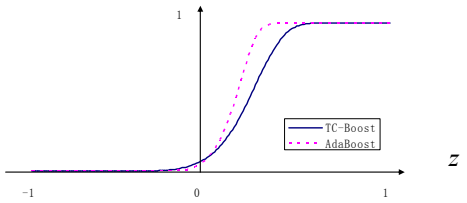

Fig.5 Margin distributions of TC-Boost and AdaBoost (T=100)

Then the commercials detection results on the test set were presented in Fig.6. We find that the results of TC-Boost are better than those of AdaBoost. More importantly, the test performance of TC-Boost is almost as good as its training performance. This is to say, its generalization capability is very good.

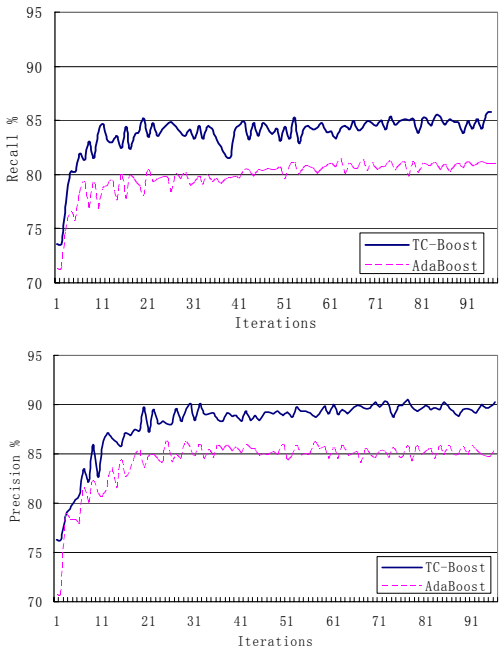

Fig.6 Generalization performance of TC-Boost and AdaBoost

Lastly, we highlighted TC-Boost's detection performance on the test set after 100 iterations in Table.1. From the table we can see that TC-Boost has promising detection accuracy (85~90%, which is about 5% better than AdaBoost). And as just mentioned, no strong but non-robust features such as black frame distribution are used in the experiments, so the result has its general meaning.

Table.1 Commercials detection accuracy on test Set

Algorithm	Recall	Precision
TC-Boost	86.79%	90.25%
AdaBoost	81.00%	85.60%

4. CONCLUSION

Commercials detection plays an important role for TV broadcast analysis. However, it is very difficult to detect commercials because some commercial shots look like program very much. To tackle this problem, the authors discussed in this paper how to employ the dependence of the successive video shots. A new algorithm, named Time-Constraint Boost is proposed, in which by the hybrid adoption of the majority-based windowing and minority-based merging techniques, the training and generalization performances of the statistical classifier are both improved. Simulation results show that the final commercials detection accuracy is promising.

REFERENCES

[1] J. G. Lourens, "Detection and logging advertisements using its sound," *IEEE Trans. Broadcasting*, vol.36, no.3, pp.231-233, 1990.

[2] R. Lienhart, C. Kuhmünch, W. Effelsberg, "On the detection and recognition of Television commercials," *IEEE Intl. Conf. Multimedia Computing and Systems*, pp.509-516, 1997.

[3] J. M. Sánchez, X. Binefa, "AudiCom: a video analysis system for auditing commercial broadcasts," *IEEE Intl. Conf. Multimedia Computing and Systems*, vol.2, pp.272-276, June 1999.

[4] A. G. Hauptmann, and M. J. Witbrock, "Story segmentation and detection of commercials in broadcast news video," *ADL-98 Advances in Digital Libraries*, pp.168-179, 1998.

[5] L. Agnihotri, N. Dimitrova, T. McGee, S. Jeannin, D. Schaffer, J. Nesvadba, "Evolvable visual commercial detector," *IEEE Intl Conf. Computer Vision and Pattern Recognition*, vol.2, pp.79-84, 2003.

[6] C. Colombo, A. D. Bimbo, P. Pala, "Retrieval of commercials by semantic content: the semiotic perspective," *Multimedia Tools and Applications*, vol.13, no.1, pp.93-118, 2001.

[7] Y. Freund and R. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *Journal of Computer and System Sciences*, vol.55, no.1, pp.119-139, Aug. 1997.

[8] V. N. Vapnik, *The nature of statistical theory*. Springer, 1995.

[9] T. Y. Liu, X. D. Zhang, etc. "Constant False-alarm Ratio Processing for Video Cut Detection," *IEEE Intl. Conf. Image Processing*, vol. I, pp.121-124, Sept 2002.

[10] Y. Freund and R. Schapire. "A short introduction to boosting," *Journal of Japanese Society for Artificial Intelligence*, vol.14, no.5, pp.771-780, 1999.