Fuzhou University at TRECVID 2009

Jianjun Huang, Changjin Gao, Qijian Zheng, Yulong Yan, Yongqing Ren

College of Mathematics & Computer Science, Fuzhou University

Fuzhou, Fujian 350108, P.R. China

Abstract

In this paper we describe our approaches to the high-level feature extraction task in TRECVID 2009. Our semantic detection system is based on 6 basic low-level features as well as merged features. We have experimented with several fusion strategies, especially the low-level feature concatenation and Borda fusion. Our experiments showed that the second Borda fusion could increase the number of the true shots returned, but not inevitably improve the mean inferred average precision. Experiments also revealed that the selection of various low-level features and fusion schemes is very crucial for achieving a good system performance. The description of our 6 runs for the high-level feature detection is as follows:

ID	Description
FZU_Run1	AutoCorrelagram and Canny Edge Histogram
FZU_Run2	Color Histogram and Color Coherence Vector
FZU_Run3	Color moment Global and Wavelet texture
FZU_Run4	The average fusion of 6 SVM classification results
FZU_Run5	The weighted Borda fusion of six modals
FZU_Run6	The weighted second Borda fusion result

1. Introduction

Our team took part in the high-level feature extraction task TRECVID 2009, and submitted six runs of type A. Since this is our first time to participate in TRECVID, our major goal is to develop an effective detection system, and to explore various ways to achieve a good result. We present an overview of our participation in the high level feature extraction task, including annotation, key frame extraction, feature representation, learning, fusion and final results.

2. Annotation and Key Frame Extraction

We participated in the collaborative annotation for TRECVID 2009 organized by the LIG, (Laboratoire d'Informatique de Grenoble) and LIF (Laboratoire d'Informatique

Fondamentale de Marseille), and annotated more than 31800 key frames (about 6% of the total labels to annotate). The annotation data on the TRECVID 2007 dataset were used in our systems.
Key frames are extracted to represent the visual content for each shot. On the TRECVID 2007 dataset, the key frames are provided by LIG-LIRIS. The key frames on the TRECVID 2009 test dataset were to be extracted by each participant. To avoid a large set of key frames, we extracted only one key frame per shot according to the shot and sub-shot segmentation provided by Christian Petersohn at the Fraunhofer Institute [3].

3. Feature representation

There are six basic types of global low-level features that we use to build models. They include Color Moments, Color Histogram, Color Coherence Vector, Auto Correlagram, Canny Edge Histogram and Wavelet texture. In our experiments, we found that the simple use of these feature led to poor performance, so we merged them to improve system performance. The mixed features are derived from the concatenation of two basic features, for instance, Auto Correlagram and Canny Edge Histogram, and also from the concatenation of three basic features, such as the combination of Color Moment, Wavelet texture and Color Coherence Vector.

4. Learning and Fusion

We use Support Vector Machine for supervised classification of semantic concepts with RBF Kernel and probabilistic output scores. All our models are trained exclusively on TRECVID’07 data. We have built a set of SVM models using different features mentioned above and model parameters. We use one value of cost factor (1) and two values of gamma parameter (0.2, and 4) of the RBF kernel function. Some models are learned by using all positive examples and twice as many negative examples from standard annotation provided by LIG and LIF Collaborative Annotation.

Beside early fusion, late fusion schemes are also applied to get a final result. In our experiments, we select only a few classifiers (about 6 for each concept) to fuse by using both weighted average scheme and Borda fusion. We select the six best classifiers of the various low-level features to fuse, and different weights are assigned to each classifier. The second Borda fusion is carried out by assigning to each shot a score according to its rank resulted from different weighted sums of the first Borda fusion and then recalculating the score of their weighted sum. This score is then used to re-rank the shots. We observed a slight improvement in the FZU_run6 run over the FZU_run5 run in the number of the true shots returned (1192, up by 100). But the mean inferred average precision of the FZU_run6 decreased a little (0.023, down by

0.005). The experiment showed that the second Borda fusion could increase the number of the true shots returned, but not inevitably enhance mean inferred average precision.

5. Result

Our team has submitted six runs for the High-Level Feature Detection task. Figure 1 and Figure2 show the performances of our typical runs.

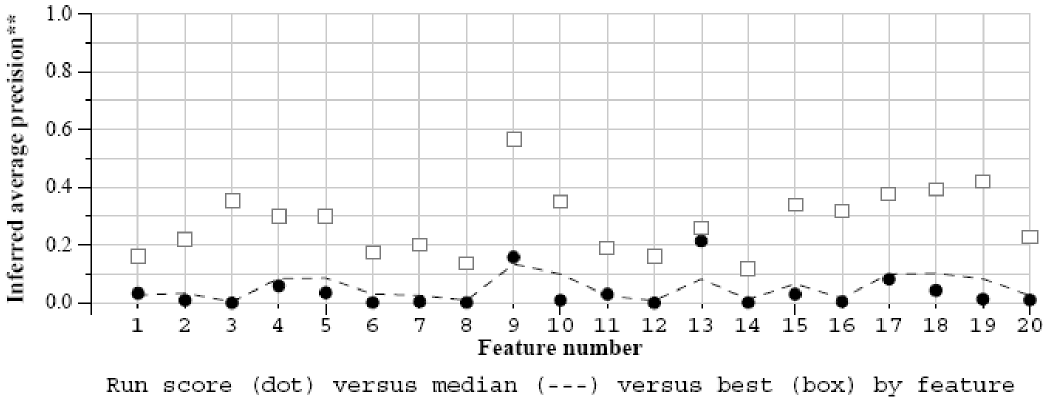

Figure 1: The inferred AP of FZU_Run1

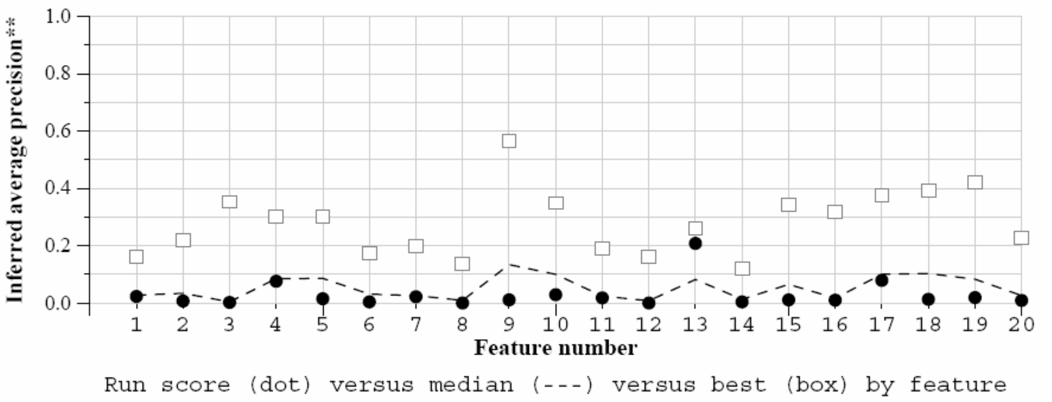

Figure 2: The inferred AP of FZU_Run4

The evaluation results of our submission are better than what we expected. For instance, one of our runs has an Inferred Average Precision of 0.214 for the person-eating concept (13th), which is significantly above TRECVID 2009 HLF detection task median performance for the concept (0.082). But detection results for most of other concepts are not so satisfactory. Due to restriction in time and resources we didn’t train SVM models with more SVM parameters, therefore affected the overall performance of our system.

6. Conclusion and Future Work

In this paper we present our participation in the high-level feature extraction task in TRECVID 2009. We have explored with the low-level feature concatenation and Borda fusion. Our experiments showed that the second Borda fusion could increase the number of the true shots returned, but not inevitably improve the mean inferred average precision. Experiments also revealed that the selection of various low-level features and fusion schemes is vital for achieving a good system performance.

We realize that there are a few areas where improvement can be achieved. Based on our results we believe more models should be trained with different SVM parameters and low-level feature presentation. We also believe that adding to the detection system other low-level and intermediate features, especially local features, is essential to enhance the system performance. Moreover, other fusion strategies need to be explored. The scarcity of positive examples and the data imbalance are all the problems that need to be addressed. In the case of those concepts that have a small number of positive examples in the development data, other resources need to be included. These measures are definitely important for improving the efficiency of our concept detection system.

7. References

[1] Yingyu Liang, et al. THU and ICRC at TRECVID 2008. In Proceedings of TRECVID 2008 workshop.

[2] St´ephane Ayache and Georges Qu´enot, LIG and LIRIS at TRECVID 2008: High Level Feature Extraction and Collaborative Annotation. In Proceedings of TRECVID 2008 workshop

[3] C. Petersohn. Fraunhofer HHI at TRECVID 2004: Shot Boundary Detection System. In Proc. Of TRECVID Workshop, 15-16 Nov. 2004.

[4] Jingen Liu, Yusuf Aytar, Bilal Orhan, Jenny Han, Mubarak Shah, University of Central Florida at TRECVID 2007 Semantic Video Classification and Automatic Search. In Proceedings of TRECVID 2007 workshop

[5] Herv´e Glotin, Zhongqiu Zhao, et al. IRIM at TRECVID 2008: High Level Feature Extraction. In Proceedings of TRECVID 2008 workshop.

[6] D. Wang, X. Liu, L. Luo, J. Li, B. Zhang. Video Diver: Generic Video Indexing with Diverse Features. MIR workshop at ACM Multimedia, 2007.

[7]Shih-Fu Chang , et al. Columbia University/VIREO-CityU/IRIT TRECVID2008 High-Level Feature Extraction and Interactive Video Search. In Proceedings of TRECVID 2008 workshop.

[8] C.C. Chang and C.J. Lin. LIBSVM: A Library for Support Vector Machines, 2001. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm

[9] Apostol Natsev, Wei Jiangy, et al IBM Research TRECVID-2008 Video Retrieval System. In Proceedings of TRECVID 2008 workshop.