BBC rush summarization and High-Level Feature extraction In TRECVID2008

Ken Ishihara
Information Technology Laboratory
Asahi Kasei Corporation
3050 Okata
Atsugi city Kanagawa 243-0021
+81-230-4891

Yuuta Sakai
Information Technology Laboratory
Asahi Kasei Corporation
3050 Okata
Atsugi city Kanagawa 243-0021
+81-230-4891

ABSTRACT

In this paper, first we describe rushes summarization system which is made this year for the TRECVID2008 BBC rushes task. Our aim this year is to build up base system using minimum information and to see how it works. In the second we will describe about our High-level feature extraction system briefly.

1. INTRODUCTION

These days video information is over flooding in the net , hard disk recorders, PC's and portable devices. As video data is hard to overview in limited amount of time, appropriate tags, or labels is needed. Most of the tags used is hand labeled, which is costly and mostly not objectively-defined which is gives limitation in quality to the systems based on .

2. Summarization System Overview

Block diagram of our summarization system is shown on Fig1. An input video is decoded and averaged YCbCr is calculated for each frame as picture parameters. Then movie is cut into scenes based on changes of picture parameters, duplicated scenes are removed based on average color of scenes. Each scene is skimmed to keep motion of video constant. As summarized video is limited to be 2% in time, scene selection and speed adjuster is used to reduce duration if required. Color bar detector, white / black frame detector and Edge detector eliminates unwanted scenes like Color bars pictures out of focus, etc. just before making summarized video.

2.1 Scene cut detection

Our cut detection method for BBC rushes is done as follows. Euclid distance of average color of successive pictures is calculated. Then calculate differential coefficient of it and detect by finding the place where it goes under -threshold to over +threshold. Euclid distance itself worked as well but not in case like fast pans , fast fades etc.

2.2 Removing Redundancy

To get rid of picture redundancy, we used averaged color in scene to compare. All motion action in scene added up to a single YCbCr value. As this value does not have detailed information is not suitable for universal comparison, but just good enough for comparison like rushes where redundancy is limited. i.e. same scene retake repeats immediately after.

Up to second adjacent scene is compared and if the distance between two scenes is below threshold shorter scene is removed.

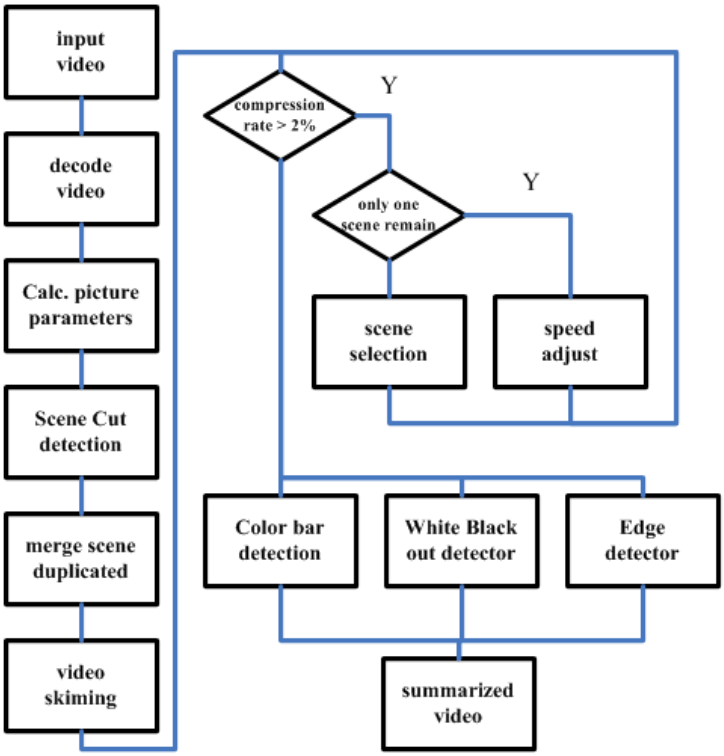

Figure 1: Block diagram of summarization system

2.3 Motion equalized movie

Motion equalize is to keep movements constant. As our system just look at its first order parameter 'average color'. Small motion might be not detected, we use summary of all changes and keep it constant. In this way stopping movie like COLOR BARS and/or slideshows will be a single picture movie which cannot be seen. To avoid this constant frame rate limitation to make it work in any case.

2.4 Reducing Movie

As there is limitation in compression rate, we put secondary reduction of scene. Makes a list of distance between all the scenes and remove scene based on shortest distance we have on the list. Choice of scene is done based on longer-is-better rule.

If there is no more scenes to be removed and still duration is too long, we change threshold value of skimming until it fits.

3. High-level feature extractions

As shown in Fig.2, our High-level feature extraction system is has multiple SVM which is combined together to get the final result. We have additional face detection block.

In the low-level feature calculation part , we calculate , Y histogram, CbCr histogram, CbCr Grid Color Moment, Luv Grid Color Moment, edge direction histogram, linear edge ratio , SIFT parameters, SHIFT pattern histogram

To avoid over fitting problem, we've removed sample data which is overlapping in data space in the training process, to keep cross validation value over 40 %.(in second case 80%)

We provided four feature extraction result output which is CV40% data with/without face detection, CV80% data with/without face detection.

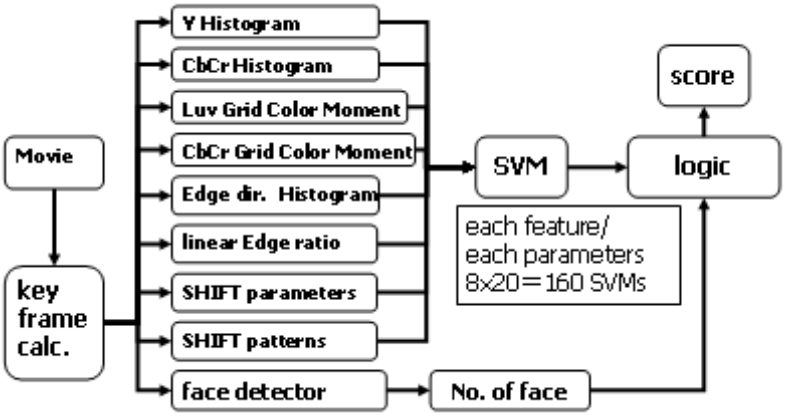

Figure 2: Block diagram of HLF system

4. REFERENCES

[1] Smeaton, A. F., Over, P., and Kraaij, W. 2006. Evaluation campaigns and TRECVid. In Proceedings of the 8th ACM International Workshop on Multimedia Information Retrieval (Santa Barbara, California, USA, October 26 - 27, 2006). MIR '06. ACM Press, New York, NY,321-330. DOI= http://doi.acm.org/10.1145/1178677.1178722

[2] D.-D. Le, S. Satoh. National Institute of Informatics, Japan at TRECVID 2007:BBC Rushes Summarization. NIST, USA, 2007.

[3] K. A. Peker and A. Divakaran. Adaptive Fast Playback-Based Video Skimming Using A Compressed-Domain Visual Complexity Measure. In. Proc. Int. Conf. on Multimedia and Expo, 2004.

[4] K. A. Peker, A. Divakaran and H. Sun. Constant pace skimming and temporal sub-sampling of video using motion activity. Proc. ICIP 2001, Greece.

[5] C.-C. Chang and C.-J. Lin, "LIBSVM: a Library for Support Vector Machines," 2001.

[6] A. Yanagawa, S.-F. Chang, L. Kennedy, and W. Hsu "Columbia University's Baseline detectors for 374 LSCOM Semantic Visual Concepts," Columbia University ADVENT Technical Report # 222-2006-8, March 20, 2007.

[7] A. Yanagawa, W. Hsu, and S.-F. Chang, "Brief Descriptions of Visual Features for Baseline TRECVID Concept Detectors," Columbia University ADVENT Technical Report #219-2006-5, July 2006.

[8] Lowe, D.G. Object recognition from local scale invariant features, Proc. of IEEE International Conference on Computer Vision, pp. 1150.1157(1999).

[9] A. Vailaya, A. Jain, H.J. Zhang. On Image Classification: City vs. Landscape. cbaivl,pp.3, IEEE Workshop on Content - Based Access of Image and Video Libraries, 1998