HuaZhong University of Science and Technology at Trecvid 2006

Kefeng Zhang ,Fan Xu, Xiaojun Zhang, Meng Cai
School of Electronic Science and Technology,
HuaZhong University of Science and Technology
WuHan 430074 , P.R.China
kfzhang@whicc.com ,afanty2000@126.com ,xjzhang518@gmail.com ,01-dream@sohu.com

1、Overview

This is the first TRECVID participation for HuaZhong University of Science and Technology . we participated in the Shot Boundary Detection(SBD).

In the paper, we will present our approaches for SBD . Our main focus was extracting new features based on the luminance to represent the visual content , and combine the traditional method and sliding windows[7] to build our algorithm. In our algorithm we put forward some post processing modules to make system more robust. From the evaluation results our system shows competitive results.

2、Shot boundary detection

Various automatic shot boundary detection algorithms have been proposed ([5],[6],[9]). Upon on the recent TRECVID video test set , we find that short gradual transitions and OTHSs become more . Short gradual transitions , OTHS, abrupt illumination change or object/camera's large movement often lead to significant change of visual content . So in our cut detection algorithm , in order to reduce the disturbance of these things we extract some expressive features , and add post processing modules to process the short gradual transitions and flashlight etc.

Detecting gradual transitions is a Challenge work . Our algorithm is based on the Twin comparison method proposed by [4],but it has some shortcomings ,we give a improved twin comparison algorithm.

2.1 Features to measure visual discontinuity

Various features have been extracted to represent the visual content of each frame[10]. We extract some simple features based on the luminance , trying to reflect visual content from different aspects.

2.1.1 Improved pixel difference

Traditional pixel difference proposed by ([1],[2]) was often effected by noise . To overcoming such shortcomings we give improvements on it . New feature is not sensitive when the whole picture's luminance become abrupt. In our implementation we compute the pixel difference as:

$$C(k) = \frac{1}{XY}\sum_{x=1}^{X}\sum_{Y=1}^{Y} D_{k,k+1}(x,y) \qquad (1)$$

$$D_{k,k+1}(x,y) = |\,(I_k(x,y) - G[k]) - (I_{k+1}(x,y) - G[k+1])\,| \qquad (2)$$

Where G[k] refers to the mean value of luminance of frame K , $I_k(x,y)$ refers to the pixel luminance ,X、Y refer to the dimension of the picture。

2.1.2 Difference degree coefficient

Similarity coefficient is proposed by [12] ,and based on that we put forward the Difference degree coefficient DS as follow:.

$$DS[k]=1-\exp(-10*(1-EX/\sqrt{E(k)*E(k+1)})) \qquad (3)$$

Where EX, E(k) and E(k+1) are defined as:

$$EX=\frac{1}{XY}\sum_{x=1}^{X}\sum_{y=1}^{Y}|I_k(x,y)-G[k]||I_{k+1}(x,y)-G[k+1]| \qquad (4)$$

$$E(k)=\frac{1}{XY}\sum_{x=1}^{X}\sum_{y=1}^{Y}(I_k(x,y)-G[k])^2 \qquad (5)$$

$$E(k+1)=\frac{1}{XY}\sum_{x=1}^{X}\sum_{y=1}^{Y}(I_{k+1}(x,y)-G[k+1])^2 \qquad (6)$$

2.1.3 Monochrome feature

This features is mainly used to find monochrome frame.

$$D(k) = \frac{1}{XY}\sum_{x=1}^{X}\sum_{y=1}^{Y}|I_k(x,y) - G(k)| \qquad (7)$$

2.1.4 Similarity degree

Based on the pixel intensity ,we put forward another features ,Similarity degree. This a variable of statistics ,and this feature is mainly used for short gradual transitions, our experiments proves it is effective for short gradual transitions.

$$SD(K) = \frac{1}{XY}\sum_{x=1}^{X}\sum_{y=1}^{Y}\left[\frac{\min(I_k(x,y), I_{k+1}(x,y))}{\max(I_k(x,y), I_{k+1}(x,y))}\right] \quad (8)$$

2.1.5 YUV histogram

$Hist(k)$ refers to the YUV histogram of frame K, this feature is proposed by [3],and it is robust to noise.

2.2 Cut detection

Various automatic shot boundary detection algorithms have been proposed by ([5],[9]).

Traditional threshold method, in which feature variation between adjacent frames is directly compared with a global threshold *Tc*,and our algorithm is based on the adaptive threshold method , but we find many short gradual transitions or OTHS or flashlight or large movement can lead to feature variation above *Tc.* Our cut detection algorithm divide into two steps: first step is rough detection , second step is doing post processing to avoiding such disturbances.

2.2.1 Rough detection

(1) Many experiments show that the feature $C(k)$ is reach the maximum value when cuts take place . we can use the condition max(C[k-3],C[k-2],C[k-1],C[k],C[k+1],C[k+2],C[k+3])=C[k] to select all potential cut positions .

(2) Besides conditions mentioned above ,we use another sliding window to build adaptive thresholds

C[k]>para1*（C[k-2]+C[k-1]+C[k+1]+C[k+2]）

C[k]>para2 *（C[k-1]+C[k+1]）

Where para1,para2 are variable parameters. Upon on the two conditions above and another luminance conditions, our Recall can reach 99%-100%,and for some test video, precision can reach 65%-70%.Precision is low which is mainly caused by short gradual transitions,OTHS,large movement,flashlight etc.

2.2.2 Post Processing

Besides cut transition, abrupt illumination change like flashlight or large movement of object/camera sometimes or short gradual transition will also lead to feature variation above *Tc*. To reduce these false alarms, we design special processing modules, including flashlight processing and gradual transition filter, to sift the cut candidates.

Flashlight processing

In [11] an ideal flash model and ideal cut model are described,we design a simple module.

Short gradual processing

From experiments , we find that these short gradual are almost 2 frames dissolve .so our main focus is to check out the 2 frames dissolve.

A dissolve sequence is defined as the mixture of two video sequences[6], and the corresponding equation is below:

$$F(x,y,t)=f_1(t)\, S_1(x, y, t) + f_2(t)\, S_2(x, y, t) \quad (9)$$

We put forward a concept that on one hand when $|f_1(t) - f_2(t)|$ is small , call it standard 2-frames dissolve, on the other hand when $|f_1(t) - f_2(t)|$ is very large, we call it non-standard 2-frames dissolve.so we use the feature C(k) to build mathematics model to detect the 2-frames dissolve. Results show that the model is good for standard 2-frames dissolve but not good for non-standard 2-frames dissolve as we expected.

2.3 Gradual transitions detection

Gradual transitions detection is really a challenge work,one reason is there are so many types : dissolve, fade , wipe ,OTHS etc ; the other is a transition process may last from 2 frames to more than 100 frames.our gradual transitions detection divide into several parts : fade detection , short gradual transitions detection, long gradual transitions detection.

2.3.1Fade

The function of this part is detecting the boundary of fade in , fade out and FOI .During experiments ,we find some characteristic of D(k) with the Frame Number increasing.

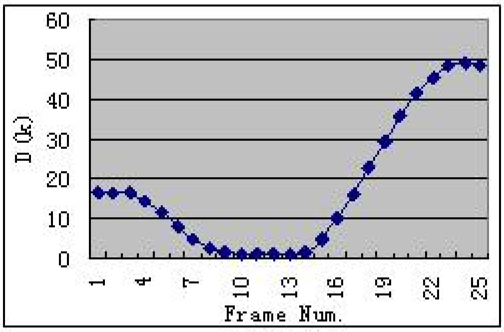

Figure1 ： FOI

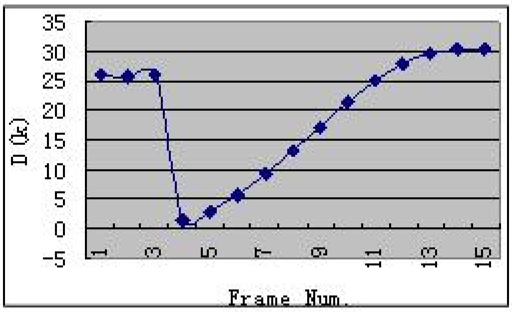

Figure2 ： Fade in

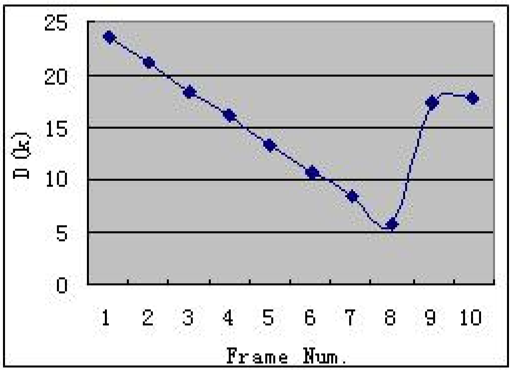

Figure3 ： Fade out

Two steps for Fade detection

(1) use the feature *D(k)* to find out the monochrome frame .

(2) seach boundary

From the figure1,2,3 ,we can see that the curve is monotonic increasing or monotonic decreasing or a parabola [8], so we build a mathematics model to detect the fade transitions.And this method is very effective ,both recall and precision are very high, what's more ,the boundary is very accurate.

2.3.2 Short gradual transitions detection

Considering a transition process may last from 2 frames to more than 100 frames. ,we divide gradual transitions into two categories short gradual transition and long gradual transition. Short gradual transition is divided into two categories : whose process length is 2 frames calls 2-frames short gradual transitions and whose process length is 3-8 frames short gradual transitions calls 3-8 frames ,others are long gradual transitions.

(1) 2-frames short gradual transitions

2-frames short gradual transitions have been processed in cut detection.you can see them in 2.2.2

(2) 3-8 frames short gradual transitions

Algorithm of this part is also based on the twin comparison method , Two threshold value tl and th which is a adaptive value is determined by its forehead value.

tl = para3* (Hist[k-1]+ Hist[k-2])

th= para4* (Hist[k-1]+ Hist[k-2])

and we add a new parameters ds as a supplyment to the traditional two threshold value ,experiments show it is useful.

ds= para5* (DS [k-1]+ DS [k-2])

2.3.3 Long gradual transition detection

About long gradual transitions detection ,our algorithm is also based on the Twin comparison method , and we make some improvements on it . we select the feature Hist[k] and DS[k] . Two threshold value tl and th which is a adaptive value is determined by its forehead value.

tl = para6* (Hist[k-1]+ Hist[k-2]+ Hist[k-3])

th= para7* (Hist[k-1]+ Hist[k-2]+ Hist[k-3])

ds= para8* (DS [k-1]+ DS [k-2]+ DS [k-3])

our improvements :

(1) add a new condition ds, which is a

supplyment to the twin comparison method.

(2) add anti-diturbance processing

During an ideal long gradual transition , all the Hist[k] is larger than tl , but actually there are many cases that Hist[k] is less than tl. so we statistic the number of such cases , when the number reachs a threshould, we abandon it.

3、Results and Discussion

In this section, we discuss results of our system for shot boundary detection when applied to the trecvid 2006 test set.

The shot boundary detection test set consisted of 13 video files with a total duration of approximately 5 hours and 30 minutes. There are 3785 transitions, of which 1844 are annotated as cuts, and 1941 are annotated as gradual transitions.

Figure 4 shows our evaluation results ,and each runs with different threshold.Figure 5 and Figure 6 show the performance of our system for cut detection and gradual transitions , measured in recall and precision , and compared to all other submissions. Figure 7 shows frame recall and frame precision to measure how accurately we detect the start and end of gradual transitions . Mean runtime results are shown in Figure 8.From the evaluation results our system shows a competitive result and our algorithm is efficient .

We examined the detection results and analyzed the causes of misclassification. We found that many reasons cause the performance is not as good as we expected:

(1) The distinction between CUTs and very short gradual transitions is blur, especially the non-standard 2-frames dissolves. Our system is based on mathematics model.,but It is unsuccessfully distinguish CUTs and the non-standard 2-frames dissolves. Therefore, the system annotate some of the non-standard 2-frames dissolves as CUTs. As a result, the performance is not as good as we expected.

(2)In order to decrease the motions' affection, we have improved pixel difference to remove such false alarms. However, the video-in-video scenes and some gradual transitions could not find efficienctly in our system.

Figure 4: Evaluation results of the ten submissions

ALL		CUTS		GRADUAL			
						Frame	
Recall	Prec	Recall	Prec	Recall	Prec	Recall	Prec
0.762	0.788	0.797	0.881	0.669	0.59	0.849	0.757
0.81	0.775	0.843	0.877	0.722	0.567	0.883	0.731
0.787	0.774	0.809	0.883	0.728	0.565	0.88	0.732
0.789	0.771	0.809	0.882	0.734	0.562	0.877	0.733
0.791	0.771	0.812	0.88	0.733	0.562	0.877	0.733
0.835	0.743	0.843	0.878	0.812	0.52	0.89	0.715
0.801	0.76	0.812	0.88	0.773	0.548	0.877	0.725
0.808	0.762	0.811	0.879	0.799	0.56	0.865	0.739
0.796	0.779	0.811	0.879	0.755	0.585	0.853	0.742
0.788	0.785	0.811	0.879	0.726	0.594	0.849	0.748

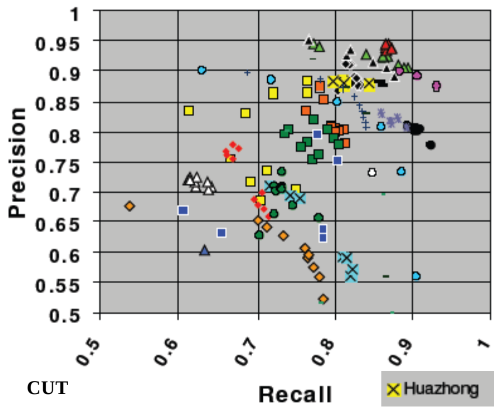

Figure 5: cut results

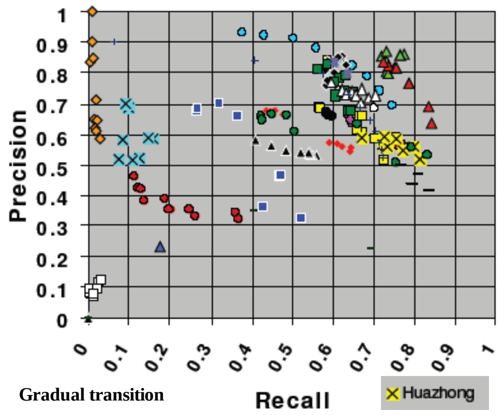

Figure 6: Gradual results

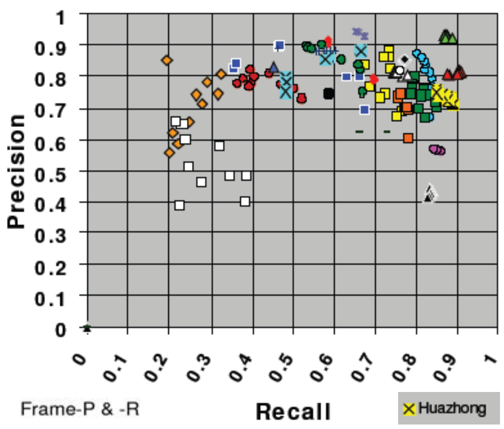

Figure 7: Frame P & R results

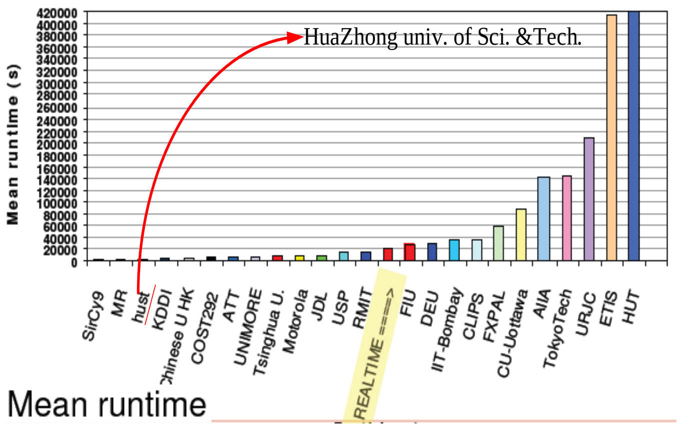

Figure 8: Mean runtime

References

[1] Kikukawa T., Karafuto S. Development of an automatic summary editing system for the audio visual resources. Transactions of the Institute of Electronics , Information and Communication

Engineers , 1992 , J752A(2) : 204 ~ 212

[2] Liu Qianlei, Yang Luxi, Zou Cairong, Twi-difference algorithm and pixel matching twi-difference algorithm[J]. Journal of Image and Graphics, 2003,8(2):161 – 168 (in Chinese)

[3] Nagasaka A., Tanaka Y., Automatic video indexing and full video search for object appearances. In: Knuth E., Wegner L.M. eds. IFIP Proceedings of Visual Database Systems. Amsterdam , The Netherlands : North2Holland ,

1992 , 113 ~ 127

[4] H. J. Zhang, A. Kankanhalli, S. W. Smoliar, Automatic Partitioning of Full-motion Video, ACM Multimedia System, Vol. 1, No.1, pp. 10-28, 1993.

[5] R. Lienhart. Comparison of Automatic Shot Boundary Detection Algorithms. SPIE Storage and Retrieval for Still Image and Video Databases VII 1999, Vol. 3656, pp. 290-301, Jan. 1999.

[6] Lienhart, R. “Reliable Transition Detection In Videos: A Survey and Practitioners Guide,”

International Journal of Image and Graphics (IJIG), 1(3):469–486, 2001
[7] B. T. Truong, C. Dorai and S. Venkatesh. New Enhancements to Cut, Fade, and Dissolve Detection Processes in Video Segmentation. ACM Multimedia 2000, pp. 219-227, Nov. 2000.
[8] J. Meng, Y. Huan, and S.F. Chang. Scene Change Detection in a MPEG-Compressed Video Sequence. IS&T/SPIE Proceedings, Vol. 2419, San Jose, CA, Feb. 1995.
[9] S.Lefevre, J. Holler, and N. Vincent. A review of real-time segmentation of uncompressed video sequences for contentbased search and retrieval. *Real-Time Imaging*, 9:73–98, 2003.
[10] U. Gargi, R. Kasturi, and S. Antani. Performance characterization and comparison of video indexing algorithms. In *Proc. IEEE* Conference on Computer Vision and Pattern Recognition, Santa Barbara, CA, pages 559–565, june 1998.
[11] Zhang Dong, Qi Wei, Zhang Hong Jiang. A new shot boundary detection algorithm[A]. In:Proceedings of 2nd IEEE Pacific-Rim Conference on Multimedia[C], Beijing, China, 2001:Page 63-70
[12] HAN Bing , J I Hong2bing , GAO Xin2bo. Cut-before-detection and hierarchical detection algorithm for video shot segmentation. Systems Engineering and Electronics. Vol1 27 .No12. Feb. 2005