Published as: Che, X., Yang, H., & Meinel, C. (2013, October). Lecture video segmentation by automatically analyzing the synchronized slides. In Proceedings of the 21st ACM international conference on Multimedia (pp. 345-348). ACM.

Table 1: Segments Overview of Public Dataset

Video	Borders	Segs	GLS	PLS	TS	AveLength
d-g-u	Yes	4	3	0	1	13:29
k-e-d	No	7	0	1	6	05:53
k-k-s	No	6	0	1	5	08:41
k-r-n	Yes	5	3	0	2	12:27
k-s-n	Yes	3	2	0	1	15:02
s-d-s	No	4	0	2	2	11:45
s-e-r-c	Yes	4	2	0	2	14:12
s-k-s	No	5	0	0	5	09:26
s-l-n	Yes	5	2	1	2	09:44
s-w-i	Yes	7	2	2	3	05:06

Table 2: Segments Overview of Additional Dataset

Video	Borders	Segs	GLS	PLS	TS	AveLength
5626	No	5	0	1	4	08:05
6011	No	4	0	0	4	07:11
6031	No	5	0	2	3	06:06
6098	Yes	6	4	0	2	07:58
6102	Yes	6	3	1	2	03:19
6104	No	5	0	1	4	07:12
6106	Yes	5	4	0	1	06:22
6196	No	8	0	4	4	07:20
6201	No	3	0	2	1	16:26
6212	Yes	6	2	1	3	10:21

reasons. Firstly, we have to manually achieve ground truth data before comparing, but this may be too subjective to convince people. Secondly, it is not fair to compare results if they derive from completely different resources, especially when the accuracy of OCR is much better than ASR (*Automatic Speech Recognition*), which is well acknowledged.

The overview segmentation result of the public and additional dataset are depicted in Table 1 and Table 2 respectively. For public dataset, we use the initials of the folder as the video name in 'Video' column, while a series number for the video from additional dataset. 'Border' means whether the synchronized slides of this video contain obvious subtopic border, which is obtained manually. Next we present the numbers of total segments, GLS, PLS and TS generated by our solution. And finally an average length of the segments in each lecture video is calculated.

From the segments overview we can easily find out that a presentation gets commonly split into 3∼8 segments, with the average length controlled inside 5∼15 minutes. It means that a presentation will never be cut too fragmentary, for which each segment will keep plenty of information as complete knowledge points, and the duration of a segment is comfortable for the learners behind the screen, neither too long to feel tired, nor too short to be confused.

Based on the test datasets, the proposed segmentation method is well capable in exploring the subtopics border when they exist (*10 in 10*). According to the further analysis showed in Table 3, over 50% segments we got from this kind of videos is GLS, in addition with almost 10% PLS, makes the segmentation result highly logical. For other videos, there are also over 1/4 segments achieved as PLS, from in fact comparatively lineal organized or discrete consisted presentations. In general, nearly half segments derived from our 20-videos dataset are logical.

Table 3: Segments Ratio Analysis

Video Type	GLS	PLS	All LS	TS
with Border	52.9%	9.8%	62.7%	37.3%
no Border	-	26.9%	26.9%	73.1%
All	26.2%	18.4%	44.6%	55.4%

And to be noticed, each presentation automatically contains one TS because the first slide is always a front cover and never a part of any logical structure. And many presentations will have one more TS at the end due to the large probability that the presenter will use the final slide to express the gratitude. Regarding these two facts, the actual ratio of logical segments should be even higher.

4. CONCLUSION

The proposed solution for lecture video segmentation has been proven effective by the evaluation results. Mainly by comparing text, we have successfully explored lots of logical correlation between slides and apply them into segmentation process. To go further in the future, the method of analyzing must involve more artificial intelligence factor, which is what we will attempt next.

5. REFERENCES

[1] V. Badrinarayanan, I. Budvytis, and R. Cipolla. Semi-supervised video segmentation using tree structured graphical models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, page 1, 2013.

[2] G. Friedland and R. Rojas. Anthropocentric video segmentation for lecture webcasts. *Journal on Image and Video Processing*, 2008:9, 2008.

[3] A. S. Imran, L. Rahadianti, F. A. Cheikh, and S. Y. Yayilgan. Semantic tags for lecture videos. In *Semantic Computing (ICSC), 2012 IEEE Sixth International Conference on*, pages 117–120. IEEE, 2012.

[4] I. Koprinska and S. Carrato. Temporal video segmentation: A survey. *Signal processing: Image communication*, 16(5):477–500, 2001.

[5] M. Lin, J. F. Nunamaker Jr, M. Chau, and H. Chen. Segmentation of lecture videos based on text: a method combining multiple linguistic features. In *System Sciences, 2004. Proceedings of the 37th Annual Hawaii International Conference on*, pages 9–pp. IEEE, 2004.

[6] M. Onishi, M. Izumi, and K. Fukunaga. Blackboard segmentation using video image of lecture and its applications. In *Pattern Recognition, 2000. Proceedings. 15th International Conference on*, volume 4, pages 615–618. IEEE, 2000.

[7] N. Yamamoto, J. Ogata, and Y. Ariki. Topic segmentation and retrieval system for lecture videos based on spontaneous speech recognition. In *European Conference on Speech Communication and Technology*, pages 961–964, 2003.

[8] H. Yang, M. Siebert, P. Luhne, H. Sack, and C. Meinel. Lecture video indexing and analysis using video ocr technology. In *Signal-Image Technology and Internet-Based Systems (SITIS), 2011 Seventh International Conference on*, pages 54–61. IEEE, 2011.