SynTag: A Web-based Platform for Labeling Real-time Video

Yen-Chia Hsu **Tay-Sheng Jeng** **Yang-Ting Shen** **Po-Chun Chen**
Interactive Architecture Laboratory, Department of Architecture,
National Cheng Kung University, Tainan, Taiwan
{N76994017, tsjeng, N78941044, PA6984029}@mail.ncku.edu.tw
+886-6-275-7575 ext. 54158

ABSTRACT

Real-time video streaming has been widely used in multimedia learning environments. As production of online videos is increasing exponentially, it is becoming more difficult for users to reach relevant content. In this paper, we propose SynTag, a web-based platform that enables users to label three types of tags—*Good*, *Question*, and *Disagree*—and to make comments synchronously and asynchronously with visualization of time-stamp video previews on an interactive timeline. SynTag generates real-time thumbnails by using real-time tags for presenters to receive instant feedback and for other users to retrieve presentation videos. In a pilot study, we found our users' tagging behaviors significantly different when they were in lecture events or discussion events. We envision that enabling users to apply tags in real-time will help reduce the complexity of classification of videos.

Author Keywords

Timeline Visualization; Collaborative Tagging.

ACM Classification Keywords

D.0 **[Software]**: General; J.4 **[Computer Applications]**: Sociology; H.5.1 **[Multimedia Information Systems]**: Video; H.5.2 **[User Interfaces]**: Graphical user interfaces, Interaction styles, Screen design, User-centered design; H.3.1 **[Content Analysis and Indexing]**: Indexing methods; H.3.3 **[Information Search and Retrieval]**: Information filtering, Relevance feedback, Retrieval models; H.3.5 **[On-line Information Services]**

General Terms

Algorithms; Design; Experimentation; Human Factors.

MOTIVATION

Many learning environments record videos, but they lack efficient tools to record the interactions. The traditional education process tends to give students knowledge but not to encourage them to express their feedbacks, especially those which are implicit. This leads to information loss. Even if the feedbacks can be stored, most environments do not establish the relationships between audience feedbacks and the video. Even if presentations can be recorded properly, the current environments still lack an appropriate user interface for visualizing results. As a result, the retrieval process is complicated and time-consuming. People may need to spend tremendous time searching through videos aimlessly.

This paper focuses on the problem of how to reduce the complexity of indexing video content by labeling video presentations in real time through a web-based platform in today's multimedia learning environments.

Related Work

Some studies have addressed reducing the complexity of video classification by automatically or semi-automatically tagging the content in videos [1, 2]. However, SynTag uses collaborative tags as metadata which are designed to assist users with locating specific instances in a given resource [3]. Social tags yield an effective retrieval process, whereas automatically generated metadata do not [4]. As Marlow et al. [5] indicate, users attempt to express their opinions through the tags.

Similar research on video- or audio-sharing systems has allowed users to make real-time tags manually [6, 7]. The EVA system [8] lets users assign collaborative tags to video key frames to enable efficient retrieval. Unlike these systems, SynTag provides instant feedback to the presenter through a thumbnail. The visualization of tags makes it easier for presenters to track audience feedback instantly. SynTag also provides timeline visualization for archived timestamps and displays the density of tags inside timestamps. Timelines connote a story rather than simply a logical collection [9].

METHOD

The SynTag System

In this paper, we propose a software named SynTag (ialab.tw/legenddolphin/TagDisplay16/TagDisplay16.html), written in Flex and AmCharts (www.amcharts.com). It can record and broadcast videos, alternate between two cameras,

CSCW'12, February 11–15, 2012, Seattle, Washington, USA.

sort events, and visualize tagging results in timestamps. SynTag has three significant functions:

1. Collaborative Synchronous Tagging Scenario

In this scenario, users can react to the presentation by expressing their feedback in real time. SynTag allows three types of tags—*Good*, *Question*, and *Disagree*—allowing users to express positive, neutral, or negative feedback. We use only three simple tag types because we do not want listeners to be distracted while making tags. If listeners want to give more feedback, they can make comments. SynTag can parse comments and checks whether hyperlinks are included (e.g., http://www.google.com) by regular expressions. If so, SynTag immediately adds hyperlink nodes on the "hyperlink" curve. Listeners can click on a node and link to the hyperlink in a new browser window.

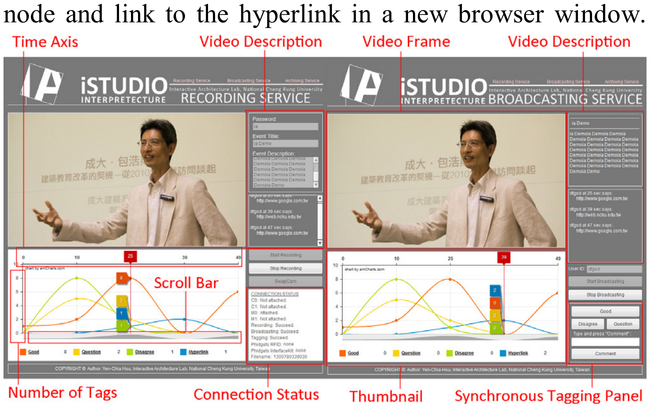

Figure 1. The screenshots of the tagging scenario's interface. Lecturing service (left) records videos and can alternate between two cameras. Listeners use broadcasting service (right) to watch live video and make tags.

2. Chronological Sorting Algorithm and Instant Feedback

SynTag first stores the information from tags in an array. The event title and description are stored in tag[0] and tag[1]. While tagging, SynTag stores the tags' properties, which include video ID, user ID, the time a tag was posted, the type of tag, and the tag content in tag[i+1], tag[i+2], tag[i+3], tag[i+4], and tag[i+5]. The variable i represents the last index of the tag array.

Second, SynTag evaluates the tag counter. If the difference between the current tagging time and the previous one is less than 10 seconds, SynTag adds a 1 to the tag counter for the type of tag node on the thumbnail. Otherwise, SynTag updates the last tagging time to the current one, resets the associated tag counter, adds 1 to the counter, and creates a new tag node on the thumbnail to provide instant feedback to the presenter.

3. Timeline Visualization

Finally, SynTag archives the results of collaborative labeling on a timeline to enable video retrieval (Figure 2, left). The color of timestamps represents the density of tags inside the timestamp. The darker the timestamp, the denser the tags inside. The timeline shows the information in timestamps when a user rolls the mouse cursor over the timeline. When users click on a timestamp, SynTag shows the content (Figure 2, right). Users can retrieve the presentation video by using the real-time tagging thumbnail inside timestamps below the video frame. For example, a user would like to watch the video segment at second 33 because he finds that there are many "Disagree" tags at that point. He can click the time node at second 33 on the thumbnail, and SynTag will adjust the video time bar to play the video segment at that point. Users can quickly watch the video segment without searching the entire video. SynTag also offers an asynchronous tagging scenario.

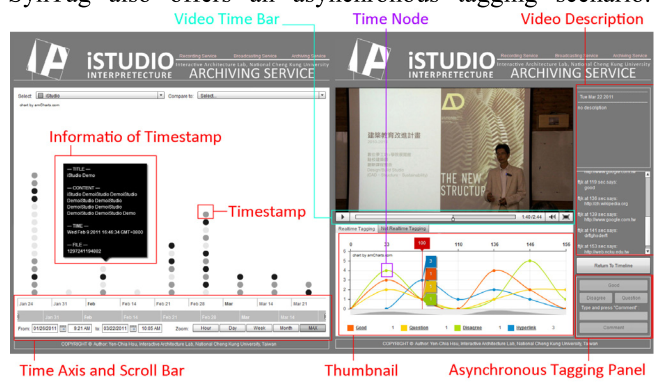

Figure 2. The screenshots of the interface of the archiving service. The timeline (left) shows all events in chronological order. The timestamp (right) shows the representation of the event and enables retrieval.

Participants

Two different types of participants are included in this study: presenters and listeners. Presenters are professors who hold lectures or students, alumni, and scholars who work in the design field. Listeners are presenters, teachers, students, and remote audiences anywhere in the world who are interested in the presentation.

Procedure

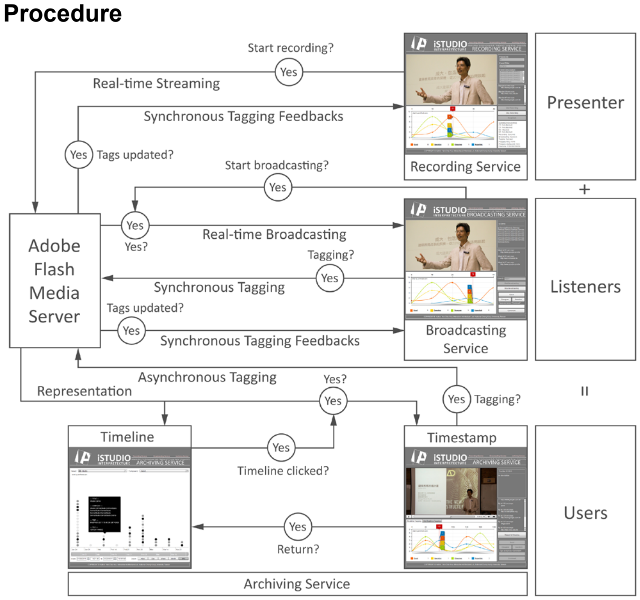

Figure 3. The procedure of SynTag.

Before beginning the recording, we asked 10 listeners willing to use SynTag to join the pilot study. The presenter then set up cameras and links to the lecturing service, entered information, and started recording. SynTag streams videos to the server for broadcasting. Listeners linked to the broadcasting service, started watching the live video, and expressed their feedback through SynTag.

During the recording, SynTag ran the sorting algorithm to add nodes to the tag curve on the thumbnail in real time. The presenter could monitor the thumbnail and respond to listeners' instant feedback, for example, by adjusting the speed of his presentation.

After the presentation, anyone could link to the archiving service and apply tags asynchronously or retrieve particular video segments by clicking time nodes on the thumbnail to adjust the video time bar.

Analysis

We recorded 51 videos from January 26 to March 31, 2011. Of these videos, 4 were lectures, 3 were discussions, and others were demonstrations and tests. In this pilot study, we collected 7 recordings of lectures and discussions. The videos are analyzed by three steps:

1. In Table 1, we classify event types and calculate the scale and total time of the event, the numbers of tags, and the effective peaks as shown in Figure 4 on the thumbnail visualization for video content indexes with the number of tags exceeding 70% of the number of people using SynTag.

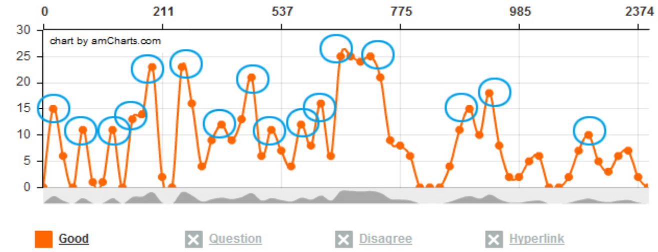

Figure 4. The circles indicate effective peaks.

2. In Table 2, we calculated the percentage of associated number of tags and the effective peaks in different events.

	Type	Description	Date	Scale	Presenter	Listener	Total number of			Tags/ min
							Time(min)	Tags	Effective Peaks	
A	Disc.	lab meeting	Mar 2	S	graduates	graduates	19.90	214	8	10.75
B		lab meeting	Mar 2	S	graduates	graduates	68.03	768	29	11.30
C		TED Tainan	Mar 29	M	students	students	177.77	366	12	2.06
D	Lect.	design lecture	Mar 1	S	teacher	sophomore	24.57	207	4	8.42
E		design lecture	Mar 10	M	alumni	sophomore	30.82	136	4	4.41
F		official speech	Mar 25	L	professor	citizens	20.24	279	10	13.78
G		official speech	Mar 25	L	professor	citizens	40.96	590	16	14.40
Average							54.61	365.71	11.86	6.70

Table 1. The properties of events. Ten listeners used SynTag in each event. *Lect.* and *Disc.* in type column represent lecture and discussion. *S*, *M*, and *L* scales represent approximately 15-20, 30-35, and 60-70 persons, respectively.

	Type	% of associated number of tags				% of associated number of effective peaks			
		Good	Que.+Dis.	Question	Disagree	Good	Que.+Dis.	Question	Disagree
A	Disc.	43.93	52.34	52.34	0.00	50.00	50.00	50.00	0.00
B		46.74	50.00	50.00	0.00	48.28	51.72	51.72	0.00
C		46.45	48.90	19.67	29.23	41.67	58.33	25.00	33.33
D	Lect.	85.50	6.28	6.28	0.00	75.00	25.00	25.00	0.00
E		74.26	5.15	5.15	0.00	100.00	0.00	0.00	0.00
F		71.68	13.26	11.83	1.43	90.00	10.00	10.00	0.00
G		91.53	3.56	1.02	2.54	100.00	0.00	0.00	0.00

Table 2. The analysis of different types of tags. *Que.+Dis.* on the top row represents combined Question and Disagree tags.

3. In Table 3, we calculated the p value (two-tailed) of two event types (lecture and discussion) by using student's t test.

Tag	% of associated number of tags	% of associated number of effective peaks
Good	0.001518531	0.001707755
Que.+Dis.	1.61673E-05	0.001707755

Table 3. The *p* value (two-tailed) of lecture events and discussion events in student's *t* test.

RESULTS

First, we found our users' tagging behaviors during lectures to be significantly different from those during discussions (the p values in student's t test in Table 3 being less than 0.05). As shown in Table 2, listeners applied more *Question* and *Disagree* tags during discussions while they tended to apply *Good* tags more during lectures. The differences between the percentages of *Good* tags and combined *Question* and *Disagree* tags during discussion events are much less than the differences during lecture events. This differentiates the characteristics of discussions and lectures.

Second, the values in the right-most column in Table 1 indicate that listeners using SynTag expressed feedback by making 6.70 tags per minute on average. We found that those people using SynTag were more willing to tag in large lectures and small discussions than in small lectures and medium discussions.

DISCUSSION AND FUTURE WORK

First, some presenters said that it may cause pressure for them to receive instant feedback through SynTag. However, we think it important for presenters to receive instant feedback from listeners and to respond by adjusting the content, the speaking speed, and manner of illustrating ideas, thereby improving the quality of a presentation.

Second, some listeners said that SynTag in some degree causes distraction at the live site because users tag by clicking buttons. The problem with the tagging methods could be improved by applying tangible interactions or pattern recognition, such as scanning spaces with webcams.

Third, SynTag is currently constrained in terms of the type of tags and tagging methods. An interpreter that can parse data from sensor networks and analyze human activities in specific spaces will be needed to augment the ability of SynTag. By applying the interpreter, SynTag would have the potential to assign videos attributes, classifying them in several sets and defining the relationships between them, thus making the retrieval process more efficient.

CONCLUSION

SynTag externalizes listeners' feedback through collaborative tagging (6.70 tags per minute on average), represents the presentation process in an algorithm, provides instant feedback to the presenter, and makes the retrieval process easier through viewing thumbnails and timelines. By applying users' real-time tags, SynTag can identify discussions and lectures (p values < 0.05 in Table 3), thereby reducing the complexity of video classification.

ACKNOWLEDGMENTS

This work was supported by the TOUCH Center and Taiwan National Science Council grant No. NSC100-2221-E-006-224.

REFERENCES

1. Bartolini, I., Patella, M., & Romani, C. (2010). Shiatsu: Semantic-based hierarchical automatic tagging of videos by segmentation using cuts. *Proceedings of the Third International Workshop on Automated Information Extraction in Media Production* (pp. 57–62).
2. Ballan, L., Bertini, M., Bimbo, A. D., Meoni, M., & Serra, G. (2010). Tag suggestion and localization in user-generated videos based on social knowledge categories and subject descriptors. *Proceedings of the Second ACM SIGMM Workshop on Social Media*.
3. Hill, J. R., & Hannafin, M. J. (2001). Teaching and learning in digital environments: The resurgence of resource-based learning. *Educational Technology Research and Development, 49*(3), 37–52.
4. Melenhorst, M., Grootveld, M., Setten, M. van, & Veenstra, M. (2008). Tag-based information retrieval of video content. *Proceedings of the First International Conference on Designing Interactive User Experiences for TV and Video* (pp. 31–40). ACM. Retrieved April 1, 2011, in press. CHI 2002, ACM Press (2002), 1-8.
5. Marlow, C., Naaman, M., Davis, M., & Hall, S. (2006). HT06, tagging paper, taxonomy, Flickr, academic article, to read. *Proceedings of the Seventeenth Conference on Hypertext and Hypermedia* (pp. 31-39).
6. Renduchintala, A., Kelkar, S., John, A., & Seligmann, D. D. (2007). Designing for persistent audio conversations in the enterprise. *Proceedings of the 2007 Conference on Designing for User Experiences*.
7. Topkara, M., Pan, S., Lai, J., Wood, S. P., Boston, J., & Heights, Y. (2010). Tag me while you can: Making online recorded meetings. Science, 25038.
8. Volkmer, T., Smith, J. R., & Natsev, A. P. (2005). A web-based system for collaborative annotation of large image and video collections: An evaluation and user study. *Proceedings of the Thirteenth Annual ACM International Conference on Multimedia* (pp. 892–901).
9. Moraveji, N., Ahmad, S., Kita, C., Chen, F., & Kamvar, S. (2011). Weblines : Enabling the Social Transfer of Web Search Expertise using User-Generated Short-form Timeline. CSCL 2011.