Multimedia Content Summarization and Application to Video Recommendation

Yi-Ting Huang, Chi-Cheng Tsai, Ching-I Chung, Chia-Hsing Shen, and Jie-Chi Yang,
Graduate Institute of Network Learning Technology, National Central University,
{coral, ken, cecilia, arcadia, yang}@cl.ncu.edu.tw
Yu-Chieh Wu, Department of Computer Science and Information Engineering, National Central University,
bcbb@db.csie.ncu.edu.tw

Abstract: In this paper, we present an automatic adaptable multimedia content recommendation system to auto-recommend suitable multimedia learning materials for learners. The proposed system firstly extracts important content as summarization from input raw video data, while the generated summarization will be auto-routed to learners according to their profiles. Video captions are initially recognized using Optical Character Recognition (OCR), then a set of key passages with corresponding frame images are extracted to form a video summary. The recommendation is achieved by calculating the relevance of the video summarization for each learner. The usage concerning for learning and instruction is presented. Also, a pilot study to evaluate the proposed system was carried out. The pilot study revealed that the proposed system has a positive effect on learners' interest of viewing video material. In the result, light video watchers were motivated to increase the frequency of watching video and heavy video watchers were motivated to spend more time to watch video through this system. Learners also agreed that summarization in video recommendation email is a good and convenient way of acquiring knowledge.

Introduction

Multimedia instruction has recently become a promising information source to the traditional instruction. Many studies reported that multimedia content is useful for learning and teaching, in comparison to traditional in-class, and text-based learning (Choi, 2005; Mackey & Ho, 2006; Rose, 2003). There are many types of multimedia learning material, but the use of video is a more effective medium for learner motivation, attention, and satisfaction (Choi, 2005; Mackey & Ho, 2006). Unlike text, reading videos requires much more time since it is displayed linearly. Video summarization therefore enables the learner to skim through the video content (Choi, 2005). According to above, the effectiveness of multimedia instruction is obvious, and it is necessary to stimulate learners' motivation by enhancing multimedia learning tool mechanism.

With the rapid growth of video nowadays, it is a difficult task of acquiring appropriate video from huge amount of videos. Adaptive recommendation is mainly designed to help the learner to filter out information. Traditionally, these tasks, for example content annotation and recommendation in e-learning are done manually, which is very time-consuming. Therefore, there is a strong demand for automatic video summarization and recommendation. Learning material recommendation can provide adaptive learning objects easily and efficiently for learners to improve the learning effect. Without recommendation mechanisms, learners will spend much time in selecting suitable learning objects. It had been studied that the automatic recommendation mechanisms is positive for learning object recommendation via comparing the learner profiles (Tsai, Chiu, Lee & Wang, 2006). The learning object features in these mechanisms are pre-defined and the content of features is usually just a simply text description. In the past, learners could select their interested films as their learning content, however, traditional text-based recommendation does adopt merely article names and titles while ignores the important contents inside the learning materials. Even though some studies (Choi, 2005; Liu & Li, 2002; Milrad, Rossmanith, & Scholz, 2005) provide the human-annotated objects, they were not automatic process. For larger video database, like digital library, the annotation process might take huge time on "watching" and "writing".

Automatic summarization is an important research topic, especially to automatic text-based and video-based summarization. Text-based summarization aims at abstracting important sentences from source document. These techniques focus on generating summaries from news-like articles that are usually short and coherent. However, video content is quite different from news texts since it is not only long (roughly 7000 words) but also contains multiple sub-topics. On the contrary, video-based summarization techniques offer a sketch with description

of an object (Liu & Li, 2002), such as color, shape, etc. Such techniques are often used in surveillance system and medical videos. Nevertheless, they may not work well for text-based video and also not useful for learners due to ignore lexical information. Besides, the traditional video-based summarization does not attach to educational purpose.

Motivation is an important factor for successful learning. The ARCS (attention, relevance, confidence, and satisfaction) model of motivation was formed in response to find more useful ways of understanding the major factors on the motivation to learn (Keller, 1983). This model defines four major conditions (attention, relevance, confidence, and satisfaction) that have to be fulfilled to become and remain motivated (Dick, Carey & Carey, 2001). Through video recommendation, we hope to attract learners' attention, recommend relevant video, and promote learners' confidence and satisfaction in effective way.

In view of the preceding research literature, we could understand that the multimedia learning is useful for learners, but there is not a customized tool for multimedia learning, and there is no a mechanism to stimulate learners motivation. In this paper, we present an automatic adaptable multimedia content recommendation system to auto-recommend suitable multimedia learning materials to encourage learners to watch and learn. The proposed system firstly extracts video content as summarization, while the generated summarization with corresponding frames was collected. These materials are combined into the hypermedia documents and auto-recommend to learners. The system also sends the hypermedia document as email to learners in response to their profiles. Unlike traditional recommendation methods, the system does not only recommend video titles, but also included extracted important content that contain summarization and corresponding image frames. While dealing with a great deal of videos, the system can extract summarization rapidly and save time. Besides, it can recommend learners video materials probably related to what they have learnt and taught. Thus, learners can quickly take the new video information they need instead of receiving a lot of unnecessary information.

For these objectives to be achieved, the article is structured as follows. Section 2 describes the proposed automatic adaptable multimedia content recommendation system. Section 3 offers examples of system usage. Section 4, 5 presents a pilot study and discussion. Finally, the conclusion and future work are drawn in the section 6.

System Architecture

An overview of the proposed automatic adaptable multimedia content recommendation system is illustrated in Figure 1. Once a new video is incoming, the *Video OCR Module* starts to recognize captions as video caption document. These documents are then passed to a *Summarization Module* and the summary document for the video is generated by extracting the key passages. Finally, the video recommendation emails are generated by the *Recommendation Module*, which estimates the relevance for each learner according to their profiles. By combining with these three modules, the system can automatically generate and send the video recommendation emails when a new video incomes. In other words, the entire process in the proposed system is automation without any human intervention. Each of the three modules is described as follows.

Video OCR Module

Video OCR Module processes the input video frame sequence and recognizes all the caption words. Video images can provide rich visual information to people. The video speech content, however, plays a more significant role for video content understanding. In many educational films, such as Discovery and National Geography, usually rich caption information is presented which is very close and sufficient to describe current video scenario. In this case, we extract video caption words as video speech content for further processing.

By employing off-the-shelf video Optical Character Recognition (video OCR) techniques, these caption words could be automatically identified. In this paper, we use the OCR systems as (Wu, Lee, Yang & Yen, 2006; Lee, Wu & Chang, 2005) for caption content extraction. As reported by (Wu, Lee, Yang & Yen, 2006), the performance of video OCR was about 70-80%. Finally, the recognized words formed the video caption document.

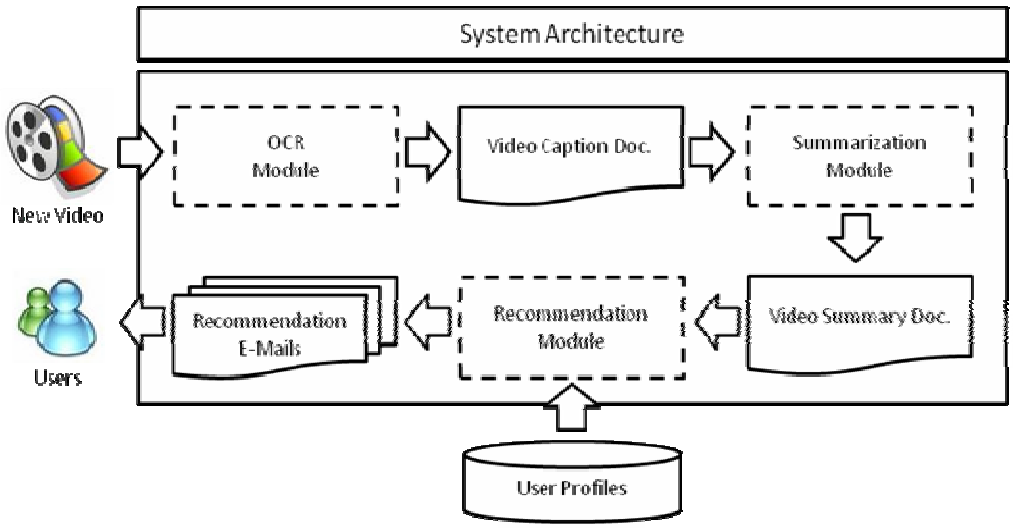

Figure 1. The proposed system overview.

Summarization Module

Summarization Module processes the video caption document from the previous module. Over the past few years, text summarization was thoroughly studied by researchers and developed well. However, most of the traditional summary generation methods aim to extract set of key sentences from documents, which are not the case for video recommendation. Because the key sentences in the video are less meaningful and usually not what users concerned with and interested in. In this paper, we adopted the Q/A-based (Question & Answering) approach (Lee, Wu & Chang, 2005), which is more likely to route what users want to know, for generating summaries.

In this module, the video caption document is initially segmented into five segments based on the time sequence analysis (Wu, Lee, Yang & Yen, 2006). For each segment, the video Q/A system (Lee, Wu & Chang, 2005) extracts passage-level answers to form the video summary document. Here, we assume that using the answers of each segment as the video summarization could be more complete and comprehensive for video content understanding. In order to enhance the readability, the top-5 ranked video summarizations are provided for users.

Recommendation Module

Recommendation Module processes the video summarization generated by the above two modules. In this module, it compares the video summarization with the user profiles, which record each user's personal information, such as name, email, interest, and so forth. An XML-based format is used to store each user's data. Figure 2 illustrates a fragment of user profiles as an example. Users can edit their profile at anytime while they want to modify it.

For the recommendation, we focus on calculating the relevance of the video summarization for users. In order to match users more effectively, we integrate video name, video description, and video summarization as the sources for comparing with the profiles. The similarity measurement is estimated with cosine value (Baeza-Yates & Ribeiro-Neto, 1999). In order to achieve adaptive personalization, the system does not comprehensively notify readers with all video. Instead, the related videos were presented to users, this reduce and prevent the system mails to become spam mails. When a new video is incoming, the recommendation module compare the content of this video with all users. By means of comparing the user profiles, and the generated video summaries, only partial of the readers will receive the recommendation mail. The key inside is the similarity measurement between the two information source. The similarity measurement used in this paper is the cosine value and describe as follow:

$$Sim(UP_i, SUM_j) = \frac{dot(UP_i, SUM_j)}{\left\|\overrightarrow{UP_i}\right\| \cdot \left\|\overrightarrow{SUM_j}\right\|}$$

Where $dot(X,Y) = (x_1 y_1 + x_2 y_2 + ... + x_n y_m)$, $\|X\|$ is the one-norm of the vector X and can be estimated as follows:

$$\|X\|=\sqrt{x_1^2+x_2^2+...+x_n^2}$$

Each of the x_i denotes as the word of either in summary or in profile. $Sim(X,Y)$ computes the similarity (i.e., cosine value) between the two variables X and Y. Term UP_i and SUM_j denote as the user profile for person i, and summary j for the j-th new videos. For each reader, we use the following function to determine relevant score to a specific summary j, SUM_j.

$$RevScore(i, SUM_j)=Sim(UP_i, SUM_j)-\theta$$

θ is a pre-define threshold. If RevScore > 0, then the system considers the video is positive-relevant the user, and will send the recommendation mail. Otherwise, the system skips the user. In other words, the higher relevant score the more likely this video summarization is what the user interests in. If the likelihood score exceeds a threshold, it will send the auto-generated video recommendation emails to the users.

```
<?xml version="1.0" encoding="big5" ?>
- <profiles>
  - <user id="VCSR_USER_0001">
      <firstname>意婷</firstname>
      <lastname>黃</lastname>
      <email>coral@cl.ncu.edu.tw</email>
    - <interests>
        <item>歷史</item>
        <item>軍事</item>
        <item>地理</item>
      </interests>
      <majorSubject>網路學習科技</majorSubject>
      <favoriteMovie>Independence Day</favoriteMovie>
      <others />
    </user>
  + <user id="VCSR_USER_0002">
  + <user id="VCSR_USER_0003">
    ...
</profiles>
```

Figure 2. XML-based description of user profiles.

System Usage

In order to explain the system usage, we provided the following scenarios as examples for how the proposed system helps to recommend new video, and how the proposed system motivates students to learn.

There are often multimedia video materials ordered newly in a library, such as Discovery videos. However, if the library wants to make a video introduction, it can only refer to the simple description of the video cover. Otherwise, the library needs to spend more efforts and time viewing and annotating the video. In addition, it is another issue that how to recommend the new video to those people who need. Using the proposed system, all of the above problems can be turned into automated process. By using the system we purposed, librarian could input new video as a source and then system automatically produce video summarization with corresponding image frames directly. This solution could save time and human efforts. Next, based on the profile information, system will automatically send emails to users who may need this information. As shown in Figure 3, the email content is composed of video cover, video description, hyperlinks to top-5 ranked video summaries, and the video summaries: summary text and key frames in corresponding to the text.

Linda is a geography major. She usually wasted most time on the Internet, BBS, and free downloadable game. She also liked video but never knew that there are many videos related to her major. These videos contained much information of geography. If she used the system and filled out her major in profile, she would get latest geographic video information from recommended mail. Such information could engage Linda in spending more time on her major through seeing video and might potentially improve her learning further.

Tom is an undergraduate student. He much likes to see the movie when he has free time. He often spent four hours in watching movies on TV a day and saw four movies a month. He enjoyed learning information from video. Through watching Discovery video, for instance, he made sense that there are two kinds of pyramid in Egypt: step pyramid and real pyramid. He understood that the honeybee's brain can marks and draw the time and space accurately from National Geographic channel. Multimedia video made a great impact on Tom. However, He also wasted much time to collect what video he wanted. After using the system he would receive recommended mail which fits for his interest. Even he might get some video information he never attended to in the past. The mail includes simple description, key frame capture, summarization of video section, hyperlink to view these video sections briefly. Thus he could avoid wasting time on uninterested video. Tom likes this system because he can gather much information of videos which match his interest. He could efficiently spend his time and attention in suitable video to improve his knowledge.

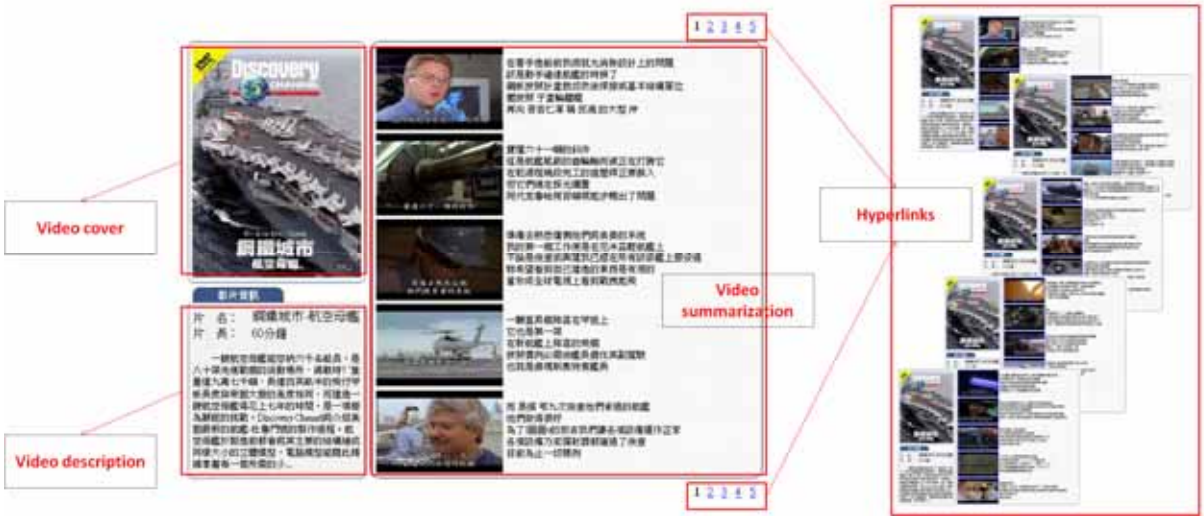

Figure 3. An illustration of the recommendation email content.

Pilot Study

Method

The Discovery video data is one of the popular video learning materials. 181 Discovery films were selected as the system video sources. Sixteen subjects were employed, all of whom were either college students or college graduates between the ages of 18 and 25 years. The subjects were first asked to fill out a questionnaire on their video watched habits. The purpose of the questionnaire was to distinguish subjects' video watched habits. We divided the subjects into three groups. One is light video watcher group, another is medium video watcher group, and the other is heavy video watcher group. We would like to know the different influence between light and heavy video watcher group. Afterwards, the subjects were beginning to use the system. First, they were asked to fill out a table on their interests and email address. The following are two sub-experiments that subjects all need to do. Experiment 1, they had got recommendation emails from the system, containing the video cover image and short descriptions (derived from the video cover) according to their interests. After reading the email, they were asked to fill out a questionnaire on their experiment one. The questionnaire was designed according to the strategies of ARCS model (Keller, 1983; Keller & Kopp, 1987). Experiment 2, they also had got recommended email. The difference of second email is contained video cover image, short descriptions, the extracted summaries (roughly < 15 sentences and the image frames), and video clips which subjects could click the images to watch the online video clips. Afterwards, the subjects were also asked to fill out a different questionnaire which was designed according to the strategies of ARCS model on their experiment two. Subjects were interviewed to understand the preferences of the functions between the two emails. The procedure of the experiment is illustrated in Table 1.

Through the questionnaire before experiment (Step 1), we expect to confirm all subjects in being experienced the approach of the traditional querying video. Also, we could divide the level between the video-

watching habits of all subjects. Therefore, we could focus on what the system affect the different subjects' motivation. In the questionnaires (Step 3, 5), they are designed by ARCS model in order to research how the system improves the motivation of multimedia learning of the subjects.

Table 1: The procedure of the experiment.

Step 1	Fill out questionnaire on video watched habits
Step 2	Experiment One
Step 3	Fill out questionnaire on using experience
Step 4	Experiment Two
Step 5	Fill out questionnaire on using experience
Step 6	Interview

The extended appliance of ARCS

According to the ARCS model, we probed into the model to apply in the questionnaire design. There are four major categories (i.e., attention, relevance, confidence, satisfaction) influence a student's motivation to learn (Keller, 1983; Keller & Kopp, 1987). We identified the features as follows: Attention (A) refers to the extent to which the learners' attention is aroused. What gained the main functions (e.g. the email title, the email format e.g.) learners' attention? Relevance (R) refers to the learners' perception that the content of the recommended email is related to personal needs or past experiments; Confidence (C) refers to the learners' perceived likelihood of achieving their expected goal after using the system; Satisfaction (S) refers to the system's preference of using experience of learners. Keller (1983) augured that design of instruction must appeal the passionate attraction of students. For students, teaching material and learning activity are engaging attention. The personal expectation of a student, motivation and the material of motivated strategy all influence how students engage in the learning activity. Thus, the study examines the effectiveness of providing the system by ARCS model. The research questions are "Would the system could improve learners' motivation?" and "if yes, what the main function of the system does improve learners' motivation?"

Results

The result of the questionnaires are presented in Table 2, we divide the categories of the questionnaires into four factors of attention and the functions of the system which are Summarization and Recommendation.

Table 2: The result of the questionnaires.

	Experiment 1		Experiment 2		
Categories	Mean	SD	Mean	SD	Variation
Attention	3.54	1.07	4.28	0.66	0.74
Relation	4.00	0.52	4.25	0.77	0.25
Confidence	4.13	0.62	4.38	0.62	0.25
Satisfy	3.69	0.71	4.21	0.67	0.52
Summarization	3.31	1.14	4.13	0.62	0.81
Recommendation	4.00	0.82	4.13	0.72	0.13

Attention The results suggested that experiment 2 on the average (M = 4.28, SD = .66) used higher attention than those in the experiment 1 (M = 3.57, SD = 1.07). Because the content from email in experiment 1 only contained the video cover image and short descriptions (derived from the video cover), but the content from email in experiment 2 contained video cover image, short descriptions, the extracted summaries (roughly < 15 sentences and the image frames), and video clips which subjects could click the images to watch the online video clips. It is more abundant in multimedia resource to catch learners' eyes.

Relation The relation of ARCS model represented that the recommendation whether can connect the prior experiences of learners. It examined the ability of the recommendation of the system and the video we recommended whether link the past experiences of learners. The results in experiment 1 (M = 4.00, SD = .52) and 2 (M = 4.25, SD = .77) showed that the relation was generally well received by students.

Confidence The confidence of ARCS model meant that the function of the system provided the content is not too simple and not too difficult, so learners would be confident of their ability to develop their interest. Thus, the ability of the recommendation and summarization of the system played important roles in facilitating to collect the information and deepen their interest. In experiment 1 (M = 4.13, SD = .62) and experiment 2 (M = 4.38, SD = .62), the results have been very positive.

Satisfy According to the results of questionnaire, 84% of subjects in experiment 2 (M = 4.21, SD = .67) strongly agreed with the system. They expressed that the system would increase frequency of watching video and extend time of watching video because of the abundance of email, containing colorful photos, detailed summaries and video clips. Therefore, they hope own the system in future.

Summarization Many subjects considered that the function of the system is most useful, because the system summarized the summarization for every video clips and learners could understand directly the story in the video without watching. They didn't need to spend their time watching, and could get information what they needed. Clearly, the findings indicated that the summarization in experiment 2 (M = 4.13, SD = .62) was significantly superior to those in experiment 1 (M = 3.31, SD = 1.14).

Recommendation The filter of recommendation is based on the categories which subjects filled out. More than 80% subjects agreed the benefit of recommendation because they could automatically get information which they are interested. Three subjects expressed that the content of recommendation was positively related to motivation. Therefore, the accuracy of recommendation could raise learners' interest.

Discussion

Multimedia content can support learners' learning by helping them identify emergent goals within a context. From a situated approach to learning, learners' knowledge is shaped by meaningful overall situation. This paper aims to motivate learners' learning desire and push them learn invisibly. As the results showed, learners were willing to receive the information they were concerned. Due to the information in their hand, learners would watch them and get knowledge imperceptibly. The system in experiment 1 was the first version, but the system in experiment 2 was the advanced version. Learners receiving the content of email showed significant improvement in the system of the experiment 2 compared to the system in the experiment 1. The functions of the system in experiment 2 composed of video cover image, short descriptions, the extracted summaries (roughly < 15 sentences and the image frames) and video clips. Those are more versatility than that of the system in the experiment 1 which consisted of the video cover image and short descriptions (derived from the video cover). The most difference between version 1 and version 2 are summarization and video clips. A large number of subjects reported favorably on the effectiveness of summarization and video clips. They expressed that the main reason which they wanted to watch the video is the detailed summarization and video clips. At first, learners could grasp the meaning of story through every summarization instead of watching whole video, so they could predominate exactly whether watch or not. Before the experiment, we have an assumption that learners would not watch the recommended video after reading the summarizations because they had known everything in the video. The result was contrary to our assumption. In fact, summarizations revealed the recommended email is closely connected to learners' motivation. Secondly, some subjects stated that they like to watch animate multimedia more than read textual description. Learners could notice more details when watching video. The colorful fames and sound effects could attract learners' attention, thus they absorb knowledge imperceptibly instead of abstract description.

The sixteen participants involved in the experiment were categorized by their level of their video watched habits, two being classified as light level subjects and two as heavy level according to the first questionnaire they had filled out on video watched habits yet. According to the results, most subjects agreed that the system would increase their motivation to watch recommended video, deepen and broaden their knowledge. What the different effect on between heavy and light level subjects? In the light level subjects, some subjects said that they were not interested in watching a video in the past. After trying the system, they expressed that they had stimulus to watch. It was easy to receive source and explore something they had not known before. The heavy level subjects highly

praised the recommended function and video clips of the system because they really needed a tool to help them finding new multimedia material and confirming what they wanted. In the past, they had spent much time finding multimedia data. They could understand the content of video only after watching them. Now, they can spend more time watching more video instead of searching them. However, they mentioned that they hoped to share their opinions of the video in a community, and could read others' ideas from a community. Although the system didn't provide this function for discussion, subjects were observed that a shift changed from passive to active. Because they felt like participate in, not being a submissive role but an energetic role in the system. In sum, the system effects light level subjects to increase the frequency of watching video and heavy level subjects to have more time to watch video.

Conclusion

In this paper, we present an automatic and adaptable multimedia content recommendation system to auto-recommend suitable multimedia learning materials for learners. In addition, it can automatically recommend videos based on user profiles. The result of the pilot study reveals that the video summarization is a good and convenient way of acquiring knowledge in particular the unstructured data, like videos. For light video watcher, the system encourages them to watch; for heavy video watcher, the system helps them collect video material. Hence both of them can obtain prior knowledge more convenient in comparison to past experience.

In the future, we plan to build a community for learners sharing their knowledge. Learners will not play a passive role in the system, but an active watcher. Based on each user's profile information, we also plan to customize the video summarization for each user. This implies that different user can obtain different summarization he is interested in even though the video source is the same. Finally, an effective algorithm to improve OCR errors is necessary for upgrading the quality of the summarization of the system.

References

Baeza-Yates, R., & Ribeiro-Neto, B., (1999). *Modern Information Retrieval*. Addison-Wesley Longman Publishing Company

Choi, H. J., & Johnson, S. D. (2005). the Effect of context-based video instruction on learning and motivation in online course, *The American Jouranl of Distance Education*, 19(4), 215-227.

Dick, W., Carey, L., & Carey, J. O., (2001). The systematic design of instruction (5th ed.) New York: Addison-Wesley.

Jin, R., Si, L., Zhai, C., & Callan, J., (2003). Collaborative Filtering with Decouple Models for Preferences and Ratings, *International Conference on Information and Knowledge Management*

Keller, J. M. (1983). Motivational design of instruction. In C.M. Reigeluth (Eds.), *Instructional-design theories and models: An overview of their current status*. Hillsdale. NJ: Lawrence Erlbaum Associates.

Keller, J. M. , & Kopp, T (1987). An application of the ARCS model of motivational design. In C. Reigeluth (Eds.), *Instructional theories in action: Lessons illustrating selected theories and models*. Hillsdale. SJ: Lawrence Erlbaum.

Lee, Y.S., Wu, Y.C., & Chang, C.H., (2005). Integrating Web Information to Generate Chinese Video Summaries, *Software Engineering and Knowledge Engineering*, 514-519.

Liu, Y., & Li, F., (2002). Semantic Extraction and Semantics-Based Annotation and Retrieval for Video Databases, *Multimedia Tools and Applications*, 5-20.

Mackey, T. P., & Ho, J., (2006). Exploring the relationships between Web usability and students' perceived learning in Web-based multimedia (WBMM) tutorials, *Computers & Education*, Article in Press, Corrected Proof.

Milrad, M., Rossmanith, P., & Scholz, M. (2005). Implementing an Educational Digital Video Library Using MPEG-4, SMIL and Web Technologies. *Educational Technology & Society*, 8 (4), 120-127.

Tsai, K.H., Chiu, T.K., Lee, M.C., & Wang, T.I., (2006). Learning Objects Recommendation Model based on the Preference and Ontological Approaches, *IEEE International Conference on Advanced Learning Technologies*, 36-40.

Rose, C., (2003). How to teach biology using the movie science of cloning people, resurrecting the dead, and combining flies and humans, *Public Understanding of Science*, 12(3), 289-296

Wu, Y.C., Lee, Y.S., Yang, J. C., & Yen, S.J., (2006). A New Passage Ranking Algorithm for Video Question Answering, *Lecture Notes in Computer Science (LNCS): Advances in Image and Video Technology*, 4319, 563-572.