Yang, J. C., Huang, Y. T., Tsai, C. C., Chung, C. I., & Wu, Y. C. (2009). An Automatic Multimedia Content Summarization System for Video Recommendation. *Educational Technology & Society*, *12* (1), 49–61.

An Automatic Multimedia Content Summarization System for Video Recommendation

Jie Chi Yang*, Yi Ting Huang*, Chi Cheng Tsai*, Ching I Chung* and Yu Chieh Wu**
*Graduate Institute of Network Learning Technology,
**Department of Computer Science and Information Engineering,
National Central University, No.300, Jungda Rd., Jhongli City 320, Taiwan // Tel: +886 3 4227151 ext.35414, Fax: +886 3 4275336 // *{yang, coral, ken, cecilia}@cl.ncu.edu.tw, **bcbb@db.csie.ncu.edu.tw

ABSTRACT

In recent years, using video as a learning resource has received a lot of attention and has been successfully applied to many learning activities. In comparison with text-based learning, video learning integrates more multimedia resources, which usually motivate learners more than texts. However, one of the major limitations of video learning is that both instructors and learners must select suitable videos from huge volume of data. The situation is worsened by the fact that video content is usually displayed linearly. In order to solve this problem, we propose a multimedia content summarization and adaptable recommendation framework which is able to extract summaries from raw videos and recommend them to learners automatically. By following the characteristics of multimedia, the generated summary contains both important abstracts and corresponding images, and can be accessed online. Only suitable videos are selected for recommendation based on user profiles. The proposed system is evaluated and compared to text-based learning in terms of ARCS model. The results demonstrate that the proposed video summarization and recommendation framework was not only positive with regard to motivating learners, but also enhanced the video learning experience significantly. Positive results are also found in relation to system usage and satisfaction.

Keywords

Video summarization, Video recommendation, Learning material recommendation, Multimedia, Video learning, ARCS model

Introduction

Multimedia-based learning has recently become a promising instructional resource. According to dual-coding theory, the human brain deals with imagery representation better than verbal representation (Clark & Paivio, 1991; Paivio, 1971). Many studies report that multimedia content is often more useful for learning and teaching than traditional text-based learning (Mackey & Ho, 2008; Mayer & Moreno, 2002; Rose, 2003). There are many types of multimedia learning materials, but video is the most representative and popular one. This is because video integrates many multimedia resources, such as text, images, sound, and speech. Based on the theory of constructivism, video provides a context wherein learners can construct their own knowledge (Brown, Collins, & Duguid, 1989). Several studies demonstrate that video is a suitable material for context-based learning (Choi & Johnson, 2005; Choi & Johnson, 2007). In addition, video learning is an effective way of providing motivation, keeping attention, and giving satisfaction to the learner (Choi & Johnson, 2005; Choi & Johnson, 2007; Mackey & Ho, 2008). However, there are some limitations that exist in video learning. For example, choosing a suitable video for instructors and learners from the rapidly growing number of videos can be a problem. A similar problem arises when recommending videos to learners and instructors. Moreover, when learners get a great deal of information in a short period of time, it causes cognitive overload (Pass & van Merrienboer, 1994; Sweller & Chandler, 1994; Sweller, van Merrienboer & Pass, 1998). The usefulness of video learning is obvious, but it is necessary to enhance the recommendation mechanism for learners in order to facilitate multimedia learning.

Unlike texts, watching videos requires much more time since video content is usually displayed linearly. Summarization is used to preserve the most informative parts of the source content. Therefore, video summarization is essential for enabling the learner to skim through video content. With the rapid growth of the video industry, acquiring the appropriate video from a huge database is a difficult task. Adaptive video recommendation could be a way to deal with this situation, as this system is specifically designed to help learners filter information. In essence, the combination of summarization and recommendation is helpful in reducing cognitive overload with regard to video learning. Traditionally, these tasks (i.e., content summarization and recommendation in e-learning) are done manually, which is a very demanding and time-consuming process. Consequently, there is a strong demand for an automatic video summarization and recommendation system.

Recently, recommendation systems have been applied to some products and information databases by making adaptive suggestions based on previous examples of a user's preference (Melville, Mooney & Nagarajan, 2002; Mooney & Roy, 2000; Wang, Tsai, Lee & Chiu, 2007). A recommendation system for learning material can provide objects easily and efficiently thereby enhancing learning activities. Without recommendation mechanisms, learners would spend more time selecting suitable learning objects and less time involved in the actual activity of learning. Several studies have shown (Tsai, Chiu, Lee & Wang, 2006; Wang et al., 2007) that automatic recommendation mechanisms that refer to learner profiles can promote the accuracy of learning object recommendation. Nevertheless, these recommendation mechanisms are only suitable for structured or semi-structured data (Popescul, Ungar, Pennock & Lawrence, 2001; Tsai et al., 2006; Wang et al., 2007). In other words, these systems may not work well with raw videos and raw texts. On the other hand, recommendation systems are also applied to movie or TV recommendations (Alspector, Kolcz & Karunaithi, 1998; Basu, Hirsh & Cohen, 1998; Cotter & Smith, 2000; Melville et al., 2002). However, these studies tend to apply very limited sets of features such as the movie title, the director, keywords, and actors, as well as like-minded user ratings. The feature that is lacking in all of these studies is the inclusion of speech content, which contains a substantial amount of information relating to the video itself. In other words, the studies mentioned above ignore important content within learning materials.

Automatic summarization is an important research topic, especially in relation to automatic text-based and video-based summarization. Text-based summarization research, such as the Document Understanding Conference (DUC) (http://duc.nist.gov/), aims at extracting important sentences from source documents. These techniques focus on generating summaries from news-like articles (i.e., newspaper and newswire data) (Dang, 2006; Dang, 2007) which are usually shorter and more coherent than video stories. Moreover, a video story usually contains multiple sub-topics. On the contrary, video-based summarization research, such as the TRECVid workshop (http://www-nlpir.nist.gov/projects/t01v/), aims at extracting key-frames and shots from source videos (Over, Ianeva, Kraaij & Smeaton, 2005; Over, Ianeva, Kraaij & Smeaton, 2006), offering a sketch that contains a description of an object (such as color, shape, or motion) (Liu & Li, 2002; Milrad, Rossmanith & Scholz, 2005; Over et al., 2005; Over et al., 2006). This technique is often used in surveillance systems (Osadchy & Keren, 2004; Piriou, Bouthemy & Yao, 2006) and medical videos (Fasquel, Agnus, Moreau, Soler & Marescaux, 2006). Nevertheless, these types of summarization may be not useful for learners due to the neglect of video content. Furthermore, the traditional video-based summarization is not generally used for educational purposes.

In order to help learners choose videos that are suitable for specific learning activities, two issues must be addressed: (a) the appropriate summarization technique must be able to extract semantic information from video content; and (b) the appropriate system must be able to recommend a suitable video to learners from a huge database. The research regarding both of the above points is very limited, and few attempts have been made to apply summarized information to recommendation mechanisms. For example, MovieLens is a well-known movie recommendation website (http://movielens.umn.edu). It generates personalized recommendations on the basis of a user preference. Nevertheless, the recommendation information provided by the site lacks an integral plot summary. Therefore, users might not be able to browse a sufficient amount of information in order to determine whether the movie is related to the knowledge that interests them. YouTube (http://www.youtube.com/), while employing a different system, still falls short with regard to providing videos that are appropriate for specific learning activities. This famous online video streaming service allows anyone to view and share videos that have been uploaded by others. Users can get videos by searching keywords on the website. Unfortunately, users will likely spend a lot of time looking for related videos through the search mechanism rather than receiving relevant information from the recommendation mechanism. Due to such situations, attaining knowledge efficiently through videos may prove to be a difficult task that could even lead to a decrease in motivation on the part of the learner.

Motivation is an important factor for learning. The ARCS model of motivation was formed in response to the necessity of finding more useful ways of understanding the major factors relating to the motivation to learn (Keller, 1983; Keller & Kopp, 1987). This model identifies four major factors: attention, relevance, confidence and satisfaction. All of these factors must be fulfilled if a learner is to become and remain motivated (Dick, Carey & Carey, 2001). Based on the effectiveness of multimedia learning, we hope to develop a video recommendation system that attracts the learner's attention, recommends relevant videos, and effectively promotes learner confidence and satisfaction.

It is clear that multimedia learning is useful for learners, but there is not a customized tool or mechanism for multimedia learning that perpetuates learner motivation. In this paper, we extend our previous works (Huang, Tsai, Chung, Shen, Yang & Wu, 2007; Tsai, Chung, Huang, Shen, Wu & Yang, 2007) and present an automatic multimedia content summarization and adaptable recommendation system, called Video Content Summarization for Recommendation (VCSR), that auto-recommends suitable multimedia material with the aim of encouraging learners to watch and assimilate knowledge within the framework. The proposed system first extracts video content as a summary and collects corresponding frames from the source. These materials are combined into a hypermedia document and auto-recommended to learners. The system also sends the hypermedia document as email (multimedia-based email) to learners in response to their profiles. Unlike traditional recommendation methods, the system not only recommends video titles, but also includes important extracted content that contains a video summary and corresponding video clips. The system can extract information rapidly from a large database of videos, saving time for the user. Moreover, the system can recommend video material to learners related to what they wish to study. Thus, learners can quickly use the new video information acquired instead of receiving a lot of unnecessary information.

In order for these objectives to be achieved, this paper is structured as follows: Section 2 describes a scenario for explaining the system usage and system architecture of the proposed VCSR system, and Section 3 describes the method and results of the experiment. Discussion of the results is presented in Section 4, and the conclusion is drawn in Section 5.

VCSR System

In this section, a scenario is given to explain how the proposed VCSR system recommends new videos to users and what kind of role the system plays. Next, the system architecture is described in detail using three modules. Finally, the inherent differences among various video recommendation systems are compared in order to identify the functionality of the VCSR system.

Scenario: An Intermediate Role in a Digital Library

Usually, there are many video resources in a library. Whenever a new video arrives, the librarian needs to make a video introduction. One way that he or she can do this is to refer to the simple description on the video cover. The only other way to extract the necessary information from the video would be to watch it. It is important to note that information taken from a video cover may not be detailed enough to make an appropriate recommendation based on a learner's interests.

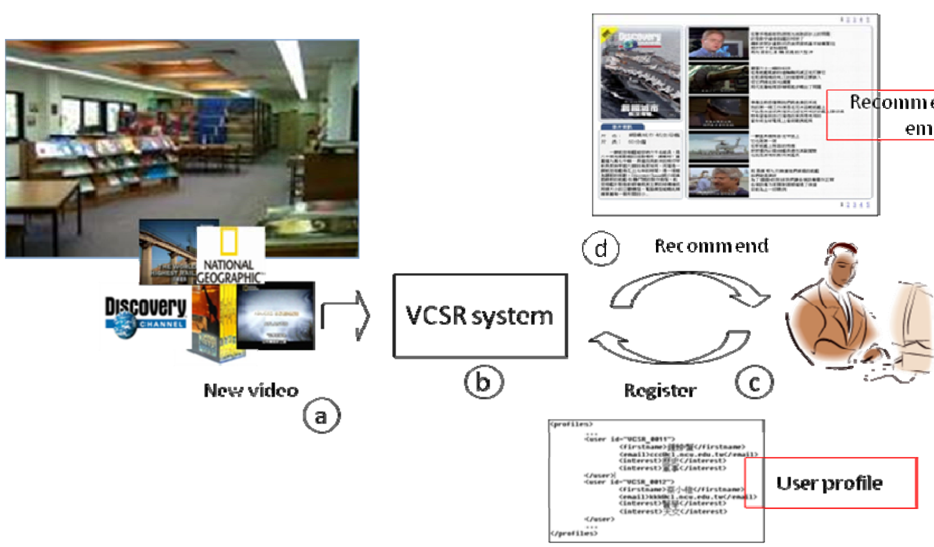

Figure 1. VCSR system workflow using a digital library as an example

Using the proposed VCSR system, all of the above problems can be turned into an automated process. Figure 1 describes the automated workflow, which is mediated by the VCSR system. First, the library receives the ordered video resource regularly. (a) When a new video is incoming, the system will receive it as an input. (b) The system will then automatically produce a video summary. (c) The summary is then compared to profile information registered by users, and (d) users receive a multimedia-based recommendation email, composed of the video cover, a video description, online video clips, hyperlinks to the top 5 ranked video summaries, and the video summaries themselves, consisting of text and key frames that correspond to the text (see Figure 2). In the recommendation email, users can click the image frames and watch the video clips online.

Figure 2. An illustration of the recommendation email content

System Architecture

The system architecture of the VCSR system is illustrated in Figure 3. Once a new video is incoming, the *Video OCR Module* recognizes captions as video caption documents. These documents are then passed to the *Summarization Module* and the summary documents for the video are 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 these three modules, the system can automatically generate and send video recommendation emails when there is a new incoming video. In other words, the entire process in the VCSR system is automated without any human intervention. Each of the three modules is described below.

Video OCR Module

The Video OCR Module processes the input video frame sequence and recognizes all the captions. Video images can provide rich visual information to people, but video speech content plays a more significant role in understanding video content. In many educational films, such as those produced by the *Discovery Channel* and *National Geographic*, rich caption information is an excellent way to describe the video content. Therefore, video captions are extracted as video speech content for further processing.

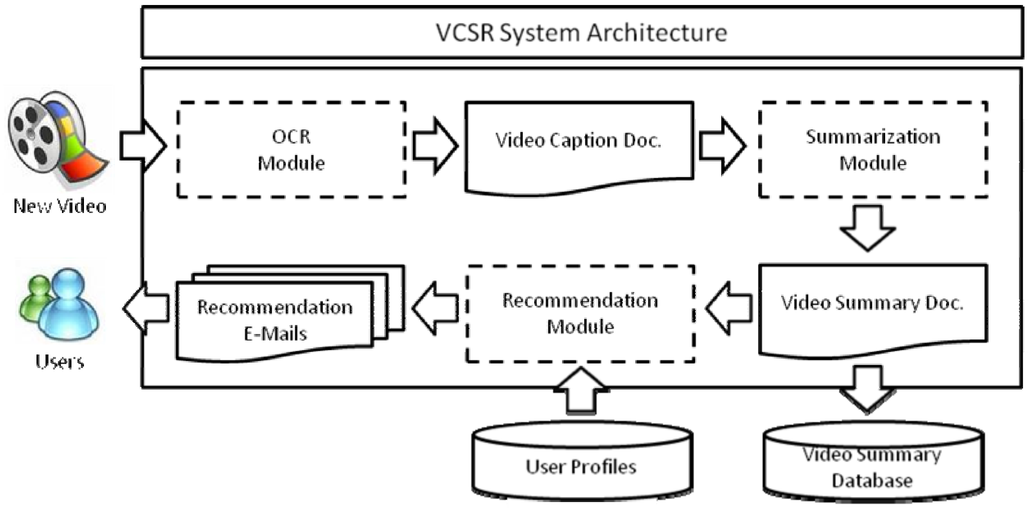

Figure 3. VCSR system architecture

VCD and DVD are the most popular video formats in the market. Since captions are embedded in these kinds of video formats, it is easy to extract them from video resources. By employing video Optical Character Recognition (video OCR) techniques, these captions can be automatically extracted and identified. Using video OCR techniques in the VCD format is a more difficult task than in DVD format due to the fact that captions and image frames are not independent. Because most of the video resources used in this study are VCD films, we have employed OCR systems (Lee, Wu & Chang, 2005; Wu, Lee, Yang & Yen, 2006; Wu, Yang & Lee, 2008) for caption content extraction. As reported in a previous study (Wu et al., 2006), the performance of such video OCR systems was about 70%-80%. The Video OCR module first decomposes the input video into a list of frames, and then performs the following steps: (a) filtering to remove the noisy blocks; (b) representing the character vector; and (c) character recognizing, in order to find the most similar characters. Finally, the recognized words form the video caption document.

Summarization Module

The Summarization Module processes the video caption document from the Video OCR Module. Over the past few years, text summarization has been thoroughly studied by researchers and is well-developed. However, most of the traditional summary generation methods aim to extract a set of key sentences from documents, but this is not the case for video summarization. This is because the key sentences in the video are less meaningful in an educational context and usually not what users are interested in. To overcome these problems, the Q/A-based (Question & Answering) approach (Wu et al., 2006; Wu et al., 2008) is adopted, which is more likely to route the information that users want to know, thereby enhancing the motivation to learn.

In this module, the video caption document is initially divided into five segments based on the time sequence analysis. Passage-level summarization is more suitable than sentence-level summarization with regard to video summary generation because video contains multiple sub-topics (Lee et al., 2005). For each segment, therefore, the video Q/A-based approach is adopted to extract passage-level answers and form the video summary document. Here, it is assumed that using each segment's answers as the video summary could be more complete and comprehensive in relation to understanding video content. In order to enhance the readability, the top 5 ranked video summaries are emailed to users. Finally, these summaries are stored in the video summary database for future browsing.

Recommendation Module

The Recommendation Module processes the video summary generated by the previous two modules. It then compares the video summary with the user profiles, which record each user's personal information (name, email address, interests, and major subject). An XML-based format is used to store each user's data. Figure 4 illustrates a fragment of user profiles as an example. Users can edit their profile at anytime through an interface.

```
  <?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 4. XML-based description of user profiles

For the recommendation, the focus is on calculating the relevance of the video summary for users. In order to match users and information more effectively, the video title, video description, and video summary are integrated as the sources for matching profiles with relevant video information. If the video content and the user profile match, the recommendation module will extract the email address in the profile and send a recommendation email automatically. In order to achieve adaptive personalization, the system does not comprehensively notify users of all videos. Instead, the related videos are presented to users, preventing the system emails from becoming spam. When a new video is incoming, the recommendation module compares the content of this video with all users. By means of comparing the user profiles to the generated video summaries, only a portion of the users will receive the recommendation email. The key is the similarity measurement between the two information sources. The similarity measurement is estimated with a cosine value (Baeza-Yates & Ribeiro-Neto, 1999) and described as follows:

$$Sim(UP_i, SUM_j) = \frac{dot(UP_i, SUM_j)}{\|\overline{UP_i}\| . \|\overline{SUM_j}\|}$$

Where $dot(X,Y) = (X_1Y_1 + X_2Y_2 + ... + X_nY_m)$, $\|X\|$ 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 x_i is a word in either the summary or the profile. $Sim(X,Y)$ computes the similarity (i.e., cosine value) between the two variables *X* and *Y*. The terms UP_i and SUM_j denote the user profile for person *i*, and summary *j* for the j-th new videos. For each user, the following function is used to determine a relevant score with regard to a specific summary *j*, SUM_j:

$$\mathrm{Re}\,vScore(i, SUM_j) = Sim(UP_i, SUM_j) - \theta$$

where θ is a pre-defined threshold. If *RevScore* > 0, then the system considers the video as positive and relevant to the user, and it sends the recommendation email. Otherwise, the system skips the user. In other words, scores of a higher relevancy are more likely to be what the user is interested in. If the likelihood score exceeds the threshold, it will send the auto-generated video recommendation emails to the users.

The Differences between Video Recommendation Systems

A comparison of the differences between the proposed VCSR system and other video recommendation systems is shown in Table 1. The VCSR system can receive raw video data and extract captions with the video OCR module. However, other systems only take artificially structured (e.g., title, genre, and user ratings) or semi-structured (e.g., description, summary, and review) data as input. The dataset is then compared with user profiles to select the suitable video. The VCSR system can automatically create video summaries from captions as a compared dataset with the summarization module. In other words, speech content is taken into account in the VCSR system. By contrast, other systems lack speech content summaries and ignore the importance of speech content when choosing videos. Moreover, other systems do not have an automatic summarization function. The manual summaries of other systems are usually subjective and influenced by advertisements. With the VCSR system, automatic summaries describe the video content more objectively. Finally, the VCSR system can play video clips online, which is another advantage over other systems.

Table 1. Comparison of video recommendation systems

System	Input data type	Compared dataset	Content summary	Video play function
MovieLens (http://movielens.umn.edu)	Structured	User ratings	No	No
Alspector et al., 1998; Cotter et al., 2000	Structured	Content feature	No	No
Basu et al., 1998	Structured	User ratings & Content feature	No	No
Melville et al., 2002	Structured & Semi-structured	User ratings & Content feature	Yes (manual summary)	No
VCSR system	Structured, Semi-structured & Unstructured (raw video)	Content feature (speech content especially)	Yes (automatic summary)	Yes

Experiment

Method

The purpose of the experiment is to examine whether or not the VCSR system can motivate learners to watch videos. The experiment was conducted to support the research hypothesis that the VCSR system can recommend adaptive videos, and that video summaries can increase the learner's motivation to watch videos. In this experiment, thirty subjects (nineteen male and eleven female) were employed, all of whom were either undergraduate students or graduate students between the ages of 18 and 25 years old. They came from different colleges of the National Central University in Taiwan, but most of the subjects were students of the College of Electrical Engineering and Computer Science. One hundred and eighty-one *Discovery* films were selected as the video sources because *Discovery* video data is one of the most popular learning materials. Two email types are used to examine the hypothesis: text-based emails and multimedia-based email. The text-based email contains a video cover image and short descriptions (derived from the video cover), as shown in Figure 5a. The multimedia-based email contains a video cover image, short descriptions, the extracted summaries (roughly 25 sentences and the corresponding image frames), and video clips (see Figure 5b). The former is mainly used as an adjunct to the typical text-based recommendation, while the latter will help us examine the impact and the effectiveness the VCSR system in comparison with the text-based recommendation.

The instruments used in this experiment are questionnaires that were designed according to the strategies of the ARCS model. The features of the ARCS model are identified as: *attention* (A), which refers to the extent to which the learners' attention is aroused; *relevance* (R), which refers to the learners' perception about whether the content of the recommendation email is related to personal needs or past experience; *confidence* (C), which refers to the learners' perceived likelihood of achieving their expected goal after using the system; and *satisfaction* (S), which refers to the system preference based on learners' user experience. Thus, this study examines the effectiveness of the proposed system by using the ARCS model. The research questions are, "Would the system improve learners' motivation?" and "Which function of the system improves motivation?"

Figure 5. Screenshots of (a) the text-based email nd (b) the multimedia-based email

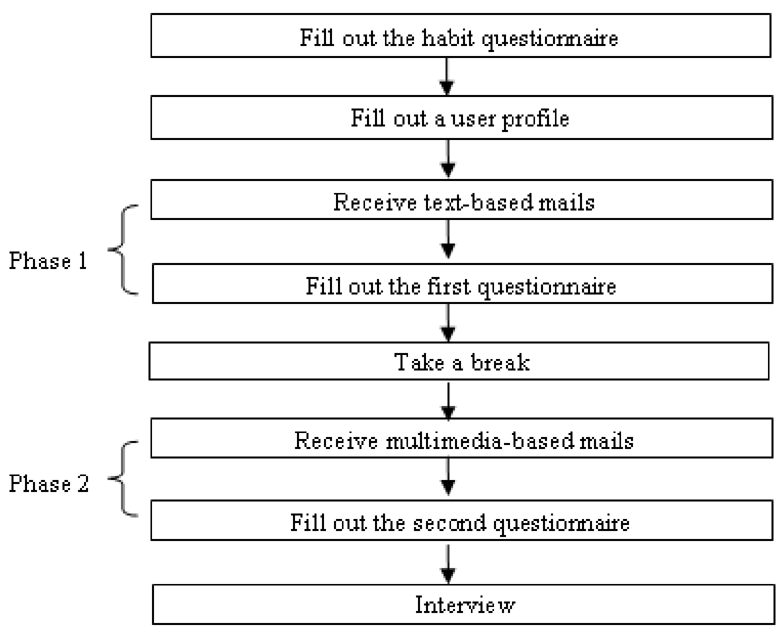

Figure 6. The experimental procedure

The experimental procedure is illustrated in Figure 6. A two-phase study was designed in order to explore learning motivation with regard to summarization and recommendation. In order to understand the subjects' video-watching habits and their personal information, they were first asked to fill out a habit questionnaire and a user profile which included their name, email address, and interests. Since the study has a within-subjects design, all subjects participated in both phases. Subjects could autonomously decide the experiment time by themselves, and the experiment was recorded by a camera for further analysis. In the first phase, subjects got text-based recommendation emails from the system. They were then asked to fill out the user experience questionnaire. Next, subjects took a rest before the beginning of the second phase. In the second phase, subjects got multimedia-based recommendation emails from the system. Similarly, they were then asked to fill out the second questionnaire after using the system. Finally, subjects were interviewed in an effort to understand their preferences in relation to the two types of emails.

Results

The results of the questionnaires are presented in Table 2. The categories of the questionnaires are divided into four factors relating to motivation (attention, relevance, confidence, and satisfaction) and two factors relating to system functions (summarization and recommendation).

Table 2. The results of the questionnaires.

	Phase 1		Phase 2			
Categories	Mean	SD	Mean	SD	Variation	p-value
Attention	3.54	.92	4.17	.75	.62	.000***
Relevance	4.03	.56	4.48	.64	.45	.029*
Confidence	4.03	.76	4.30	.65	.27	.013*
Satisfaction	3.71	.71	4.12	.68	.41	.000***
Summarization	3.37	.96	4.10	.66	.73	.000***
Recommendation	3.87	.90	3.93	.87	.07	.313

*$p < .05$; **$p < .01$; ***$p < .001$.

Attention: This factor focused on whether the recommendation emails aroused learners' attention. The results showed that the Phase 2 emails, on average (M = 4.17, SD = .75), received higher attention than the Phase 1 emails (M = 3.54, SD = .92). There is a significant difference between Phase 1 and Phase 2 with regard to the attention factor ($p < .001$).

Relevance: Relevance refers to whether or not the recommendation emails connected the learners' prior experiences. The results showed that Phase 2 had a higher mean (M = 4.48, SD = .64) than Phase 1 (M = 4.03, SD = .56). There is a significant difference between the two phases with regard to the relevance factor ($p < .05$).

Confidence: Confidence means that the content provided by the recommendation emails can enhance learners' confidence in their ability to develop their interest. Results in Phase 1 (M = 4.03, SD = .76) and Phase 2 (M = 4.30, SD = .65) were both positive, yet there is a significant difference between Phase 1 and Phase 2 ($p < .05$).

Satisfaction: Satisfaction refers to the user experience. The results showed that the difference between Phase 1 and Phase 2 is very significant ($p < .001$). 82% of subjects in Phase 2 strongly agreed that the system was easy to use (M = 4.12, SD = .68).

Summarization: Summarization refers to the system's ability to provide sufficient information for users to understand the main idea of the recommended video. The findings indicate that the summarization in Phase 2 (M = 4.10, SD = .66) was significantly superior to that of Phase 1 (M = 3.37, SD = .96). Again, there is a significant difference between the two phases ($p < .001$).

Recommendation: This factor is concerned with the accuracy of the recommendation. Because the same recommendation mechanism is employed in both phases of the experiment, it is not a surprise that there was no significant difference in the results.

Discussion

Multimedia content can support learning by helping learners to identify emergent goals within a context. This system aims to motivate learners and provide them with a useful tool for knowledge assimilation. As the results show, learners received the information they sought and assimilated knowledge unknowingly through watching the videos. Emails from Phase 2 proved to be more useful in helping learners assimilate knowledge than the emails from Phase

1, owing to the versatility of the VCSR system. The obvious differences between text-based email and multimedia-based email are the summary and the interactive video play. According to the analysis of the recorded data from the camera, subjects spent about one minute and thirty seconds reading the text-based emails in Phase 1, and they spent more than five minutes in Phase 2. The multimedia-based emails, therefore, motivated subjects more successfully. A large number of subjects reported favorably on the effectiveness of the summaries and video clips. The main reason they wanted to watch the video was because of the detailed summary and video clips. They could grasp the meaning of video through the abstract summary instead of watching the whole video, and they could then determine exactly which video to watch.

According to the results of the questionnaires, the system promotes more motivation with regard to the attention factor because the system actively sends content-rich emails to learners. When learners checked their mailboxes, they saw the video content of the recommendation email. If the video content was free, learners would read the summary and then watch the video content. On the other hand, if the video content was not free, learners could read the summary to get main idea of the video content. Furthermore, it was found that the relevance and confidence factors are related to both the accuracy of the recommendation and the abundance of summarization. Because subjects felt that the recommendation email was related to their past experience and the summary offered rich information, they had confidence in developing their interests. Subjects gave positive feedback in the two phases in relation to the recommendation mechanism, which results in a significant difference between the two phases with regard to summarization. Subjects expressed that the system increased video-watching frequency and extended their knowledge. Furthermore, they hoped to own the system in the future, which accounts for the large difference between phases in relation to the satisfaction factor.

In addition, results also showed that subjects with different backgrounds had different perceptions. Some subjects answered that they had not been interested in watching videos in the past, but after operating the system, they expressed a keen interest in watching videos. They were excited at the prospect of easily receiving video information and exploring a previously unknown area of knowledge. Some subjects highly praised the recommendation function and video clips of the system, saying that they really needed a tool to help them find new multimedia materials and confirm their video choices. In the past, they had spent a lot of time searching for multimedia data, only knowing the content of a video after watching it. By using the system, they spent more time watching relevant videos than searching and watching irrelevant media.

The original purpose of summarization is to design an objective and faithful summary that is not influenced by advertisements. The results of the experiment show that the system successfully achieved this goal. Subjects fully and quickly understood the content of the recommended videos. One subject, however, expressed that he would like to get some subjective comments about the video from others. He suggested that the system set up a forum to give feedback and form discussion groups with other system users. Although the system did not provide this function, it is important to note that this subject shifted from a passive role to an active one, wanting to participate in discussions about the videos, rather than simply using the system and accepting or rejecting recommendations. In addition, a few subjects mentioned that there were some irrelevant words in the summaries. This phenomenon results from OCR errors. Furthermore, some subjects said that they would prefer to get more information in the summary if the video contains rich educational content or knowledge they are interested in. However, if the genre of a video is commercial entertainment, the subjects expressed a lack of interest in receiving summary information, saying that more summary information would reduce their imagination as it relates to the story. In this experiment, of course, all of the materials are educational in nature, and the summaries give learners a clearer understanding of the content of the videos. Learners are also motivated to watch the entire video in an effort to gain more knowledge with respect to the topic of the summary.

Generally, the more information learners get, the more motivation is generated. If learners receive more detailed information, there is a better chance that their interest will be piqued. However, the fact that more recommendation information might inundate learners with redundant information is certainly a concern. Some subjects might treat the recommendation email as junk, thereby lowering their interest in receiving future recommendation emails from the system. On the contrary, most subjects expressed that they had dreaded searching videos in the past, and that the system considerably reduced the effort required to find the desired media. One subject expected to have a function that allowed him to save and manage recommendation emails. This indicates that the subject was pleased with the system. He wanted to keep the information and have his own accessible store of summaries and recommendations. In addition, some subjects suggested that the recommendation categories need to be diversified. For example, if someone likes animals, but especially insects, and the system recommends a video about dolphins, he will surely reject it. Moreover, subjects suggested that the system needs to analyze the learners' personality. If someone is curious about a certain topic, then the system should give him a specific part of the suitable video. It is important to note that the employed video data is limited and may not cover all subjects. Therefore, there is little likelihood that the system did not find a suitable video to match subjects' tastes in this experiment. This might be why the recommendation scores in both phases are so close: the system reinforced the subjects' knowledge, but it could not change an uninteresting item into an interesting one. Some subjects also pointed out that the recommendation accuracy of the system is very important. If most of the content in the recommendation email refers to learners' specific interests, learners will be pleasantly surprised and have higher a confidence level. On the contrary, if the system gives an inaccurate recommendation, learners would slowly lose faith in the accuracy of the system.

The VCSR system can extract a summary from numerous unstructured data (raw video), in turn constructing a recommendation. Meanwhile, the summary content was compared to learners' profiles in order to generate recommendation emails. One subject expressed that he preferred multimedia emails because the detailed summaries. He said he would feel no preference for either of the two email types if the multimedia-based email did not include a summary. Hence, the recommendation email is not fascinating for the user if it lacks a summary. Another subject said that he was not interested in *Discovery* films, so he did not pay particular attention to the summary. The recommendation mechanism must necessarily conform to learners' interests. If the recommendation mechanism is not robust, learners are less likely to read the summary. The above observations indicate that summarization and recommendation are intimately linked in the sense that the summary is important to the perceived validity of the recommendation.

Conclusion

In this paper, we presented an automatic multimedia content summarization and adaptable recommendation system, called VCSR, which is able to auto-recommend suitable video content for learners. The system can automatically create a video summary from raw video data and send it to learners based on their user profiles. The results of the experiment demonstrated that the system positively promoted learners' motivation in comparison with text-based recommendation method. Furthermore, the experiment suggests that there could be a relationship between summarization and recommendation in the sense that if there is no summary in the recommendation email, the recommendation email is likely not appealing enough to motivate learners. On the other hand, learners do not like to read the summary if the recommendation mechanism is not robust enough to fit their needs.

In the future, we plan to collect the personal information of more learners. The more the system understands learners' personalities, the better the recommendations it can create. Based on each learner profile, we also plan to customize the video summary for each learner. This implies that different learners can obtain different summaries even though the video source is the same. In addition, we are going to create a forum where learners can give their feedback and discuss issues relating to the VCSR system. Learners will play an active role rather than a passive role in the system. Moreover, it is necessary to improve the accuracy of the OCR module and the quality of the summaries by employing a more effective algorithm. Finally, we expect that the VCSR system could be applied to a real educational setting, such as a digital library, in the near future.

Acknowledgements

The authors would like to thank Mr. Chia-Hsing Shen for assisting in the experiment. The authors would also like to thank all the subjects who participated and cooperated in the experiment. This study was partially supported by a grant (NSC 97-2628-S-008-001-MY3) from the National Science Council of Taiwan.

References

Alspector, J., Kolcz, A. & Karunaithi, N. (1998). Comparing Feature-Based and Clique-Based User Models for Movie Selection. *Proceedings of the Third ACM Conference on Digital Libraries*, 11–18.

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

Basu, C., Hirsh, H. & Cohen, W. W. (1998). Recommendation as Classification: Using Social and Content-Based Information in Recommendation. *Proceedings of the Fifteenth National Conference on Artificial Intelligence*, 714–720.

Brown, J. C., Collins, A. & Duguid, P. (1989). Situated Cognition and the Culture of Learning. *Educational Researcher, 18*(1), 32–42.

Choi, H. J., & Johnson S. D. (2005). The Effect of Context-Based Video Instruction on Learning and Motivation in Online Courses. *The American Journal of Distance Education, 19*(4), 215–227.

Choi, H. J. & Johnson, S. D. (2007). The Effect of Problem-Based Video Instruction on Learner Satisfaction, Comprehension and Retention in College Courses. *British Journal of Educational Technology, 38*(5), 885–895.

Clark, J. M. & Paivio, A. (1991). Dual Coding Theory and Education. *Educational Psychology Review, 3*(3), 149–210.

Cotter, P. & Smyth, B. (2000). PTV: Intelligent Personalized TV Guides. *Proceedings of the Seventeenth National Conference on Artificial Intelligence and Twelfth Conference on Innovative Applications of Artificial Intelligence*, 957–964.

Dang, H. T. (2006). Overview of DUC 2006. *Document Understanding Conference*, 1–8.

Dang, H. T. (2007). Overview of DUC 2007. *Document Understanding Conference*, 1–10.

Dick, W., Carey, L. & Carey, J. O. (2001). *The systematic design of instruction* (5th Ed.). New York: Longman.

Fasquel, J. B., Agnus, V., Moreau, J., Soler, L. & Marescaux, J. (2006). An Interactive Medical Image Segmentation System Based on the Optimal Management of Regions of Interest Using Topological Medical Knowledge. *Computer Methods and Programs in Biomedicine, 82*(3), 216–230.

Huang, Y. T., Tsai, C. C., Chung, C. I., Shen, C. H., Yang, J. C. & Wu, Y. C. (2007). Multimedia Content Summarization and Application to Video Recommendation. *Proceedings of Technology Enhanced Learning Conference*, 23–30.

Keller, J. M. (1983). Motivational design of instruction. In C. M. Reigeluth (Ed.), *Instructional-Design Theories and Models: An Overview of Their Current Status* (pp. 383–433). New York: Lawrence Erlbaum Associates.

Keller, J. M. & Kopp, T. W. (1987). An Application of the ARCS Model of Motivational Design. In C. M. Reigeluth (Ed.), *Instructional Theories in Action: Lessons Illustrating Selected Theories and Models* (pp. 289–320). New York: Lawrence Erlbaum Associates.

Lee, Y. S., Wu, Y. C. & Chang, C. H. (2005). Integrating Web Information to Generate Chinese Video Summaries. *Proceedings of the Seventeenth International Conference on 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, 17*(1), 5–20.

Mackey, T. P. & Ho, J. (2008). Exploring the Relationships Between Web Usability and Students' Perceived Learning in Web-Based Multimedia (WBMM) Tutorials. *Computers & Education, 50*(1), 386–409.

Mayer, R. E. & Moreno, R. (2002). Animation as an Aid to Multimedia Learning. *Educational Psychology Review, 14*(1), 87–99.

Melville, P., Mooney R. J. & Nargarajan, R. (2002). Content-Boosted Collaborative Filtering for Improved Recommendations. *Proceedings of the Eighteenth National Conference on Artificial Intelligence*, 187–192.

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.

Mooney, R. J. & Roy, L. (2000). Content-Based Book Recommending Using Learning for Text Categorization. *Proceedings of the Fifth ACM conference on Digital Libraries*, 195–204.

Osadchy, M., & Keren, D. (2004). A Rejection-Based Method for Event Detection in Video. *IEEE Transactions on Circuits and Systems for Video Technology, 14*(4), 534–541.

Over, P., Ianeva T., Kraaijz W. & Smeaton, A. F. (2005). TRECVID 2005 – An Overview. *Proceedings of TRECVID 2005 Workshop*, 1–27.

Over, P., Ianeva T., Kraaijz W. & Smeaton, A. F. (2006). TRECVID 2006 – An Overview. *Proceedings of TRECVID 2006 Workshop*, 1–29.

Paivio, A. (1971). *Imagery and Verbal Processes*. New York: Holt, Rinehart and Winston.

Pass, F. G. W. C. & van Merrienboer, J. J. G. (1994). Variability of Worked Examples and Transfer of Geometrical Problem-Solving Skills: A Cognitive-Load Approach, *Journal of Educational Psychology, 86*(1), 122–133.

Piriou, G., Bouthemy, P. & Yao, J. F. (2006). Recognition of Dynamic Video Contents with Global Probabilistic Models of Visual Motion. *IEEE Transactions on Image Processing, 15*(11), 3418–3431.

Popescul, A., Ungar, L. H., Pennock, D. M. & Lawrence, S. (2001). Probabilistic Models for Unified Collaborative and Content-Based Recommendation in Sparse-Data Environments. *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence*, 437–444.

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.

Sweller, J. & Chandler, P. (1994). Why Some Material Is Difficult to Learn. *Cognition and Instruction, 12*(3), 185–233.

Sweller, J., van Merrienboer, J. J. G. & Paas, F. G. W. C. (1998). Cognitive Architecture and Instructional Design. *Educational Psychology Review, 10*(3), 251–296.

Tsai, C. C., Chung, C. I., Huang, Y. T., Shen, C. H., Wu, Y. C. & Yang, J. C. (2007). VCSR: Video Content Summarization for Recommendation. *Proceedings of the Seventh IEEE International Conference on Advanced Learning Technologies*, 862–864.

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

Wang, T. I., Tsai, K. H., Lee, M. C. & Chiu, T. K. (2007). Personalized Learning Objects Recommendation Based on the Semantic-Aware Discovery and Learner Preference Pattern. *Educational Technology & Society, 10*(3), 84–105.

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: Advances in Image and Video Technology, 4319*, 563–572.

Wu, Y. C., Yang, J. C. & Lee, Y. S. (2008). A Weighted String Pattern Matching-based Passage Ranking Algorithm for Video Question Answering. *Expert Systems with Applications, 36*(1), 1–32.