

Virtual Reference for Video Collections: System Infrastructure, User Interface and Pilot User Study

Xiangming Mu

School of Communication, Information Studies, University of Wisconsin-Milwaukee, Bolton Hall 510, 3210 N Maryland Ave, Milwaukee, WI 53211. mux@uwm.edu

Lili Luo

School of Information and Library Science, University of North Carolina at Chapel Hill, CB 3360, 100 Manning Hall, Chapel Hill, NC 27599-3360. luo@unc.edu

A new video-based Virtual Reference (VR) tool called VideoHelp was designed and developed to support video navigation escorting, a function that enables librarians to co-navigate a digital video with patrons in the web-based environment. A client/server infrastructure was adopted for the VideoHelp system and timestamps were used to achieve the video synchronization between the librarians and patrons. A pilot usability study of using VideoHelp prototype in video seeking was conducted and the preliminary results demonstrated that the system is easy to learn and use, and real-time assistance from virtual librarians in video navigation is desirable on a conditional basis.

Introduction

In the past several years, the Virtual Reference (VR) service software has been installed in many libraries to provide real-time chat reference service that allows patrons to “chat” directly with a librarian in a web-based environment. The central concept of the VR is

to use software and the Internet to facilitate human intermediation in information seeking at a distance (Lankes, 2004). The Internet is a bridge that connects patrons and librarians without the constraints of library hours or physical location. The VR software delivers real-time assistance to patrons on the web.

One advantage of the VR service over a phone or email reference service is the “escorting” process, namely, librarians can show patrons the information seeking process virtually and remotely. The information escorting is usually achieved by web page “pushing,” web co-browsing, and query co-inputting. With “escorting”-supported VR software, a librarian can “push” a specific web resource to remote patron’s web browser, share controls of the web browsing, and directly modify the patron’s query input. In this way, both the librarian and the patron go through the same information- seeking process, as if they were both physically in the reference section of the library. Such web browser-based escorting for information navigation and seeking is supported by most of the popular VR systems, including 24/7 reference (Keys & Wilson, 2004), Tutor.com, and OCLC QuestionPoint (O’Leary, 2003). However, the current VR tools do not support navigation escorting for multi-media resources, such as video, while the multi-media collection and the need for searching multi-media resources efficiently have been growing rapidly.

To understand issues involved in virtual reference service for video collection, a new prototype of the video-based Virtual Reference (VR) system, VideoHelp, was developed. One of the primary functions in the VideoHelp is to support video content- based “escorting”—online librarians can escort patrons to a specific shot or scene inside the video to facilitate their information seeking.

As a new media format, video provides rich, vivid, and unique visual/audio information that sometimes cannot be substituted by text-based information. The reduced cost of video digitalization, compression, and storage technology has stimulated the growth of digital video collections. Some dedicated video libraries are on the horizon: the Internet Archive Moving Images (<http://www.archive.org/movies>), the American Museum of the Moving Image (<http://www.movingimage.us/>), the International Federation of Film Archives (<http://www.fiafnet.org/uk>), the Motion Picture and Television Reading Room in the Library of Congress (<http://www.loc.gov/rr/mopic>) and the Open-Video Project (<http://www.open-video.org>) are some of these collections. Digital video has proliferated to such an extent (i.e., a collection of special topic in the Internet Archive only, “Prelinger Archives,” contains over 2,000 films) that the major search engines such as Google and Yahoo are planning to develop dedicated tools for digital video searching (Olsen, 2004).

Unlike conventional information search and retrieval, video seeking may require more content and visual detail-based exploration. For example, most search topics in the TRECVID 2004 (Kraaij, et al. 2004) required users for visual content searching, such as “Find shots of a street scene with multiple pedestrians in motion and multiple vehicles in motion somewhere in the shot” (topic 125), “Find shots of a person hitting a golf ball that then goes into the hole” (topic 136), and “Find shots of one or more buildings on fire, with flames and smoke visible” (topic 147). As a result, simply using a video title and text description may not effectively answer these questions, as shown in the results of the TRECVID studies (Smeaton, 2004). For example, for the 25 topics in the TRECVID 2004, the average precision [the accuracy of answers or information] was less than 0.1 (Smeaton, 2004). In this sense, it is reasonable to assume that when patrons access a web-based video collection and look for shots on a specific subject, video-content-based help from the librarians would be desirable, which means, librarians familiar with the collection will assist patrons in retrieving relevant video clips and locating relevant shots embedded in them. This service can be accomplished in the next generation of VR: the new VR system that supports video navigation escorting.

The paper is presented in the following sequence: after a literature review, the system requirements, system infrastructure, and user interface of the VideoHelp will be illustrated. Next, the initial results of a pilot study of applying the VideoHelp prototype in an experimental environment will be presented and discussed. Finally a list some of future research plans on VideoHelp will be provided.

Literature Review

The present development of network and multimedia technology increased needs for digital video applications in education, commercial, and digital library. Compared to text document, video retrieval is a more complex task. Visual information and audio information are different from the traditional text media. Traditional text-based retrieval technology can not be directly applied to retrieval visual content based on color, textual, or visual objects. In addition, a growing number of researchers have realized that the importance of content-based video retrieval. As a result, a specific TREC for video retrieval (TRECVID) has been formed to explore issues involved in content-based video retrieval (<http://www-nlpir.nist.gov/projects/trecvid>). As all the current online catalogue services provided by library can only support text-based search for video (i.e., key words, title), helps from librarians are desired for visual content-based information needs such as “I am looking for a shot that contains at least three walking people, a running car

and flooding water”.

To provide content-based video reference service and information-seeking “escorting”, specific software is needed. Several commercial VR software are currently available, for example, the QuestionPoint (<http://www.questionpoint.org>), 24/7 Reference (<http://www.247ref.org>), Docutek VRLplus (<http://www.docutek.com>), eLibrarian ([eLibrarian.digi-net.com](http://www.digi-net.com)), LivePerson Pro (<http://www.liveperson.com>), and Virtual Reference Toolkit (<http://www.vrtoolkit.net>). Olivvares (2004) offered a review and comparison of these systems in 21 categories, covering cost, administration, primary functions, and application scope of current VR systems. All of these systems support real-time chatting, web page pushing, web co-browsing, and web navigation escorting. However, none of them support video content browsing/navigation escorting functions. We extended our literature review to the video collaboration area and assessed some video collaboration systems, including JETS, TELEP, Helpmate, DVCC, and ISEE.

JETS (Shirmohammadi et al. 1998) is a Java-Enabled TeleCollaboration System developed at the Multimedia Communications Research Laboratory (MCRLab), University of Ottawa. JETS is designed for real-time sharing of Java applets. Using any Java-enabled Web browser, multiple users in a telecollaboration session are able to share generic applications in the form of Java applets. The current supported applications include shared whiteboard and shared 3D objects. The whiteboard supports real-time sharing of images, interactive PowerPoint presentations, on-line drawings and H.263 Video. However, the restrictions on the video format make it not an appropriate platform for video navigation escorting.

TELEP, developed by Microsoft for remote education, added dialog chat functions to the one-way streaming video (Jancke et al. 2000). Remote users are able to send questions to the presenter directly, and other users can also join the discussion by replying to the question or issuing comments. However, real-time multicast technology used in the TELEP is designed for a group of clients rather than the point to point service provided in the VR service.

Helpmate (Curran 2002) is an online collaboration environment that packages a group of communication tools, including chat room, instant messaging, discussion forums, and shared whiteboard. One unique feature of Helpmate is that a real-time translation package called Systran (<http://www.systransoft.com>) for multi-language translation is integrated and it supports users to communicate in different languages. The author did not offer any accuracy data in the real-time translation applications. In addition, in Helpmate video is primarily used as a communication tool (videoconference is supported in Helpmate) rather than as content and

resources to be shared.

DVCC is a collaborative video viewing system developed by Microsoft (Cadiz, et al. 2000) that supports video content sharing. Microsoft windows media player was integrated with Microsoft Netmeeting to allow distributed users for both real-time chatting and shared video viewing. An individual user is able to use the local VCR functions to control the shared video. The video co-navigation is achieved by the sharing of the video streaming. The drawback, however, is that the streaming video may cause transmission delay, which in-turn restricts the interactions between the video and users (Cadiz, et al. 2000).

ISEE is an Interactive Shared Educational Environment specifically designed for low network bandwidth users (Mu et al. 2003). Pre-downloaded videos are used to support instant interactions between the video player and users. A mechanism of Smartlink is utilized in ISEE to achieve video synchronization across distributed users. Video annotation and group discussion are also supported (Mu, 2004).

These systems, however, are not designed either for a VR service or for video navigation escorting. In addition, the video streaming and video multicasting technologies are adopted by most of these systems (except ISEE) and thus limit the support for instant interactions between users and the video players, which is critical for quality video sharing and video navigation escorting. As a result, a new VR system is needed.

VideoHelp System

The primary goal of the VideoHelp VR system is to provide support for video navigation escorting between librarians and online patrons. Towards this goal, we believe the following functions are important:

- Video synchronization: both the virtual patron and virtual librarian need to be able to share the same video at the same time.
- Video navigation: the virtual patron needs to be able to control and navigate the video playback.
- Video control: the virtual librarian needs to be able to take control of the virtual patron's video playback in order to demonstrate video navigation.

Based on these requirements, a prototype video-based VR system called VideoHelp was developed. Java was chosen as the development language to achieve interoperability across common platforms. Java Shared Data Toolkits (JSDT) Application Program Interfaces (APIs) were utilized for information sharing and group communication functions. Java Media Frame (JMF) APIs were used for developing the video-related components, including the video player and the video storyboards.

A client/server infrastructure was adopted. On the server side, a Session Manager, implemented using Java socket, was running as a daemon to coordinate communications and video sharing. When a virtual service was activated, both the librarian and the patron were registered by the Session Manager in the same communication session (see Figure 1).

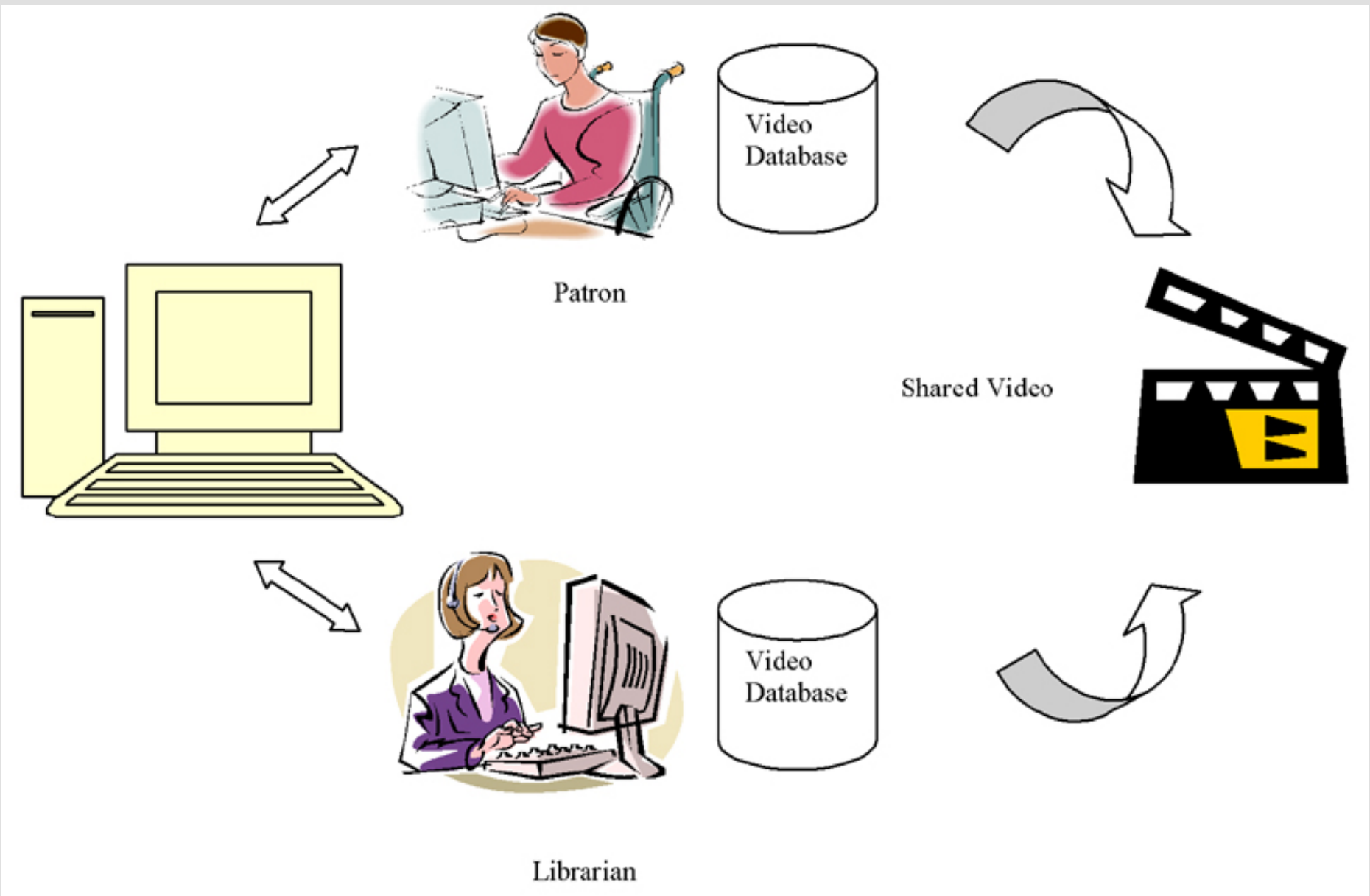


Figure 1: System infrastructure of the VideoHelp tool

On the client side, a VideoHelp client java application was running for both the librarian and the patron. The same video database was connected to each copy of the VideoHelp client. The following two reasons drove us to choose such an infrastructure over streaming video:

- Local video supports instant response for navigation controls. The streaming video will introduce transmission delay when it is manipulated (i.e., pausing/resuming or jumping).
- The storage cost is no longer a critical consideration in the design of the VideoHelp system. For example, a 250G hard drive can store up to 500 hours MPEG-1 videos and is affordable (now less than \$400).

Video sharing and synchronization were achieved by using the video timestamps, which were similar to those in the ISEE system (Mu, 2004). A manipulation of the video such as video pause or a video jump issued an action in the system, which notified the Session Manager. In addition, the related video timestamp of the manipulation was sent to the Session Manager, which, in turn, was delivered to the virtual patron. After receiving the action message, the patron's VideoHelp system automatically updated the local video playback to the new location based on the librarian's video timestamp. Similarly, if a patron issued a video navigation action, the same process was triggered and the librarian's video playback was updated. In this way, real-time video synchronization was achieved.

VideoHelp User Interface

The interface of the VideoHelp was adapted from the Interactive Shared Educational Environment (ISEE) system (Mu et al. 2003). There were three major components: a shared video player (component A), a shared web browser (component B), and a real-time chat box (component c) (see Figure 2).



Chat Panel

LIBRARIAN: Welcome to the VideoHelp ! I am a librarian. What can I do for you today?

PATRON 01: Hi, I am glad to meet you here

PATRON 01: I am looking for some video shots here

PATRON 01: particularly about earthquake in Memphis

PATRON 01: for example, the history of earthquake in this city and possibility in the future

LIBRARIAN: I see. if you put "Memphis earthquake" as the key words to search in your video database

LIBRARIAN: you will see a result like this

LIBRARIAN: Have you got the page?

PATRON 01 : Yes, I see the result

LIBRARIAN: How about we take a look at the "tolden fury" first

PATRON 01: sure

PATRON 01: You can see, this video talked about "Madrid Seismic Zone"

PATRON 01: which do include Memphis

LIBRARIAN: Yea. and it compared this zone with the great San Francisco earthquake in 1906

LIBRARIAN: Let me take you to the section that predicts the future for memphis earthquake

PATRON 01: sure

LIBRARIAN: Look at here,

PATRON 01: yea. I think this video is exactly what I want

PATRON 01: thank you

shared browser

http://www.uwm.edu/~muc/video/help/demo/pilots/ud/

Result for your query-- memphis earthquake

Hidden Fury (1993)	Documentary
Earthquake: Risk to the Central U.S (1983)	Documentary
Computer Animation of Loma Prieta Aftershocks (1989)	Documentary
San Francisco Earthquake Aftermath	Documentary
Your Chance to Live: Earthquake (1977)	Documentary

Send Clear Mark

You are welcome.

B

C

Figure 2: A snapshot of the VideoHelp user interface

- **Shared Video Player:** This is the key component of the VideoHelp system. Java Media Framework APIs were used to build the video player. To encourage interaction with the video, the librarian and the patron only viewed a local copy of the same video (with the same video ID), rather than a streaming video. Timestamp was used for video synchronization. For example, in Figure 2, after the librarian said “let me take you to the section ...” he pulled the video playback to a new position by using the control slider which was located just beneath the video player. An action message with the timestamp (5’33”, or 333”) was then sent to the Session Manager. XML was used to encode this metadata information. A part of a sample message is given below.

```
<video>
<id>34 </id>
<timestamp> 333 </timestamp>
</video>
<sender> Librarian </sender>
<receiver> Patron </receiver>
```

After receiving the message, the patron’s VideoHelp decoded it according to the same encoding protocol. The timestamp (333 second) extracted was used to update the video player. In this way, the librarian was able to “escort” the patron’s video content navigation.

- **Shared web browser:** It provided supports for web resource sharing and functions that were similar to the conventional VR systems such as the 24/7 Reference system. Java EditPane APIs were used to implement this component.
- **Real-time chat** was a two-way communication chat room that enabled text chatting.

A Pilot User Study

To improve our knowledge of video escorting in the VR service and to evaluate the performance of the VideoHelp prototype, a pilot user study was conducted to explore the following two questions:

- If the video escorting function is currently available, will the patron use it to aid his/her video information seeking?
- How satisfied are users with finding relevant video content via the VideoHelp system?

The study was conducted in the Student Union of the University of Wisconsin at Milwaukee. A statement describing the study was posted in front of the Union bookstore, asking for volunteers to evaluate a new virtual library software program. Twelve volunteers joined the study. Among them, seven were female and the other five were male. Eleven were students and one was a staff member. Free snacks and drinks were provided for participants during the study.

The Session Manager server was running in one of the researchers' offices throughout the study. Two laptop computers installed with Windows XP and the VideoHelp client system were placed on two adjacent desks (around two meters apart) in front of the Union bookstore for the study: one for the virtual librarian and one for the virtual patron. One researcher logged on as the virtual librarian, and the volunteer logged on as a virtual patron. The VideoHelp system had been set up in advance so the volunteers did not have to login each time. The campus wireless network was used to connect the librarian and the patron to the Session Manager server. A sample video database containing 1,200 video clips obtained from the Open-Video Project (<http://www.open-video.org>) was installed in each of the client computers for local video access.

The volunteers were asked to play the role of online virtual patron looking for a video clip related to "Memphis earthquake" and a shot containing "Memphis" as text on the screen. After a short demonstration of how to use the VideoHelp system, the volunteer and the researcher were seated in front of their respective computers at different desks. The volunteer was asked to imagine that he/she was using the VideoHelp system to "remotely" chat with an online librarian. Headphones were provided for both the researcher and the volunteers. It took the volunteers an average of 20 minutes to finish a study.

Immediately after the task was completed, volunteers were interviewed to answer some questions (see Appendix) about their experience with the VideoHelp system. The questions were designed on the basis of the Questionnaire for User Interaction Satisfaction (QUIS) (Chin, et al., 1988) version 7.0.

Results and Discussions

From the responses (see Table 1) received from the 12 volunteers the VideoHelp system for video content seeking, we found that VideoHelp was interesting to use (6.67 on a scale of 9) and not difficult to learn (6.33 out of 9). Negative feedback was also reported. One participant mentioned that she was frustrated because the system crashed and had to be restarted during the study (we are still looking for the reasons for the system crash).

Table 1: Descriptive Statistics for Using VideoHelp system (Number of users: 12, Scale Range: 1-9)

	Mean	Standard Deviation
Terrible/Wonderful	5.83	1.99
Frustrating/Satisfying	4.58	2.19
Dull/Stimulating	6.67	1.30
Difficult/Easy	6.33	2.10
Inadequate/Adequate power	5.17	1.99
Rigid/Flexible	6.33	1.67

In addition, all 12 users indicated that they had never used such a service before and would like to use it if it became available. Some users mentioned that they would not use it each time for video content seeking. Four users stated that the usage would depend on the video searching tasks. If they were familiar with the content of the video and knew where to find it, they would not spend time talking with the virtual librarian. Two users claimed that the virtual librarian's knowledge of the video content would affect their decision of whether or not to continue using the system.

Participants also offered suggestions for the improvement of the system. One user asked if he could use the shared browser (provided by the VideoHelp, see Component B in Figure 1) to search the Internet for more videos. Our current infrastructure of the VideoHelp system did not support this function, but we plan to incorporate it into our future design.

Another direction for the future work is to explore the users' real needs for help in their video seeking and retrieval tasks. Previous studies have demonstrated that users were frustrated when using the OPAC system (Christine, 1986; Christine, 1996) and the Internet (Jansen, et al. 2001) in information seeking due to the high zero hit rates, inappropriate interface design, confusion on Boolean logic,

and incapability to use advanced features. But we still experienced low usage after the VR services were widely available (Coffman & Arret, 2004; Boyer, 2001; Tenopir, 2004). Content-based video searching is more complex than generic information seeking. Further studies are needed to improve knowledge of users' needs in their video content seeking.

Conclusion

The primary contribution of this paper is that we have designed and developed a novel video-based VR tool -- the VideoHelp system. The prototype of VideoHelp allows a librarian to escort a remote patron's video content navigating in a web-based environment. The video escorting function is achieved through the sharing of video timestamps between the librarian and the patron. One advantage of the VideoHelp is availability of instant interactions between the users and the video player.

The pilot user study demonstrated that the VideoHelp was interesting and easy to learn. All the participants mentioned that they would use to try such a service if it becomes available in the future. However, to fully evaluate the performance of a video-based VR system, a more comprehensive user study is needed. We will continue to work on a new version of the VideoHelp and plan to conduct a larger scale and more rigorous user study in the near future. The new system will continue to improve the usability. No in-person training will be needed. In addition, the "virtual" librarian will stay in a different room from the online users.

Acknowledgements

Thanks to Open-Video project (<http://www.open-video.org>) for providing video clips for the pilot user study.

References

Boyer , J. (2001). Virtual Reference at the NCSU Libraries: The First One Hundred Days. *Information Technology and Libraries*, 20(3), 122-128.

Cadiz, J.J., Balachandran, A., Sanocki, E., Gupta, A., Grudin, J., & Jancke, G. (2000). Distance learning through distributed collaborative video viewing. In *Proceedings of the 2000 ACM conference on computer supported cooperative work* (pp. 135-144). New York: ACM Press.

Chin, J.P., Diehl, V.A., & Norman, K.L. (1988). Development of an instrument measuring user satisfaction of the human-computer interface. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 213-218). New York: ACM Press.

Borgman, C.L. (1986) Why are online catalogs still hard to use? Lessons Learned from Information-Retrieval Studies. *Journal of the American Society for Information Science*, 37(6), 387-400.

Borgman, C.L. (1996) Why are online catalogs still hard to use? *Journal of the American Society for Information Science*, 47(7), 493-503.

Coffman, S., & Arret, L. (2004). To Chat Or Not to Chat -- Taking Another Look at Virtual Reference, Part I. *Searcher*, 12(7), 38-46.

Curran, K. (2002). An Online Collaboration Environment. *Education and Information Technologies*, 7(1), 41-53.

Jansen, B. J., & Pooch, U. (2001) Web user studies: A review and framework for future work. *Journal of the American Society of Information Science and Technology*, 52(3), 235 - 246.

Jancke, G., Grudin, J., & Gupta, A. (2000). Presenting to local and remote audiences: design and use of the TELEP system. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 384-391). New York: ACM Press.

Keys, J., & Wilson, F. (2004). AskNow! - Evaluating an Australian Collaborative Chat Reference Service: A Project Manager's Perspective. *Australian Academic & Research Libraries*, 35(2), 81-94.

Kraaij, W., Smeaton, A.F., & Over, P., (2004). TRECVID2004-An Introduction. In *TREC video retrieval evaluation online*

proceedings. Retrieved Jan. 10, 2005 from <http://www-nlpir.nist.gov/projects/tvpubs/tvpapers04/tv4intro.pdf>

Lankers, R. D. (2004). The digital reference research agenda. *Journal of the American Society for Information Science and Technology*, 55(4), 301 - 311.

Mu, X., Marchionini, G., and Pattee, A., (2003). The Interactive Shared Educational Environment: User interface, system architecture and field study. In *Proceedings of the third ACM/IEEE-CS joint conference on Digital libraries* (pp. 291-300). New York: ACM Press.

Mu, X. (2004). *Smartlinks in a video-based collaborative distance learning system: a cognitive model and evaluation study*. (Doctoral dissertation, University of North Carolina at Chapel Hill)

O'Leary, M. (2003). QuestionPoint Fortifies Libraries in Internet Age. *Online*, 27(3), 70-72.

Olivares, O. (2004). Helping You Buy: Virtual Reference Systems. *Computers in Libraries*, 24(5), 25-29.

Olsen, S. (2004). *Striking up digital video search*. Retrieved Jan. 10, 2005, from http://news.com.com/Striking+up+digital+video+search/2100-1032_3-5466491.html

Shirmohammadi, S., Oliveira, J.C. & Georganas, N.D. (1998). Java-based Multimedia Collaboration: Approaches and Issues. In *Proceeding of International Conference On Telecommunications. Vol. I* (pp.127-131). Porto Carras: Greece.

Smeaton, A., & Over, P. (2004). TRECVID-2004: Search Task Overview. In *TREC video retrieval evaluation online proceedings*. Retrieved Jan. 10, 2005, from <http://www-nlpir.nist.gov/projects/tvpubs/tv.pubs.org.html>

Tenopir, C. (2004). Rethinking Virtual Reference. *Library Journal*, 129(18), 34.

Appendix: Post-test questionnaire

	Mean	Standard Deviation
Terrible/Wonderful	5.83	1.99
Frustrating/Satisfying	4.58	2.19
Dull/Stimulating	6.67	1.30
Difficult/Easy	6.33	2.10
Inadequate/Adequate power	5.17	1.99
Rigid/Flexible	6.33	1.67