Task-dependent video relevance criteria – A proposed naturalistic study

Meng Yang
School of Information and Library Science
University of North Carolina at Chapel Hill
Advisor: Gary Marchionini

1. Problem statement

Relevance is one of the fundamental and central concepts in information science. Relevance-based measures – recall and precision are the two most commonly used criteria to evaluate the effectiveness of information retrieval systems. In addition, relevance is also a necessary part of understanding human information behavior (Schamber, 1994, p. 36), since people's information seeking process is also an iterative relevance judgment process.

Despite its importance in information science, relevance has also been a controversial concept since its first appearance. The earlier system-oriented definitions of relevance consider it as a relationship between users' queries and documents, and have aroused many criticisms such as "inappropriate", "inadequate" and "ambiguous" (e.g., Doyle, 1963 and Cooper, 1973). Later on, research on relevance focused on "What users think relevance means", or a user-oriented definition of relevance, which considers relevance as a relationship between users' needs and the documents. Since the 1990s, many naturalistic relevance studies have elicited criteria beyond these topicality and user-defined criteria. However, among the large body of user relevance studies, very few of them focus on how people search videos and make video selections. This is an important gap considering the rapid development of video digitization, compression and retrieval techniques (especially the content-based video retrieval ones).

1.1. Research questions

The study described in this proposal will explore video relevance. The main question to be answered in this study is: *How do people make relevance judgments when they search videos?* There are also some sub-questions as follows:

- What criteria do people use to make relevance judgments when searching videos?

- What visual criteria do they apply, if any?
- How should these relevance criteria be categorized?
- What factors affect users' video relevance judgments?
- What are the relationships between video surrogates/metadata and those criteria?
- How will different video searching tasks influence people's relevance judgments?

1.2. Significance of the study

Research on users' video relevance judgment processes will not only enrich the video information seeking literature and the general relevance literature as well, but also provide insights into the design of effective video retrieval systems to serve users' information needs. This study will add to our understandings of multi-dimensional nature of relevance as a concept (e.g., topical, cognitive, dynamic and situational) in the video searching domain. Additionally, the relevance criteria from this study will also provide information about users' video information needs and seeking behavior, and thus inform video indexing practice. Furthermore, as Shatford (1994) states, relevance criteria suggested from empirical user studies can be regarded as possible access points to images and videos, and consequently inform metadata selection and design. The criteria could also help video retrieval researchers to test the applicability of the current content-based retrieval techniques that they have developed (e.g., color, shape, texture, specific objects and events) and also give them more clues for future developments.

2. Related literature

Since its first appearance, relevance has been one of the most confusing and debated concepts in information science, in spite of its importance in the field. Intuitively, people understand what relevance means and often use it in their daily information retrieval activities, yet researchers cannot reach a consensus on a scientific definition of relevance. Normally speaking, there are two types of definitions for relevance: *system-oriented* relevance or *user-oriented* relevance. The system-oriented definition focuses on the relations between a specified search request and the retrieved documents, whereas the user-oriented definition concentrates on the relations between

users' information needs and the retrieved documents. Many researchers have expressed this dichotomous view of relevance concept: Vickery's (1959a, 1959b) *relevance to a subject* and *user relevance*; Schutz's (1970) *topical relevance* vs. *interpretational relevance*; and most notably, Swanson's (1986) *objective relevance* and *subjective relevance*.

Since the 1990s, there has been an increase of studies regarding relevance judgment processes by real users, with real tasks and applying various kinds of naturalistic methods such as interviews (e.g., Schamber 1991, Park 1992), case studies (Tang & Solomon 1998) and quasi-experiments (Barry, 1994). Many of the criteria found in those studies are similar to each other. Considering the differences in research methodologies and environments applied in different studies, it can be said that user relevance criteria studies strongly support the notion that "there is a finite range of relevance criteria that is shared across users and situations" (Barry 1994, p.157). Based on those empirical studies, it is apparent that criteria related to information content (e.g., topicality) are the ones most frequently mentioned by users. For instance, topicality accounted for 65.3% in Wang's (1994) study, information content of the document 35.1% in Barry's (1994) study, and topical relatedness 68% in Tang & Solomon's (1998) Study.

However, among those various relevance criteria studies, very few of them looked at how users search video retrieval systems. Several studies on image relevance judgments (Markkula & Sormumen 1998, Choi & Rasmussen's 2002) could have some implications, since images share many similar characteristics with videos. The relevance criteria about image searching gathered from these studies have some similarities with those about textual information searching; for instance, topicality is still the most important and often the first criteria people use. However, there are also some criteria that are unique to image searches, such as affectiveness and technical quality of the image. As for videos, the relevance criteria could be even more complicated, since videos contain both spatial and temporal information. The study proposed here would do some exploratory work toward this direction.

3. Methodology

Qualitative data collection methods will be applied in this study, since currently there is a lack of existing literature on video relevance research and the collection of qualitative data is more suitable for exploration and understanding purposes than experimental methods. Qualitative data collection plus grounded theory data analysis approaches can be used either to create or develop new models or theories in the video relevance research field, or perhaps to further develop existing ones. Additionally, naturalistic methods will be applied in this study, since users' relevance criteria are situation-dependent (Schamber 1990), and they can only satisfactorily be identified when users are engaged in their own tasks. As Wilson (1994) indicates in his review paper of information needs and uses, "Our concern is with uncovering the facts of everyday life of the people being investigated" (p.72).

In the relevance literature, many researchers have applied interviews (e.g., Park 1993, Schamber 2000), and quasi-experimental methods using think-aloud protocols (e.g., Barry 1994, Tang & Solomon 1998). Both types of method have been proved to be viable ways to elicit user-defined relevance criteria; however, each method alone also has its own deficiencies. Retrospective interviews can provide a more general view of participants' information needs, tasks and their searching strategies, but asking participants to recall what they were thinking during a prior experience might also give inconsistent and incomplete data due to users' memory inaccuracy. These "inconsistent" and "incomplete" data can be complemented by asking participants to do real searches while thinking aloud, which can give concurrent and more accurate information on the participants' cognitive processes. The triangulation of these two methods can provide a more complete description of the participants' relevance judgment processes.

Therefore, this study will be broken down into two parts: a semi-structured interview and a real video search with the think-aloud method. The goal is to elicit users' video relevance criteria, especially their audiovisual relevance criteria, since those are unique to video searches.

3.1. A preliminary study

A pilot study (Yang & Marchionini, 2004) was conducted in early 2003 and a semi-structured interview was applied in the study due to its exploratory and descriptive nature. Four participants were interviewed, including a professor in communication studies, an art professor, a news video librarian and a video editor. Using the grounded theory method, three categories of relevance judgment criteria (Figure 1) were identified: textual criteria (e.g., topicality, recency, and nationality), visual criteria (e.g., cinematography, objects/events, and style) and implicit criteria (e.g., interest, familiarity, and appropriateness). Some preliminary findings include: relevance criteria were very task-dependent; topicality was still considered the most important criterion for video relevance judgments, however, users also wanted to see more visual surrogates, especially those surrogates that contained motion (see Yang & Marchionini, 2004 for more details). This pilot study has provided a viable interview schedule to be improved and used in the dissertation research, and the preliminary coding scheme will also be tested and further developed.

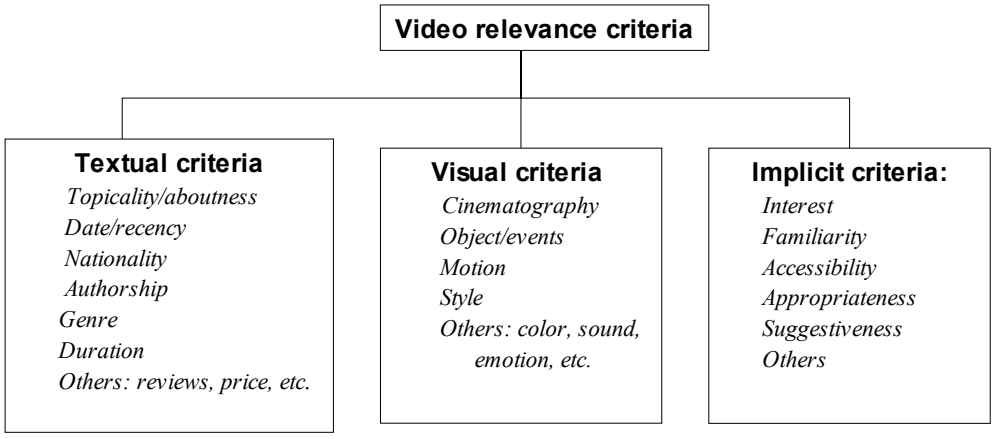

Figure 1. A summary of video relevance criteria from the pilot study

3.2. Participant selection

As suggested in the literature, people's relevance criteria could depend on various factors, such as their specific situations or tasks in hand, their professions, their cognitive and psychological features, their background knowledge and experiences, and video characteristics such as genres and quality. Some of these factors actually interrelate with each other. For

instance, their professions might affect what types of videos they choose and what types of tasks they would accomplish. From the earlier pilot study it was found that situational factors, especially the tasks people want to accomplish, would influence the criteria they apply for relevance judgments. For instance, appropriateness was an important criterion for the two professors who wanted to find some video clips to illustrate class topics, but the video editors cared more about the topicality of video scenes.

In this proposed study, the participant selection will be based on the different tasks. Derived from the previous pilot study, three types of tasks will be considered: *illustration, production* and *collection*. The *illustration* tasks happen when teachers or professors want to find some videos to illustrate their class topics. The participants will be chosen among professors from local universities. They can be from departments such as communication studies, American studies, and Asian studies. The *production* tasks happen when video editors want to find clips to produce stories. The participants will be chosen from some video production companies and TV stations, such as UNC-TV. The *collection* tasks happen when librarians, especially media specialists, want to find videos to enrich their collections, sometimes based upon teachers' or professors' requests, and other times based on their collection development policies. Approximately 10 participants will be chosen for each of these three types of task. Other factors such as video genres and user characteristics will not be considered during the participant selection process, but will be analyzed and compared during the data analysis stage.

3.3. Data collection

3.3.1. Semi-structured interview session

In the first part of the study, a semi-structured interview method will be applied to ask the participants to describe one of their job-related video search situations in sequential order (see Appendix A for the interview protocol). The participants will be asked to describe their specific information needs, the information sources selected, video searching questions, results selection

process, and their final video uses. The interview protocol has been revised based on the investigator's pilot study and will continue to be refined during the future interview process.

The interview session will be semi-structured and the investigator will constantly check whether all the questions on the protocol have been answered. The interviews will last about 30-40 minutes and will be recorded on audiotape.

3.3.2. Video search session with think-aloud

In the second part of the study, the participants will be asked to do a real video search while thinking aloud their thoughts, using the task they described in the first interview session. The participants will first be asked to go to a website they use regularly to search videos. If there is no such online resource available, the investigator will recommend several online video resources to the participants, based on the types of videos they are looking for. Then they will be asked to do a real video search, based on the tasks they discussed in the interview session. At the same time, they will also be asked to think aloud during their searching and browsing processes.

During the search session, the investigator will remind the participant to think aloud if this does not happen and might also ask some further questions about her/his video selections. Since visual relevance criteria are one focus of this study, the participants might be guided to view some visual surrogates if they fail to do so. The search session will last about 20-30 minutes and will be recorded on audiotape.

3.4. Data analysis

Both the interview and video search sessions will be audiotaped and transcribed. The preliminary coding schema (textual, visual and implicit criteria; see Figure 1) developed from the pilot study will be tested and further developed in this formal study. Grounded theory will again be used to code those categories that are not covered by the preliminary coding scheme. The coding scheme development will observe the five general steps suggested by Weber (1985): determine the recording unit (word, phrase, sentence), develop categories, code data samples, test for inter-coder agreement and revise and retest. As suggested in the grounded theory literature

(Strauss & Cobin, 1990), the data collection and data analysis processes are deliberately fused, and the initial data analysis is used to shape the future data collection. Therefore, the investigator will try to transcribe and code the content right after each interview in order to revise and improve the interview questionnaire. Additionally, the investigator will also send a brief summary to the participant after the transcription, in order to check the accuracy and completeness of the investigator's interpretation. Finally, additional coders will code a sample of the data set. The inter-coder reliability will be examined to verify and refine the coding scheme. The scheme will be discussed and revised until there is an 80% coding agreement between the investigator and the two other coders (Robson, 2002).

Once the coding process is finished, relevance criteria applied by the participants will be compared between the three different user tasks (illustration, production and collection). It is expected that the results will indicate how different tasks affect people's video relevance judgment criteria. The criteria will also be compared across other factors, such as video genres, and user characteristics, if possible. The relationship between metadata/surrogates and the corresponding relevance criteria will also be considered. Finally, the user-elicited video relevance criteria will be compared and contrasted to those identified in the relevance literature, which were mostly developed from people's searches on textual documents.

3.5. Validity and reliability of the study

The naturalistic and qualitative methods applied in this study will increase the validity of the results, since the participants will be asked to describe their own search experiences. Triangulation of the semi-structured interviewing method and think-aloud protocols during searching will also compensate for each other's deficiencies, providing more accurate and complete information about the participants' video relevance judgment processes, and thus improving the validity of this study. After each interview, brief summary will be sent to the participant to check the accuracy of the transcription, thus further enhancing the validity of the data. Additionally, revising the interview questionnaires and coding scheme through a pilot study

and during the proposed study, and testing inter-coder agreements can also improve the reliability of these data collection and data analysis processes.

3.6. Limitations of the study

Like many other relevance studies, the generalizability of the results from this study cannot be assured, since the grounded theory approach focuses on theorizing and applicability. The results might only be applied to those types of tasks which the three user groups have done, and the criteria applied by people accomplishing other types of tasks might be quite different. However, since this study will explore a new area which has not been studied before, it is hoped that more similar studies on other user groups will further test the generalizability of the results to other situations and tasks.

3.7. Schedule of completion

	Data collection	**Data transcription/coding**	**Data analysis**	**Paper writing**
April, 2004				
May, 2004				
June, 2004				
July, 2004				
August, 2004				
September, 2004				
October, 2004				
November, 2004				
December, 2004				
January, 2005				
February, 2005				
March, 2005				

3.8. Budget justification

The budget for this project is mainly devoted to data collection. Funds are requested for audiotape recording equipment, travel to Atlanta or Boston to interview some producers from big TV stations or video production companies, interview transcription, statistical consulting.

The author also receives a research assistantship by working for the Open Video Project, which is funded by NSF IIS 0099638.

Reference:

Barry C. (1994) User-defined relevance criteria: An exploratory study**,** *Journal of the American Society for Information Science*, *45 (3)*, pp.149-159.
Choi, Y. & Rasmussen**,** E. M. (2002). Users' relevance criteria in image retrieval in American history**,** *Information Processing & Management,* 38(5), pp. 695-726.
Cooper, W.S. (1973). On selecting a measure of retrieval effectiveness. *Journal of the American Society for Information Science*. *24(2),* pp. 87-100.
Doyle, L.B. (1963), Is relevance an adequate criterion for retrieval system evaluation? *Automation and Scientific Communication*, *Part 2*, p. 199-200. Washington, American Documentation Institute.
Markkula, M. and Sormumen, E. (1998), ``Searching for photos: journalists' practices in pictorial IR'', *The Challenge of ImageRetrieval Research Workshop*, Newcastle upon Tyne
Park, T. K. (1993). The nature of relevance in information retrieval: An empirical study. *Library Quarterly, 63(3),* pp. 318–351.
Schamber, L., Eisenberg, M. B., & Nilan, M. S. (1990). A re-examination of relevance: Toward a dynamic, situational definition*. Information Processing & Management*, *26(6),* pp. 755–776.
Schamber, L. (1991). Users' criteria for evaluation in a multimedia information environment. *Proceedings of theAmerican Society for Information Science (AISI) 54th Annual Meething,* Voume 28, pp. 126-133.
Schamber, L. (1994). Relevance and information behavior. *Annual Review of Information Science and Technology, 29, pp.* 3–48.
Shatford, L. S. (1994). Some issues in the indexing of images. *Journal of the American Society for Information Science, 45(8)*, pp. 583-588.
Strauss, A. & Corbin, J. (1990). *Basics of Qualitative Research: Grounded Theory Procedures and Techniques.* Newbury Park, CA: Sage Publications
Robson, C. (2002). *Real World Research: A Resource for Social Scientists and Practitioner-Researchers.* Second Edition. Blackwell Publishers
Schutz, A. (1970), *Reflections on the Problem of Relevance*. New Haven, Yale University Press
Swanson, D. R. (1986). Subjective vs. objective relevance in bibliographic retrieval systems. *Library Quarterly,* 56, pp. 389-398.
Tang, R.& Solomon, P. (1998). Toward an Understanding of the Dynamics of Relevance Judgment: An Analysis of One Person's Search Behavior, *Information Processing & Management, 34*, pp. 237-256
Vickery, B.C. (1959a). The structure of information retrieval systems, *Proceedings of the International Conference on Scientific Information,* Vol. 2, National Academy of Sciences, Washington, DC, pp. 1275-1290.
Vickery, B.C. (1959b). Subject analysis for information retrieval, *Proceedings of the International Conference on ScientificInformation,* Vol. 2, National Academy of Sciences, Washington, DC, pp. 855-865.
Wang, P. (1994) A cognitive model of document selection of real users of information retrieval systems (Doctoral dissertation), University of Maryland, College of Library and Information Science, College Park, MD.
Weber, P.W. (1985). Basic content analysis. Beverly Hills, CA: Sage

Wilson, T. (1994) Information Needs and Uses: Fifty Years of Progress? In B.C. Vickery *(Eds.), Fifty Years of Information Progress.* London: Aslib.
Yang, M. & Marchionini, G. (2004, to appear). Exploring Users' Video Relevance Criteria ---- A Pilot Study. *Proceeding of the Annual Conference of American Society for Information Science and Technology*. Providence, Rhode Island. November 12-17, 2004.

Appendix A: Interview Questions

1. Demographic information (profession, status, frequencies of video search.)
   - Profession:
   - Frequencies of video search
2. Information need analysis:
   - Please think about one recent job-related situation in which you needed to find some videos (not including the situation if you knew exactly what you wanted.)
   - Could you please describe in detail why you needed the video information? What kind of video information did you want?
3. Information source analysis
   - Where did you go to search for the video information? Why? (online or physical library?)
4. Video search questions
   - How did you begin your search? (e.g., the questions you asked the librarian, the queries you input into the system, the actions you took?)
   - Can you think of other search methods? (e.g., visual search methods?) Will they work or not? Why?
5. Browsing surrogates/relevance judgment questions
   - What kind of video information did the system provide? What did you look for first? Why?
   - What kind of video information (e.g., title, author, date) seemed most important to you? Why?
   - Did the system have some pictures or video segments to preview? If so, did you think they helped? Why? If not, what kind of visual information did you expect to have? Would you like to have? Why?
   - If both textual and visual information was available, which information was more important to you, visual or textual information? Why?
   - Was it difficult for you find what you wanted? Why?
   - Do you think that the system provided enough information for you to search? If not, what other kinds of information did you want to have? Why?
6. After watching full videos
   - After you watched/downloaded the videos, were you satisfied with the results? If not, what disappointed you? What information did you expect to have in the surrogates/metadata?
7. Video use
   - How did you use the video?
8. Can you think of another situation that you search videos? Any difference?
9. In general, when you search videos, what criteria/metadata/surrogate do you think are most important?

Meng Yang

School of Information and Library Science
University of North Carolina at Chapel Hill
303 Manning Hall, CB#3360
Chapel Hill / NC / USA / 27599-3360
TEL: 919.966.3589 / FAX: 919.962.8071
Email: yangm@ils.unc.edu

Education:

- *Ph.D. student: Sep. 2000 - present:*
  School of Information and Library Science at University of North Carolina at Chapel Hill
  Advisor: Gary Marchionini

- *Bachelor of Engineering: July 1998.*
  Major: Management of Information Systems
  School of Management at University of Aeronautics and Astronautics, Beijing, P. R. China.

Research projects involved:

- *Open Video Project* (*www.open-video.org*)
  Involved in designing and conducting usability tests.

- *TRECVID 2003 workshop participation* (*see the demo*)
  System and website design (php/mysql)
  Involved into designing and conducting usability tests.
  See the related publication

- *VIVO*
  System design (java)
  See the related publication

Teaching experience:

- Summer 2004:
  INLS 110-125 -------Video indexing (self-designed course)

Publications:

2004

Yang, M. & Marchionini, G. (2004, to appear). Exploring Users' Video Relevance Criteria ---- A Pilot Study. *Proceeding of the Annual Conference of American Society*

for Information Science and Technology. Providence, Rhode Island. November 12-17, 2004

2003

Yang, M., Wildemuth, B., Marchionini, G., Wilkens, T., Geisler, G., Hughes, A., & Gruss, R. (2003). Measures of User Performance in Video Retrieval Research. *Proceedings of the Annual Meeting of the American Society for Information Science* (Long Beach, LA, Oct. 19-22, 2003), pp. 3-11

Wildemuth, B. M., **Yang, M.**, Hughes, A., Gruss, R., Geisler, G., and Marchionini, G. (2003). "Access via Features versus Access via Transcripts: User Performance and Satisfaction. *Proceedings of TRECVID 2003 workshop notebook*. pp. 189-198

Yang, M., Marchionini, G., Wildemuth, B. & Mu. X. (2003) The Open Video Project: Building an open source digital video archive. *DigiCULT newsletter. Issue 5*, pp22-25. Available online http://www.digicult.info/downloads/digicultinfo_issue5_october_2003.pdf

Yang, M., Mu, X., & Marchionini (2003) (poster), VIVO - A Video Indexing and Visualization Organizer, *Proceedings of the Third ACM+IEEE Joint Conference on Digital Libraries*,(Houston, Texas, June27-31, 2003), pp. 403

Wildemuth, B. M., Marchionini, G., **Yang, M.**, Geisler, G., Wilkens, T., Hughes, A., and Gruss, R. (2003). "How Fast Is Too Fast? Evaluating Fast Forward Surrogates for Digital Video." *Proceedings of the 3rd ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL 2003)*, pp. 221-230. (**Vannevar Bush Award Winner for Best Paper at JCDL 2003**)

2002:

Wildemuth, B., Marchionini, G., Wilkens, T., **Yang, M.**, Geisler, G., Fowler, B., Hughes, A. & Mu, X. (2002). "Alternative Surrogates for Video Objects in a Digital Library: Users' Perspectives on Their Relative Usability." *Proceedings of the 6th European Conference on Digital Libraries,* September 16 - 18, 2002, Rome, Italy

Geisler, G., Marchionini, G., Wildemuth, B. M., Hughes, A., **Yang, M.**, Wilkens, T., and Spinks, R. (2002). "Video Browsing Interfaces for the Open Video Project." *Proceedings of CHI 2002*, Extended Abstracts

2001

Geisler, G., Marchionini, G., Nelson, M.l, Spinks, R. and **Yang, M**. (2001). "Interface Concepts for the Open Video Project." *Proceedings of the Annual Conference of the American Society for Information Science*. DC,

Nelson, M., Marchionini, G., Geisler, G., and **Yang, M.** (2001). "A Bucket Architecture for the Open Video Project [short paper]." *ACM - IEEE Joint Conference on Digital Libraries* (June 24-28, 2001, Roanoke, Virginia).