STUDY ON REQUIREMENT SPECIFICATIONS FOR PERSONALIZED MULTIMEDIA SUMMARIZATION

Lalitha Agnihotri[1], Nevenka Dimitrova[1], John Kender[2], John Zimmerman[3]
{lalitha.agnihotri, nevenka.dimitrova}@philips.com, jrk@cs.columbia.edu, johnz@cs.cmu.edu

[1]Philips Research USA
345 Scarborough Road
Briarcliff Manor, NY 10510

[2]Dept. of Computer Science
Columbia University
New York, NY 10027

[3]HCI Institute
Carnegie Mellon University
Pittsburgh, PA 15213

ABSTRACT

The ability to summarize and abstract information will be an essential part of intelligent behavior in consumer devices. However, even for the emerging area of video content analysis, user preferences on summarization have not been explored. This paper reports on a panel which asked: who, why and when summarization is needed; what information should be summarized; and what forms should summaries take. In particular, we investigated the requirements with sensitivity to user needs, user context, media content, device capabilities, and the methods by which user and environment profiles can be assembled and exploited. We organize our findings as a wish list containing four major user questions, and we report the current and near-term states of the art and the major technical challenges in satisfying them. Our study suggests that user preferences should be derived from explicit user statements and from implicit trends inferred from viewing histories and summary usage.

1. INTRODUCTION

Access to information continues to increase while human cognitive abilities remain the same. Therefore, it is necessary to provide technologies that select and summarize information in a personalized but content-preserving way. Such summarizations can serve as friendly indices into the fuller content, as trained sentinels watching for unusual but desired events, or as gentle reminders of content that has migrated into media that are more difficult to browse.

Automatic video summarization attempts to shorten video or create another representation such as images, collages, etc. Video summarization has been an active area of research in the last several years. Most of it is specialized by genre, with some emphasis on news [8,10], although a few researchers have investigated the issues more generically [1,4,7,9,11,12]. Summarization appears to be active among researchers working on speech [2] and printed text [5]; in particular, there have been several workshops hosted by the Association for Computational Linguistics on scalable text summarization.

With all this extensive research effort in automatic summary extraction there has been little attempt to explore the user aspects and analyze their requirements. To approach the user requirements problem, we organized a panel, which gathered together five researchers with extensive experience in video presentation, indexing, recommender systems, and user studies [1,6] in order to explore user needs. Over four separate brainstorming sessions the panel examined four questions: (i) Who needs summarization and why? (ii) What should summaries contain? (iii) How should summaries be personalized? And (iv) how can results be validated? The resulting ideas on what was essential (and, as yet, essentially missing) in summarization systems were then structured by using "affinity diagrams" [3]. Participants first generated ideas of Post-It notes that were placed on the wall. Next, they clustered and linked the ideas to reveal the underlying themes and relationships. We analyzed the responses and synthesized and generalized the emerging patterns. This paper reports the resulting consensus.

Following the organization of the workshop, sections of this paper relate to our specific questions: Section 2 addresses who and why; Section 3 explores summary elements; Section 4 covers personalized summaries; and Section 5 looks at validation. Section 6 of the paper concludes with a summary table of the current and near-term state of the art, as well as (many) key issues and obstacles that remain to be researched.

2. WHO NEEDS SUMMARIZATION & WHY

2.1 Who needs summarization?

In this session, we elaborated on potential users of multimedia summarization. Some popular responses to the "who" question were: TV viewers in general; people without the time or patience to fast forward; executives; researcher exploring conference presentations; tourist shopping for vacation spots; mothers searching baby care; homeowners exploring "how-to" videos; sports fans reviewing highlights; people exploring cooking video (example, search for cake icing); police examining street activity; home video editing. Other users included TV producers looking for story patterns; anthropologists studying trends; and architects exploring how people move through space.

A pattern emerged that summarization is needed by people who either do not have the time or do not want to invest time in searching extensively for information. The two main categories that appear are professionals and home users. The two groups have different demands on the system and need different kinds of summaries.

2.2 What content needs to be summarized?

The different contents and genres that can and should be summarized varied across the board. The participants suggested contents from web search results to baby care TV shows. A summarizer is needed for the home, supermarket, and traffic surveillance videos. Lecture tapes and presentations have to be summarized for quick digest or for searching for specific information. Wedding and home videos need to be summarized to enable editing. Different TV genres that need summarization include news programs, sports, how-to, travel, cooking, music videos, talk shows, etc.

The main categories are produced content such as TV programs and film and live capture content such as home videos, lectures, and surveillance. Also there is amount of detectable structure. There are many types of TV genres that came up. Figure

I shows level of editing vs. detectable structure in different types of programs. Informative programs (e.g. news) have both high levels of detectable structure and lots of editing. Narrative content like action, comedic, or dramatic films and TV shows have much less detectable structure but are highly edited.

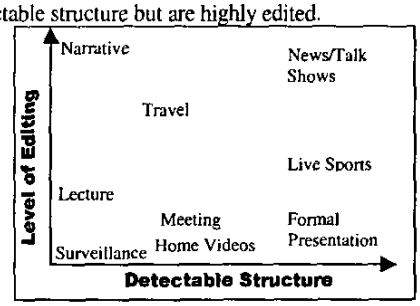

Figure 1. Level of Editing vs. Detectable Structure

2.3 When is summarization needed and used?

Participants identified several contexts when summarization is needed: browsing, searching, editing, and organizing. When browsing for video to watch, summaries (previews) can help identify both content of interest as well as previously viewed content. These summaries can help users plan and organize their viewing. When searching for specific information, summaries can both help identify appropriate content and provide access points into the content. When editing content, summaries can help users both search, plan, and organize content. Recommenders and personalized filters work with all of these contexts. Figure 2 shows the different types of summary users in a known intent ("know what I want") vs. available time graph along with different contexts. Figure 3 shows the different types of users who use the summaries along the same axes.

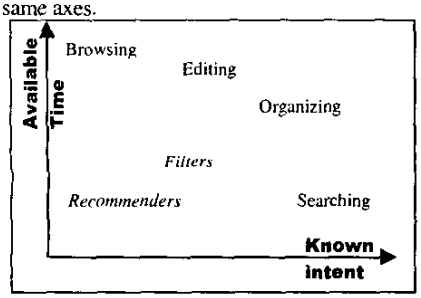

Figure 2. Modes of summary usage

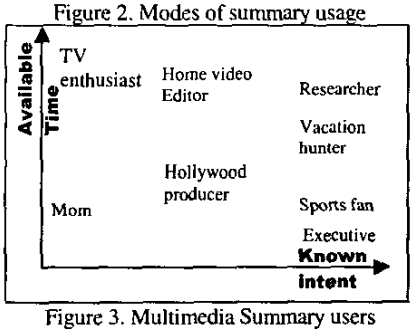

Figure 3. Multimedia Summary users

2.4 Where will the summarization be viewed?

Participants identified the need for summaries in a variety of places: home, office, on the move, and on various storage devices including TVs, PCs, PDAs and phones. Summaries would primarily be watched in "spare" time while commuting (car/train), or while spending time between scheduled events (e.g. waiting room). Full programs would preferably be watched at home. Summaries could be browsed at anyplace magazines are currently read. Summaries should be on a network and accessible everywhere.

2.5 How often would the summaries be viewed and used?

With respect to frequency of access, participants' responses ranged from: every time I access via a mobile device to some preset number of defaults based on time. Other responses included: in the morning, every evening, and everyday. The bottom line was that summaries would be viewed at some regular interval based on an individual's routine. For commuters, traffic and weather summaries would be useful in the morning and evening. Talk show, instructional videos, and news programs would be accessed at leisure time.

3. CONTENT IN A SUMMARY

3.1 Elements of a summary

Summaries contain two types of elements: *media elements* such as audio visual, textual items and *content elements* such as people, locations, topics, goals, etc. Orthogonal to the media and content elements, participants indicated that summaries should be hierarchical and contain multiple levels. The summary hierarchy should provide a "meta program summary": e.g. 21 stories in a news program or 4 goals in a soccer game. At the next level, there should be a detailed description of the program's content: e.g. keyframe and headline for each story. The level after that should include a scene summary and farther down still a shot summary.

3.2 Is the summary genre-dependent?

The answer to this question was a unanimous YES. For each genre different specific elements are important and the summary should capture this aspect. This is related particularly to the media and content elements.

Participants enumerated *content elements* based on genre. For example, summaries of sports should include content elements such as events, participants, and final scores. Summaries for award programs should include a list of nominees, winners, and presenters. Talk show summaries should contain guests and topics. News summaries should contain the topic, the participants (who), the event and outcome (what), the time of the event (when), and the location (where) for each story. Movie summaries should include information on characters and plot points and should not reveal fully the plot as to diminish the enjoyment of watching.

The *media elements* present for different genres are also different. For sports, video summaries should have more visual and audio information. For financial programs, a text overview is essential. Music video summaries could display a list of songs along with audio clips. Instructional video summaries should include visuals on the specified topic. This suggests that a full taxonomy of genres and their media components is needed and would be a valuable contribution to the area.

4. SUMMARY PERSONALIZATION

4.1 Is the summary user context dependent?

According to participants, the summary is not only user dependent but also dependent on time, task, and environment. It is user dependent as particular segments of movies are important to different people. The summaries are also user context dependent based on the task. For example, news summaries that are needed when watching is a secondary task (e.g. getting dressed and using the bathroom) are bound to be different than news summaries they might use when watching is a primary task (e.g. on a train). In the first context, users cannot devote their attention 100% to the task.

In the second context, users can devote their eyes, but may not want audio. The display (playback) of the summary for the same content should be quite different.

Summarization systems should preferably monitor and adapt to user's usage: time and location as well as the display device and content genre. By matching these elements to formats selected by the user, the system should be able to accurately predict which summary format users most likely want.

The responses in this session suggest that personalized summaries based on user's preferences that take user context into account will be highly desirable.

4.2 Personal Profile

According to participants, the personal profile consists of an *implicit* and an *explicit* section. The implicit section is derived from the user behavior and should include the following elements: previous search criteria; summary access patterns and past usage of summary; shows watched and their genre; list of topics consumed; and user's watching habits.

Users provide the information for the explicit section, and it should contain two parts: personal information and general program-related information specific to the user. The personal information should include: gender; age; home and work location; profession; time allocated for watching television including user's schedule, busy times and vacation schedule; hobbies; visual or auditory preference; abstract or concrete preference.

Additionally there should be general program related information specific to the explicit profile. This should include: explicit ratings of topics; preferred summary based on genre and task; preference of media for summary presentation; favorite and disliked: celebrities, politicians, stocks, topics, places, sports teams/types, music; target devices; tolerance for violence, sex, and other possibly objectionable content.

Our panel showed that for summarization purpose we need an extensive personal profile. Certain fields are required for filtering of videos and others are required for personalization of summaries. A computational model to match this personal profile with the video content needs to be researched.

5. VALIDATION

The final session's primary focus was summary validation. How do we validate the automatically generated summaries and their usage? What type of user testing will facilitate this process? Validation and user testing is the toughest part of the summarization process.

The participants suggested a number of processes for validation. Two important aspects came up: *user aspects* and *content element aspect*. For the user aspect validation, the usage of summaries should be compiled and measured against how useful the summaries were to users. Also the informative value, level of personalization, time fitting, task fitting, and enjoyment should be evaluated. For the content elements validation, we can enumerate important items according to genre and then validate if these elements are present in the summary by using recall and precision. Talk shows summaries should include a list of guests. Narrative program summaries should list actors and story threads.

For performing user tests we could use the talk aloud method: we have a person read the summary first and then present them with actual footage and ask if it is what they expected. Did users expect something to be summarized and it was not? We should ask user's expectations prior-to and after watching summaries.

6. SYNTHESIS & CONCLUSION

We synthesize the requirements and the key obstacles in each of the areas explored in our workshop in Table 1. In Table 2, we present a comparison of each of the four sessions versus the state of the art and speculate on the near term possibilities.

The result from first session "who/what/when/where/how-often" yielded people from varied walks of life need different genres summarized based on their context. The second session concluded that different genres need different kinds of summaries. Also, it was uncovered that we need to differentiate between *media elements* vs. *content specific elements* summarization. The personalization session concluded unanimously that summary is user context dependent. The personal profile should be comprehensive and include *implicitly* derived elements from user behavior and *explicitly* provided personal information. The validation session revealed that both *content requirements* and *user requirements* should be validated.

This panel suggests that the more important research advances will result from tackling the "more human" issues related to personalization and validation. In particular, it is imperative to develop a methodology for evaluating multimedia summaries.

In the literature the methods have covered the content aspects of summarization for generic summaries. Our panel concluded that it is important for summarization to be personalized to cover user aspects: preferences, context, time, location, and task.

References:

[1] L. Agnihotri, K. Devara, T. McGee, N. Dimitrova, "Summarization of Video Programs Based on Closed Captioning," SPIE 2001, January, San Jose, California

[2] B. Arons. "SpeechSkimmer: A system for interactively skimming recorded speech," ACM Trans. on Computer-Human Interaction, Vol. 4, No. 1, March 1997, pp 3-38

[3] H. Beyer and K. Holtzblatt," Contextual Design: Defining Customer-Centered Systems," Morgan and Kaufmann Publishers, Inc. San Francisco, California, 1998

[4] M.G. Christel, M.A. Smith, C.R. Taylor, D.B. Winkler, "Evolving Video Skims into Useful Multimedia Abstractions," In proc. of CHI, 1998, Los Angeles, pp 171-178

[5] T. Firmin and M. J. Chrzanowski, "An Evaluation of Automatic Text Summarization Systems," In Advances in Automatic Text Summarization, 1999,pp 391-401

[6] K. Kurapati, S. Gutta, D. Schaffer, J. Martino, J. Zimmerman, "A Multi-Agent TV Recommender", workshop on Personalization in Future TV, July 2001, Germany.

[7] H. Lee, A. Smeaton, P. McCann, N. Murphy, N.E. O'Connor, S. Marlow, "Fishlar on a PDA: A Handheld user interface for video indexing, browsing and playback system," CNR-IROE, Florence.

[8] B. Merialdo, K. Tak Lee, D. Luparello, J. Roudaire, Automatic Construction of Personalized TV News Program, ACM Multimedia, 1999, pp 323-331.

[9] Y-F. Ma, L. Lu, H-J. Zhang, M. Li, A User Attention Model for Video Summarization, ACM Multimedia Dec 1-5, 2002.

[10] A. Merlino, M. Maybury, "An Empirical Study of the Optimal Presentation of Multimedia Summaries of Broadcast News," In Advances in Automatic Text Summarization, 1999, pp 391-401

[11] H. Sundaram, L. Xie, S-F. Chang, A Utility Framework for the Automatic Generation of Audio-Visual Skims, ACM Multimedia 2002, Juan Les Pin, France, Dec 1-5, 2002.

[12] S. Uchihashi, J. Foote, A. Girgensohn and J. Boreczky, "Video Manga: generating semantically meaningful video summaries," ACM Multimedia, pp 383-392, Nov. 1999.

Session Questions	What is needed	Key Obstacles
Who needs summarization?		
When will summaries be used?	Several modes of usage identified: quick, refresh, recommender, browsing, editing, personalized filters	Consumer devices (both mobile and stationary storage) are unable to support complex audio-video analysis required to perform summarization.
What content needs to be summarized?	Surveillance, home videos, TV programs: news, cooking, music, talk shows, sports, narrative.	There is no available method for all genre summarization. Taxonomy is required.
Who needs it?	Business people, tourists, sport fans, news buffs, police looking at street activity.	Currently not available because summary of video content needs semantic understanding.
Where will it be viewed/used?	Multimedia server, on PDAs, mobile phones.	Summaries are not available yet.
How often?	Regularly.	Doesn't exist.
What information do summaries require?		
What elements should be present in a summary?	Include different media and content elements based on genre and user context.	Difficult to find important elements in all genres of television programs.
Should summaries be different for different genres?	Genre-sensitive summarizer.	Same as above.
User Dependent?	Personalized summaries	Difficult to map user profiles to user needs.
Personal profile		
What are the elements of a personal profile?	Profile to include list of topics, watching habits, viewing history, location, time, personal information.	Mapping of various attributes to the content not straightforward. There may be some intermediate psychological construct.
Validation?		
How to validate and setup user testing?	Validate if summary satisfies user and content requirements. Validate by elements and important items for elements in each genre. User study.	Not trivial to estimate what users need. No established methodology yet.

Table 1. The different requirements and the key obstacles in each of the areas

Session Questions	State of the art	Near State of Art
Who needs summarization?		
When are summaries used today?	Select results of web search; Electronic Program Guide (EPG) one-line summaries	Push technology from service provider
What content needs to be summarized?	Lectures (VideoManga), talkshows (SPIE 2001) Home videos, narratives (ACM MM 2002)	Sports, Movies
Who needs summaries?	Web surfers, TV viewers (TV Guide/TiVo/Satellite/ Cable)	Digital broadcast
Where are summaries currently viewed?	TVs, laptops	Cell phone, PDAs, web pads
How often summaries needed?	Daily	As often as the summary is updated
What information should be contained in a summary?		
What elements are present in a summary?	Images (keyframe), text, video, audio only and also some with combination of these	Multimedia summaries with all video, audio, and text elements.
Summaries different for different genres?	Sports highlights, lecture summary, Talk show summary, Handcrafted summaries of EPG	More individual genres summarized. Different summaries for different genres in EPG.
Summaries user dependent?	Current summaries are user independent	Still user independent
Personal profile		
Elements	Topics, Viewing history, Story board	Active inference from usage
Validation?		
How to validate?	Informedia, Movie Skims	Topic specific validation methodology

Table 2 State of the art and the near state of the art