Browsing Large Video Collections on Personal Video Recorders: Remote Control Navigation Keys do not Scale

Frode Eika Sandnes
Faculty of Engineering
Oslo Univ. College
N-0130 Oslo, Norway
frodes@hio.no

Yo-Ping Huang
Dept. Electrical Eng,
Natl. Taipei Univ. of Tech.
Taipei 106, Taiwan,
yphuang@ntut.edu.tw

Yueh-Min Huang
Eng. Science Dept., Natl.
Cheng Kung University
Tainan 701, Taiwan.
huang@mail.ncku.edu.tw

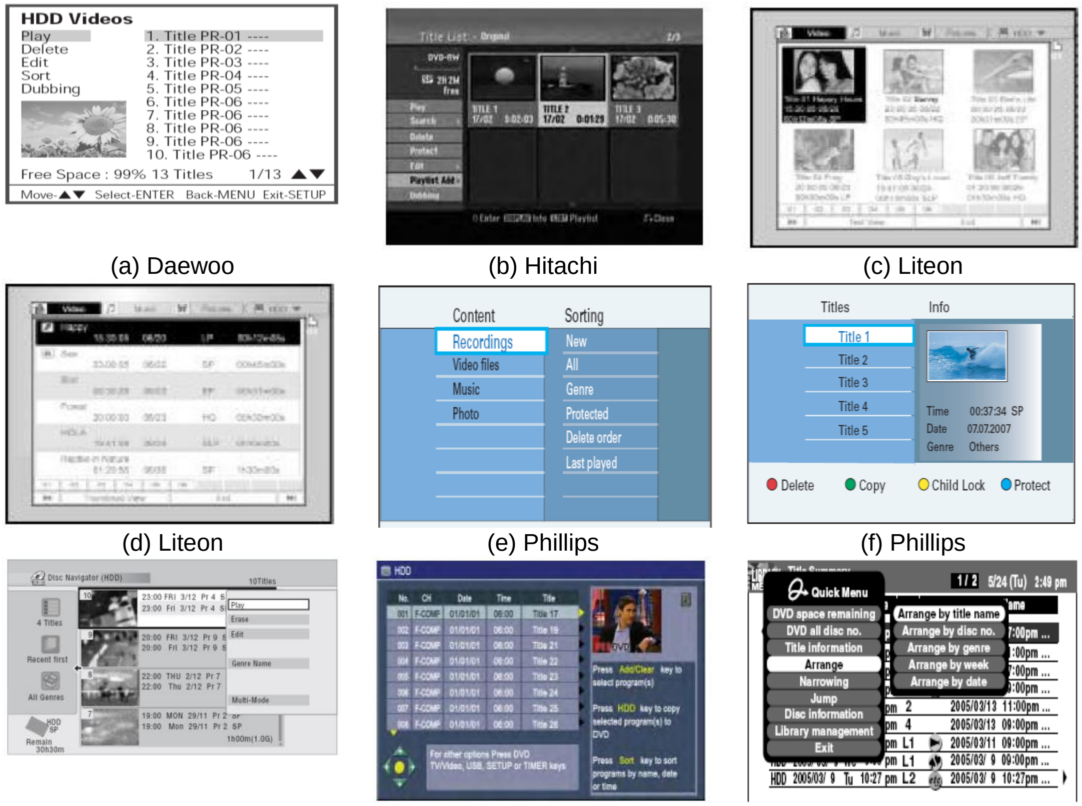
Abstract

Hard disk-based personal video recorders have the capacity to store large quantities of video contents. A personal video recorder serving one household may have to cater for the diverse entertainment tastes imposed by different family members. Timer event programmed video recordings have a tendency to accumulate and it is difficult to access large video collections on current generation personal video recorders. This is because the recordings often are presented linearly in a chronologically ordered list that is accessed using the navigation keys on a remote control. In this study an alternative strategy for accessing large video collections via the navigation keys of a remote control is investigated. The strategy automatically clusters the video recording using cuts that are perceptually meaningful to the viewers and that minimize the retrieval effort. However, the results suggest that navigation keys do not meet the needs of large video collection browsing.

1. Introduction

Most HDD video recorders present stored recordings in a chronologically ordered list which is navigated by scrolling up or down using a remote control (see Figure 1). This list can be time-consuming and difficult to navigate if it grows large. With repeated programs scheduled such lists will continue to grow if they are not continuously managed. Another problem is that it is time-consuming and difficult to locate a particular recording, especially if it has not been given a particular identifiable text string. Then the user must remember approximately when the recording was made or sequentially look through the recordings and guess the contents of each recording based on the time-stamp, duration and index image.

Increasing low cost storage means that multimedia collections are growing. Unlike text documents which are easily searched using keywords, it is harder to search through multimedia contents. Multimedia retrieval strategies can be divided into three classes, namely meta-information based [1], contents based and mixed [2]. Multimedia information which is tagged with meta information, i.e., information about the contents can be used for search and retrieval purposes. Meta information is usually manually added textual information that describes the contents. Content based multimedia retrieval involves finding content without meta information. For instance, in content based image retrieval low level image features are used to search through the image database [3]. Automatic content based video browsing has also been proposed [4-6].



(a) Daewoo (b) Hitachi (c) Liteon

(d) Liteon (e) Phillips (f) Phillips

(g) Pioneer (h) Polaroid (i) Toshiba

Figure 1. Screenshots of title selections from HDD-based personal video recorders. The screenshots are taken from the instruction manuals of Daewoo, Hitachi, Liteon, Phillips, Pioneer, Polaroid and Toshiba respectively (see Table 1). Screenshots a, d, g, h, and i show one-dimensional lists, screenshots b and c show two-dimensional lists that display 3 x 2 titles per screen. Screenshots e and I show models that allow contents to be sorted according to different attributes.

An alternative direction in multimedia retrieval is content browsing. There has been a significant amount of research into image browsing [7] and especially browsing on mobile terminals [8]. Most mobile phones have cameras and pictures tend to accumulate. Efficient browsing mechanisms are needed so that desired images can easily be found. It is often unrealistic to rely on manual meta information tagging. Browsing therefore has to be based on the available information such as date of creation [9], size and geospatial information [10]. Research has found that users prefer to browse digital photo libraries and rarely use advanced content based data retrieval features [11]. The strategy presented herein falls into this category. Another related problem is the exploration of music collections [12], and especially music collections on small terminals [13].

Another challenge is that unlike desktop computers which are controlled using sophisticated and accurate control devices personal video recorders are usually controlled using low cost IR-remote controls due to their low cost and convention [14-16]. Users usually have to control the device using four navigation keys and a select key. Users therefore have to perform many steps in order to reach their goals. These steps will be analyzed in the subsequent sections.

Table 1. Video browsing characteristics on common HDD personal video recorders.

#	manufacturer	models	Navigation	Preview	Library
1	BenQ	DE350	1D	Picture	
2	DAEWOO	DHR-8105	1D	Picture	
3	Eltax		1D	Picture	
4	Hitachi	DV-DS253A	2D	Picture	
5	JVC	DR-MH20S, DR-MH30S	2D	Picture	Category marking
6	LG	RH177 RH188	2D	Picture	
7	LiteON	LVW-5045	1D, 2D	Picture	
8	Panasonic	DMR-EZ47V	2D	Picture	
9	Phillips	HDMI 1080i	1D	Picture	Sorting on attribute
10	Pioneer	DVR-550H-S	1D	Picture	Category marking
11	Polaroid	DRA-101601A	1D	Picture	
12	Sony	SVR-S500	1D	Storyboard	
13	Thomson	DTH 8740E, DTH 8750E	1D	Picture	
14	Toshiba	RD-XS34SU, RD-XS34SC	1D		Sorting on attribute
15	Yamaha	DVR-9300HX	1D		

2. Analysis

2.1. Current practice and conventions

To establish an understanding of current practice, 15 HDD DVD recorders from leading manufacturers were surveyed and the results are shown in Table 1. The study is based on the information available in owners' manuals downloaded from the manufacturers' websites. This selection is a representative sample of the personal video recorders on the market. The support websites confirmed that it is common for manufacturers to adopt similar user interfaces for large parts of their product lines.

Most HDD video recorders surveyed provided a screenshot of the first image in the recording which can provide a tremendously useful visual clue to the viewers, but may still not be sufficient if the recording has started early during a commercial break as the image will be of the commercial and not the intended program. However, this problem was elegantly solved by model 12 in Table I which provides a thumbnail sequence for each recording at 5 minute intervals throughout the video. The user then gets an instantaneous visual impression of the entire storyline. Model 5 allows users to change the index image. However, active effort from the user is required.

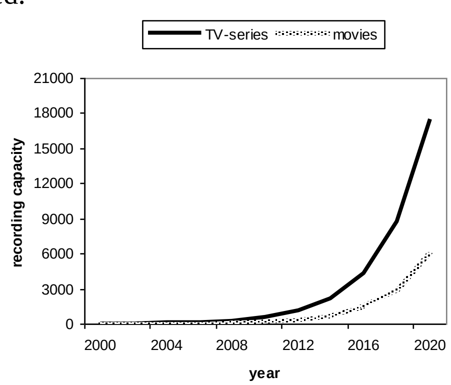


Figure 2. Personal video recorder storage capacity doubles approximately every two years. The estimate is based on standard play, i.e., 2 hours of video per 4.7 GB (DVD).

Model 5 in Table I was the only device with explicit video library management. In addition to supporting textual labeling of recordings the device also allows users to choose a category from a list of preset topics including movies, music drama, animation, etc. However, the problem with this strategy is that it requires manual intervention. Furthermore, preset categories are not likely to match the viewing behavior of all users.

Some of the models provide two-dimensional scrolling through the video collection which greatly speeds up access. However, two dimensional scrolling does not solve the problem of locating recordings in large collections.

2.2. Expected video collection growth

Figure 2 shows the rapid growth in personal video recorder storage capacity. Current HDD-based personal video recorders with 320GB HDDs can store approximately 272 30-minute TV-series or 45 90-minute movies in standard quality, i.e., approximately two hours of video per DVD-R (4.7GB). With a doubling in HDD capacity every two years, we can expect to find that by 2020 an off-the-shelves personal video recorder can store more than 17000 TV-series and nearly 6000 movies. The emerging need for efficient video management and retrieval systems is obvious.

2.3. Linear search

Imagine that a personal video recorder stores N video recordings. If these are accessed using a chronologically ordered list then the minimum number of steps needed to reach a recording is 1 and the maximum is N if the list does not wrap around. The mean number of steps s needed is

$$\bar{s} = \left\lceil \frac{N}{2} \right\rceil \tag{1}$$

assuming a uniformly distributed access pattern. If the list wraps around the maximum number of steps is

$$s_{MAX} = \left\lceil \frac{N}{2} \right\rceil \tag{2}$$

and the mean number of steps

$$\bar{s} = \left\lceil \frac{N}{4} \right\rceil \tag{3}$$

With just 100 recordings the number of steps to access a particular recording is therefore approximately 25 on average. With a library of 1000 recordings the mean number of steps to reach a recording is 250. This is clearly too inefficient to provide a pleasant user experience. Imagine further that the user is able to press each navigation button three times per second. It will therefore take more than one minute on average to locate the desired recording.

2.4. Two-dimensional search

For the recorders that display the videos in a two-dimensional grid with a width of 3 the situation is a bit better. Given a wraparound mesh, the minimum number of steps needed is 1, the maximum is

$$s_{MAX} = \left\lceil \frac{N}{6} \right\rceil \tag{4}$$

and the mean number of steps is

$$\bar{s} = \left\lceil \frac{N}{12} \right\rceil \tag{5}$$

Although this is a significant improvement the strategy does not scale well for large video collections. For instance, a collection of 1000 video recordings require 84 steps to reach a recording on average which would take about half a minute with three keystrokes per second.

2.5. Direct access

A theoretically efficient strategy would be to give each recording a number, and the video could be accessed directly using the numeric keypad that is often available on remote controls. With such a strategy the mean number of steps would in most cases be identical to the maximum number of steps namely.

$$\bar{s} \approx s_{MAX} \approx \lceil \log_{10} N \rceil \tag{6}$$

This strategy is problematic for two reasons. Firstly, the user would need to know the code for a particular recording. This is not realistic. One important user interface design principle is recall over memory and one should hence not rely on users remembering facts such as codes. Furthermore, one is dependent on a numeric keypad.

2.6. Hierarchal access

In this study we propose to automatically organize the video recordings into a tree structure which are accessed using a hierarchal menu structure. For a balanced tree structure the number of steps would be:

$$s_{MIN} \approx \bar{s} \approx s_{MAX} \approx \lceil \log_4 N \rceil \tag{7}$$

This would give a relatively narrow and deep tree-structure. According to Hick's law humans prefer shallow and wide structures instead of deep and shallow structures since the time it takes for a human to select from a list of items is logarithmic of the number of items. For example, imagine Hick's law is used with the constants a=50 and b=150, such that

$$T = a + b \log_2 (n+1) \tag{8}$$

where T is the response time, n the number of menu items, and a library of 1000 video recordings. Using an narrow four-way tree structure, $log_4(1000) = 5$, then the total decision time is 5(50+150 $log_2(4+1)$) = 1991 ms. Instead, we use a shallow three-level structure with 10 items on each level then the total decision time is 3(50+150 $log_2(10+1)$) = 1706 ms. This represents a less cognitively demanding strategy. However, the number of actual steps would be larger. While only 5 steps would be needed for the optimal strategy, the shallow strategy would require 3 $log_3(N)/4$ = 9 steps.

Note that it is not realistic to achieve a completely balanced four-way hierarchy as the one outlined here. According to [17] an interface of the type that is presented on a regular TV-set should not have more than 12 items. Therefore, the first two levels should be limited to 12 categories and the last levels should be as balanced (equally distributed) as possible. The next section outlines strategies for achieving this.

3. Viewing Behavior

Video recordings are usually associated with information that is important to successful classification. This information includes start-time of a recording, date of a recording, end-time of a recording (or duration) and the channel recorded. Additionally, a recording may have been tagged manually or automatically. However, we will assume that no such labels are available, i.e., unsupervised cataloguing of video recordings. Based on these pieces of information the following can be assumed:

Content type is often linked to the channel where the recording is made, i.e., certain channels broadcast specialized contents. There are often dedicated move-channels, sports channels, news channels and cartoon channels.

Content type is often linked to the time of day when the recording is made. I.e., some multi-purpose channels may broadcast children's programs during early evenings, news programs in the middle of the evening and feature films late at night.

Content type is often linked to the day of the week when the recording is made. First, there is a distinction between weekends and weekdays as channels typically broadcast different types of programs during the weekdays than during the weekends. Furthermore, certain programs such as TV-series maybe broadcast once a week at fixed timeslots.

Content type is often linked to the duration of a recording. For example, TV-series such as comedy shows often last about half an hour, other TV-series, such as dramas may last from half an hour to one hour. A typical feature film is typically 90 minutes. Sports broadcasts and other special events may last several hours, i.e., 4 hours, or even more.

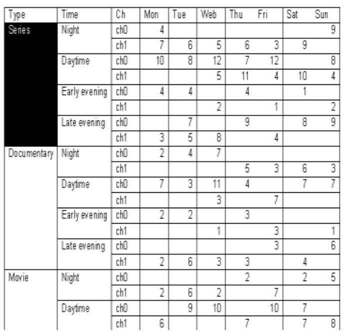
4. Organizing Video Recordings

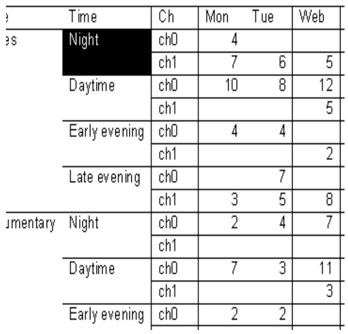
Video recordings are clustered using a dynamic subdivision algorithm. The aim is to overall reduce the number of steps needed to reach the desired recordings. Recordings are first subdivided into clusters based on their duration. Next, each of these classes is subdivided into classes according to start-time of the recording. Next these classes are again subdivided into classes according to the day of the week and finally the resulting clusters are subdivided according to channel. A cluster is only subdivided once its size exceeds a minimum threshold. Each recording is represented by a data-tuple [*ch*,*st*,*dw*,*d*], where *ch* is the channel, *st* is the start-time, *dw* is the day of week and *d* is the duration. The start-time *st* is discretized as follows:

$$st(x) = \begin{cases} morning: 0 \le x < 6 \\ daytime: 6 \le x < 17 \\ early-evening: 17 \le x < 19 \\ late-evening: 19 \le x < 24 \end{cases} \tag{9}$$

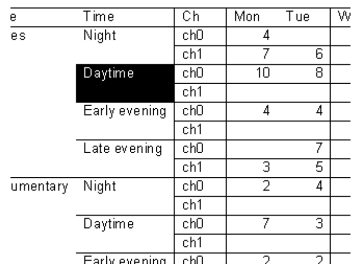
where *x* symbolizes the actual start-time hour. Moreover, the duration *d* is discretized as follows:

$$d(x) = \begin{cases} series: 0 \leq x < 45 \\ documentary: 45 \leq x < 75 \\ movie: 75 \leq x < 150 \\ sports-event: 150 \leq x \end{cases} \tag{10}$$


a) Selecting category based on recording duration, i.e., short recordings less than one hour.

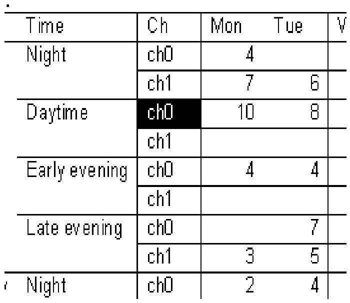


b) Selecting time of day.

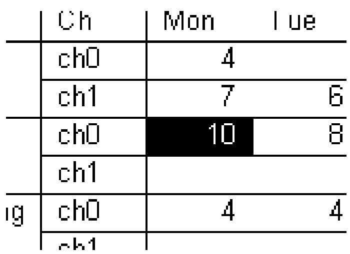


c) Selecting daytime recordings.

.



d) Select recordings on channel 0.



e) Select recordings on Monday. There are 10 recordings in this category.

f) The list of recordings for the chosen category as displayed. The user selects the desired recording. Recordings are sorted according to age.

Figure 3. A low-fidelity prototype visualizes how selections are made using remote control navigation keys with large video collections. The data in this collection represent a simulation of 500 recordings over the duration of a year. Up down is mapped to the up down buttons and left and right are mapped to the left and right buttons, respectively.

where x symbolizes the actual duration in minutes. The subdivision strategy is outlined in the following algorithm.

```
ClusterRecordings(Input: v: list of recordings.
                   Output: o: list of recording clusters)
begin
  Cluster(v,durations,t1);
  Cluster(t1,start-times,t2) ;
  Cluster(t2,days-of-week,t3);
  Cluster(t3,channels,o);
End

Cluster(intput: List L, Categories K
            Output: R list of clusters)
  begin
    for each category d in K
       if |L| > minClusterSize
        then
          Cluster c = all elements in L with category d
          Add cluster c to list R
       else
          R := L
       end
```

In our experiments the minimum cluster size minClusterSize was set to 10. The prototype in Figure 3 shows these clusters could be used in the user interface.

5. Experimental Evaluation

To assess the retrieval improvements with an automatic video organization scheme a series of simulations were run where recordings were randomly generated during the course of a year. For comparison the mean number of retrieval steps for the generated recordings were used as the performance measure. Namely:

$$\bar{s} = \frac{1}{N}\sum_{i=1}^{N}|s_i| \tag{11}$$

where $|s_i|$ is the number of steps needed to retrieve recording i and N is the total number of recordings. Figures 4 and 5 show the results of the simulations. The results reveals that for the next few years the strategy described herein would allow efficient access to medium sized video collection of up to 800 recordings in less than 7 steps. However, for a longer time perspective the results are less encouraging. Figure 5 shows the results up to 20,000 recordings which one could be expect being the norm in year 2020. Clearly, the number of retrieval steps grows linearly with the recording capacity although the strategy is an improvement by a factor of 100 compared to linear access. However, for 20,000 recordings a mean of 140 steps are required, which is totally unacceptable form a usability perspective. Although, many simplifications are made in the simulations it is clear that conventional navigation keys are too limited for the emerging interaction tasks, and that other interaction paradigms may be more suitable, such as touch sensitivity.

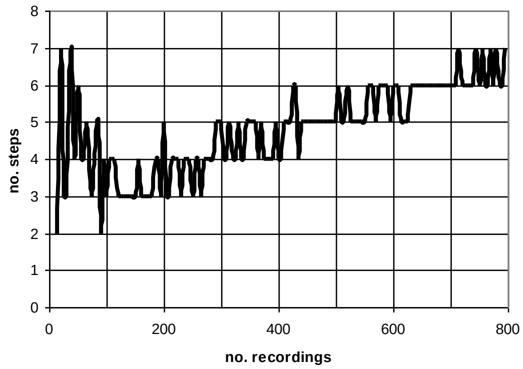


Figure 4. Simulation of mean number of steps to access recordings on personal video recordings as capacity increases for the next five years.

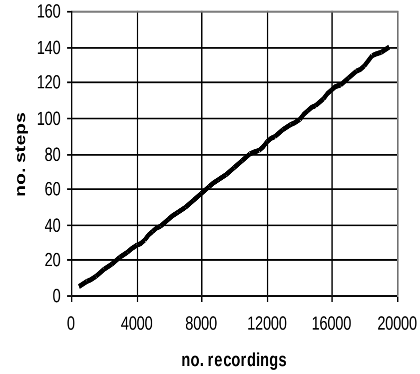


Figure 5. Simulation of mean number of steps to access recordings on personal video recordings as capacity increases for the next ten years.

6. Future Thoughts

Different family members may have different viewing habits. Children may view recordings during the day, young adult early in the evening and parents late at night or during weekends. It would be interesting to look at whether a recommendation engine could be employed to recommend recordings, and especially filter out irrelevant recordings based on the time of access. For example, if the system identifies that certain recordings of certain move channels are viewed in the evening and cartoons are viewed during day then the options for viewing the movie channels are made less accessible during the day and more accessible during night, while the cartoon content is made more accessible during the day than during the evening. Another important issue that needs exploring is that of fuzzy clusters, i.e., that a recording can co-exist in several clusters simultaneously to ease the search for recordings.

7. Conclusions

This paper addressed the emerging problem of growing video collections on personal video recorders and how to access these through remote control navigation buttons. Current practice will quickly become insufficient with the growth in storage capacity. A strategy for organizing video recordings based on human viewing patterns was proposed. Experimental evaluations reveal that the number of retrieval steps is greatly reduced. The strategy is easy to implement with a low computational cost. However, the strategy does not scale to very large collections and interaction techniques based on other technology than traditional navigation buttons is likely to meet the projected future needs in personal digital video recorder video retrieval.

8. References

[1] Y. Wang, P. Zhao, D. Zhang, M. Li, and H. Zhang, "MyVideos – A System for Home Video Management," in Proc. Multimedia'02, ACM Press, 2002, pp. 412-413.
[2] J. C. Platt, M. Czenvinski, and B. A. Field, "PhotoTOC: Automatic Clustering for Browsing Personal Photographs," in Proc. ICICS-F'CM 2003, IEEE Press, 2003, pp. 6-10.
[3] T. Hayashi and M. Hagiwara, "Image Query by Impression Words-The IQ1 System," IEEE Transactions on Consumer Electronics, vol. 44, no. 2, pp. 347-352, 1998.
[4] K. Nakane, I. Otsuka, K. Esumi, A. Divakuran, and T. Murakami, "A content-based Browsing System for A HDD and/or Recordable DVD personal Video Recorder," in Proc. IEEE International Conference on Consumer Electronics, IEEE Press, 2003, pp. 26-27.
[5] H. J. Zhang, J. Wu, D. Zhong, and S. W. Smoliar, "An Integrated System for Content-based Video Retrieval and Browsing," Pattern Recognition, vol. 30, no. 4, pp. 643-658, 1997.
[6] F. Arman, R. Depommier, A. Hsu, and M.-Y. Chiu, "Content-based Browsing of Video Sequences," in Proc. Multimedia 94, ACM Press, 1994, pp. 97-103.
[7] A. Girgensohn, J. Adcock, and M. Cooper, "Simplifying the Management of Large Photo Collections," in Proc. Human-Computer Interaction -- INTERACT'03, IOS Press, 2003, pp. 196-203.
[8] S. Harada, M. Naaman, Y. J. Song, Q. Wang, and A. Paepcke, "Lost in Memories: Interacting With Photo Collections On PDAs," in Proc. 2004 Joint ACM/IEEE Conference on Digital Libraries, ACM Press, 2004, pp. 325-333.
[9] A. Graham, H. Garcia-Molina, A. Paepcke, and T. Winograd, "Time as Essence for Photo Browsing Through Personal Digital Libraries," in Proc. JDCL'02, ACM Press, 2002, pp. 326-335.
[10] M. Naaman, S. Harada, Q. Wangy, H. GarciaMolina, and A. Paepcke, "Context Data in GeoReferenced Digital Photo Collections," in Proc. MM'04, ACM Press, 2004, pp. 196-203.
[11] K. Rodden and K. R. Wood, "How Do People Manage Their Digital Photographs?," in Proc. CHI'03, ACM Press, 2003, pp. 409-416.
[12] D. Bainbridge, C. G. Nevill-Manning, I. H. Witten, L. A. Smith, and R. J. McNab, "Towards a Digital Library of Popular Music," in Proc. DL'99, ACM Press, 1999, pp. 161-169.
[13] R. van Gulik, F. Vignoli, and H. van de Wetering, "Mapping Music in the Palm of Your Hand, Explore and Discover Your Collection," in.
[14] K. Komine, Y. Sawahata, N. Uratani, Y. Yoshida, and T. Inoue, "Evaluation of a Prototype Remote Control for Digital Broadcasting Receivers by using Semantic Differential Method," IEEE Transactions on Consumer Electronics, vol. 53, no. 2, pp. 561-568, May 2007.
[15] J. Nichols, B. A. Myers, and B. Rothrock, "UNIFORM: Automatically Generating Consistent Remote Control User Interfaces," in Proc. CHI 2006, Montreal Quebec, Canada, ACM Press, 2006, pp. 611-620.
[16] E. Diehl, "User Recognition of Standard A/V Control Pictograms," IEEE Transactions on Consumer Electronics, vol. 40, no. 3, pp. 446-450, Aug. 1994.
[17] M. C. Maguire, "A review of user-interface design guidelines for public information kiosk systems," Int. J. Human-Computer Studies, vol. 50, pp. 263-286, 1998.

Authors

Frode Eika Sandnes received a B.Sc. in computer science from the University of Newcastle upon Tyne, U.K., and a Ph.D. in computer science from the University of Reading, U.K. He is currently a Professor in the department of Computer Science at Oslo University College, Norway. His research interests include human computer interaction.

Yo-Ping Huang received a Ph.D. in electrical engineering from Texas Tech University in 1992. He is currently a Professor in the Department of Electrical Engineering and Secretary General at National Taipei University of Technology, Taiwan. His research interests include intelligent information retrieval, data mining, artificial intelligence, and application systems for handheld devices. Prof. Huang is a senior member of the IEEE and a fellow of the IET.

Yueh-Min Huang received the B.S. degree in Electrical Engineering from National Cheng Kung University, Taiwan and both the M.S. degree and Ph.D. degrees from Arizona State University. He is currently a Professor and the Chair in the department of engineering science at National Cheng Kung University. His research interests include e-learning, sensor networks and embedded systems.