Designing Socially-Aware Video Exploration from Community Assets

The previous chapter provided the basis of socially-aware multimedia authoring. Our results validated the main assumptions, showing that users appreciate the importance of video sharing for building common experiences and for increasing the feeling of togetherness with others. Our results also indicated that current video sharing services fail to meet users’ needs, because they miss useful mechanisms for navigating media and do not take into account emotional intensity and intimacy. In this chapter we argue that there is a need for useful mechanisms for navigating and sharing media, and socially-aware video management systems should provide efficient automatic processes to manage personal interests.

The wide availability of video recording devices in mobile telephones and pocket cameras has made documenting shared events easy (see Figure 3.1). The collected set of videos provides a rich archive from which users can enjoy content that matches their personal interest. Unfortunately, current browsing tools, including social networks, are not geared to supporting this form of selective consumption; these tools are geared towards throwing away unwanted content from a single collection, and not for browsing a broader community collection of temporally aligned alternatives. Current video tools often support only a high-level

---

1 This chapter is based on the following paper:

abstraction of objects and events, and do not help users to explore community videos that portray people within their social circle. Even though social networks archive media based on higher-order social relationships, they do not provide support for searching and navigating media content that was captured at a particular event by different camera people.

Most social events have an inherent structure that can be used to aid searching for content. We can take advantage of this structure for the development of socially-aware video exploration interfaces. Most participants at an event will attach different levels of importance to any given sub-event, based on their personal/social preferences. If we consider a high school concert, it has a structure (the order of the songs), a sub-structure (individual songs) and multiple levels of sub-sub-structure: solos, duets, vocal announcements and other often-unpredictable happenings. As discussed in the previous chapter, not everyone at the concert (or viewing it) will be equally interested in all parts. Parents will focus on their own children, students on their friends, and invited guests on the clock.

This chapter focuses on our efforts in designing and implementing an interface for browsing community assets, in which the relationships between users of the system and performers, featured in the videos, play an essential role in content selection. Our work, which follows the emotional intensity guideline (see Chapter 2), includes our findings and key results from the two-phased series of evaluations. In the next chapter we will show that such social bonds are key not only for navigating a shared media space, but also for authoring personalized stories users care about. Here, we focus on the importance of providing a rich representation of an event (in this case, a high school concert) in a way that helps users to navigate and explore a community repository based on their social/personal interests. The research question we address is:

**Question 1.3** Does a socially-aware video exploration system provide an identifiable improvement over current approaches for accessing and navigating a repository of shared media?

To answer this research question we first present a browsing interface, and the underlying system infrastructure, that allow for socially-aware exploration of a collection of media assets captured in an event. Users can explore and navigate (fragments of) video clips recorded by several people based on their own personal/social interests. The design, deployment and evaluation of the system resulted in the identification of key requirements for this novel type of browsing
Community-based Browsing

interfaces. In particular, our approach 1) supports exploration based on the inherent event structure; 2) it makes use of contextual information to help in the navigation process; 3) it allows for flexible searches based on combination of filters; and finally, 4) it provides a way to switch between cameras angles that might have captured different aspects of the event.

The structure of this chapter is as follows. First, Section 3.1 provides an overview of the design and evaluation of an initial prototype system for socially-aware video exploration. Based on the users’ feedback, a set of functional requirements was gathered. Then, in Section 3.2 we describe the design and implementation of the second version of the browsing interface that addresses these requirements. Next, Section 3.3 reports the evaluation of such system, analyzing the results. Finally, Section 3.4 provides a reflection on how our findings fit in the context of this thesis.

Figure 3.1. Typical interface for watching videos on the Web. It does not take into account the social affinity between viewers and subjects featured in the video.
Chapter 3. Designing Socially-Aware Video Exploration from Community Assets

3.1 Community-based Browsing

The family interviews and focus groups in the beginning of this journey (see Chapter 2) provided us valuable data for identifying a series of requirements. The conclusion was that current social media sharing interfaces are not adequate for satisfying the expectations of strong ties. In this chapter, we focus on innovative interfaces that help users to explore a shared media repository they have social affinity with (emotional intensity guideline defined in Chapter 2). The final goal is to provide interfaces that can help shaping and sharing memories of important events with family members and friends.

The starting point of our investigation was traditional video browsing interfaces, such as YouTube (see Figure 3.1). Nevertheless, early in this process,
we realized that this kind of service does not provide social filters (e.g., to select videos by a particular performer) for concert videos, and it does not take advantage of the temporal relationships between videos belonging to the same event.

To address these limitations of current video sharing services, our initial video browsing interface offered two views for exploring community contributed video clips. The **thumbnail view** (Figure 3.2) displayed media assets in a paginated grid, while the **timeline view** (Figure 3.3) showed how recorded videos temporally fitted the event timeline. In both views a user could apply six different filters to refine a query. These filters were: *all media*, *my media*, *cameras*, *people*, *instruments* and *events*. ‘All media’ referred to all videos uploaded to the system. ‘My media’ restricted navigation to only the videos uploaded by the current user. ‘Cameras’, ‘people’, ‘instruments’, and ‘events’ filters would display the respective annotated video clips based on the filter selection (e.g., ‘Julia’ or ‘Drums’).

![Figure 3.3. Initial prototype implementation for browsing videos (timeline view).](image-url)
Besides allowing for navigating the video clips, our initial interface also enabled users to annotate media assets and to correct existing annotations that could be wrong. When showing a video clip, a user could ‘flip it’ over and access all annotations related to that clip (as shown in Figure 3.4).

Using the footage recorded during the Big Band concert in Amsterdam, potential users were invited to evaluate our initial system. Details about the methodology and user assessment can be found in Chapter 2. In the remaining of this section, we discuss the results regarding media exploration obtained in the first evaluation phase. From these results a set of new requirements were elicited, and used for the design of the second phase.
3.1.1 Phase 1 Evaluation

In general, participants’ feedback for the first version of our system was positive (see Figure 3.5). Four (4) out of 7 participants said it was better than traditional tools to find people they cared about (Q1.1). We received slightly better feedback when we asked whether our system was better to browse videos recorded by other parents (Q1.2). These results are directly aligned with the requirements of emotional intensity and easiness of use.

During the evaluation session, participants were actively looking for video clips of their close friends and relatives. In particular, some participants wanted to immediately share video clips with members of their close circle. “Can I send it now?” was a common reaction after seeing a video clip they especially liked. When asked how they would share the videos, teenagers expressed they would rather download the video files to their local computers, send a link of a particular video by email or share on YouTube and/or Facebook. Parents, on the other hand, indicated that a ‘Burn to DVD’ functionality of the selected videos also would be convenient given that grandparents usually do not have Internet access at home.

When prompted about what they remembered of the concert, most participants that attended it said that they recalled superficially the spatial arrangement of the stage (see Figure 3.6). At this point, some participants

![Phase 1 Evaluation](image.png)

Figure 3.5. Results of the questionnaires from phase 1 evaluation.
mentioned that it would be interesting to have a spatial representation of the concert venue to help browsing the event footage. When inquired about particular events they remembered, participants reported on solos performed by different musicians. Among the youngest participants, an event in particular was pointed out as the most memorable of the concert.

“I think that the jamming at the end I liked the most... I found that the most memorable of the whole evening...” (Friend of some performers)

In some cases participants complained – and were desperate – when the quality of the video was not good enough or when the metadata was wrong (see Figure 3.7). Most participants expressed they would add/correct metadata with our system (Q1.3). However, they were quite resistant about the amount of time they would spend on this process, arguing that it demanded a lot of effort.

“It is not my problem (correct the wrong metadata)… people don’t have time to play with the system.” (Uncle of a performer)
When questioned about the filter functionality, participants appreciated such feature because it would allow them to retrieve only the videos related to their interest. Nevertheless, almost all participants manifested interest in using a combination of filters, when searching for videos (e.g., show all videos of the trombone player in the 3rd song). Despite being feasible using the recommendation algorithm presented in Chapter 2, such functionality was not contemplated in the first version of our user interface. At last, some participants also mentioned that a person or instrument should be considered featured in a video only if this was a prominent shot, e.g., close-up or solo. They would not be interested in a video clip in which the subject of interest barely appears.

“If he (my nephew) is in the background but he is on the shadow it is OK but I would like to see a video in which he really shows up… My mother (performer’s grandma) would not enjoy seeing this video of him because there is not much to see.” (Uncle of a performer)
3.1.2 Lessons Learned

In the first evaluation phase we followed an interactive approach, where a number of new requirements were defined. The most relevant observation was the necessity of providing contextual information for browsing, searching and watching community assets.

On the one hand, the thumbnail view did not show the temporal relationships between the video clips. On the other hand, the interface of the timeline view was considered complex. Participants were looking for a more intuitive and simple visualization model. We observed during the evaluation process that they tended to remember the inherent structure of the event (e.g., the concert program or spatial arrangement). Rather than treating each media asset as a discrete entity, archival theory and practice suggests that digital videos should be managed, preserved and presented to users in a way that reflects the social and documentary context in which they were originally embedded [8]. This argument leaded us to the specification of following requirement:

i. **Support inherent event structure**: users indicated the need for a more intuitive metaphor to organize or cluster community assets. Such approach would help them in exploring and searching for people or events of interest;

Although the interface allowed users to add/remove and correct existing annotations, these were not directly accessible. In order to see and change any information regarding a video clip (e.g., associated performers, songs or instruments), users had to click on a button to show the annotation interface (see Figure 3.4). When playing a video, the same problem was evident: annotations were again ‘hidden’ behind the media. In some situations, users would just click and watch a video in order to know more about its content. This was a time consuming process that led to frustration of the users. Based on these issues, we introduce our second requirement:

ii. **Make contextual information explicit**: feedback from users suggested that by clearly showing associated annotations, it would facilitate the browsing experience. It would also minimize the chance of ‘blind’ navigation or of getting ‘lost’ in the media space;
In the previous section we said that both thumbnail and timeline views offered a number of different filters for content selection. Despite appreciating this functionality, participants manifested interest in using more than one filter at the same time when searching for people or events of interest. The use of individual filters did not fulfill their needs. Based on this we present our next requirement:

iii. *Allow combination of filters*: users should be able to combine filters to compose robust queries. Such functionality would allow them to find videos of interest more effectively and faster;

Users feedback also suggested that they would like to have a spatial representation of the videos, in which content recorded from different angles could be activated in parallel. The work of Kennedy and Naaman [44] indicates that in a music scenario, like the one addressed in this thesis, alternative camera views could significantly reduce the required time to scan or to watch the content, while still providing a complete overview. In these lines, we introduce our last requirement:

iv. *Allow multi-camera navigation*: when watching a particular event (e.g., a solo), users should be able to switch between different camera angles (if there is any other available). Such functionality would enrich the browsing experience by providing spatial context.

In this section we introduced a set of functional requirements based on user feedback and results from the first evaluation phase. These new functional requirements motivated the design and evaluation of a second prototype system. In the next section we discuss our efforts for providing more effective socially-aware visualization mechanisms and innovative navigation paradigms.

### 3.2 Socially-Aware Media Browsing

The first prototype was helpful for better understanding user requirements for socially-aware video exploration of community assets. The evaluation results suggested we were in the right direction and helped in identifying a number of requirements for improving the user experience. With such requirements in mind, we started a new design from scratch. The browsing component discussed in this
section intends to simplify the exploration of media assets, without compromising the flexibility of query specification [47].

To address our first functional user requirement, we designed an interface based on the concert program (Figure 3.8 (2)). This digital version of the original paper-based program handed out at the day of the event (Figure 3.9) clusters songs in two columns. Rather trivial in concept, it provides a general overview of the event schedule. In this interface, performers have a prominent position at the top (Figure 3.8 (1)). After all, these ‘raising stars’ are the main reason for users (friends and family) to use the system.

For each song in the concert program, a few video clips are recommended. This design choice provides contextual information without having to select a specific song. We also implemented a clip hovering functionality that shows a key frame animation on mouse over. It aims at providing a summary of the video
without the need to watch it. These design decisions are directly aligned with our second requirement.

Hovering the mouse over interface elements (i.e., performers, songs and clips) also provides efficient and informative feedback. For instance, when a user hovers the mouse cursor over a performer thumbnail, the associated songs and media clips containing that person are highlighted in the user interface. This functionality, which has been designed to react in rapid response time, reduces the short-term memory load [47] and makes clear the relationship between performers, songs and clips.

Another functionality supported in the new prototype is the specification of queries based on performers. When the user clicks on a particular performer, the selection is sent to the server, which recalculates the recommendations considering the selection. Our design also allows for more complex query specifications such as the combination of two or more performers. In this case, a conjunction operator is used to connect the selection of performers. Thus, only songs (and the respective video clips) in which there is an intersection among the selected performers will be
Figure 3.10. Supporting combination of filters. In this example, one performer is selected and the mouse cursor over another performer highlights the songs and video clips in which both performers played together.

highlighted in the interface (see Figure 3.10). This functionality addresses our third requirement by allowing participants to search for videos using combined filters.

Next we present our efforts on integrating context information, video playback, and supporting multi-camera navigation. As aforementioned, some video clips are listed in each song of the concert program. These videos are the entry points for media playback and multi-camera navigation. The video clip recommendations are based on the selected search terms and on the user profile (as we will see in the next chapter, the user profile is computed automatically considering user recording behavior).

When the user clicks on one of the recommended video clips, the playback interface is launched as illustrated in Figure 3.11. This interface is divided in three
socially-aware media browsing

Main areas: media player (1), video clip information panel (2), and alternative views of a video clip (3). The information panel shows metadata associated to the video clip (e.g., who has recorded it, the number of views). It also provides information that is constantly updated based on the video playback (e.g., the song elapsed time). This panel also offers users a way to share a link of the current clip with someone by e-mail, to download the current clip, or to inform the system administrator that the clip has inappropriate content. The rationale behind this last functionality is to cope with the privacy concerns discussed in Chapter 2.

The area of alternative views of a video clip (Figure 3.11 (3)) presents other camera angles that happened at the same time of the main video. In other words, this area shows concurrent video clips recorded by other people during the event. By design choice, only a limited number of alternative views is presented (or

Figure 3.11. Interface for watching video clips.
recommended) to the user. It is possible that more cameras were active at that point in time. The position of each camera is set when the player interface is launched and remains unchanged during playback.

When the user clicks on an alternative camera view, this will take the place of the main video, and the playback will continue from the same point in time as if the user had changed his position or angle. Such interface provides support for watching and navigating the media space, which directly addresses our fourth and last requirement about multi-camera navigation. Due to performance limitations in a Web browser environment, alternative videos are not played at the same time as the main clip. As an elegant workaround, our design provides a camera update functionality that – during the playback of the main clip, – periodically changes the key frame of each alternative. This approach aims to minimize the blind camera navigation problem discussed in Section 3.1.

3.3 Evaluation

Using the footage recorded during the Woodbridge high school concert (UK) in the beginning of November 2011, 13 people from 6 families participated in the evaluation of the new prototype system. While this section reports on the observations from the interactions of all the 13 participants, the quantitative data shows the answers from 9 people (the others did not fill in the evaluation questionnaires). More information about the methodology and participants’ profiles can be found in Chapter 2. Next, we analyze the user responses and discuss the findings regarding socially-aware video exploration. Our results are based on a qualitative analysis of the interviews and observation of the system usage.

3.3.1 Results and Findings

Figure 3.12 and Figure 3.13 present the results of the questionnaires. Overall, participants appreciated the browsing interface (Q2.1). They indicated that it is useful for finding videos of performers and that it is better than traditional tools to explore the event media space (Q2.2 and Q2.3, respectively). Therefore, users would find videos more efficiently using our system (Q2.4 and Q2.5). If we compare with the results obtained in the initial evaluation (see Figure 3.5), there was a clear improvement, even though these were two distinct experiments.
“It (the browsing interface) has everything in one place and you can access other (people’s) videos without having to import / open them.”
(Brother of a performer)

The concert program metaphor was well assessed by our participants. In general, they expressed that this inherent event structure provides a simple and intuitive overview of what happened during the concert. Performers’ thumbnails displayed at the top of the user interface were also appreciated. Participants said this was a good way to quickly look for videos they were interested in.

“Very easy to use! Performers at top is a good idea, and the concert programme is very clear!”
(Father of a performer)

When asked how much they liked the mouse over functionality in the concert program, 8 out of 9 participants said a lot, while the other participant said some (Q2.7). This was by far the most appreciated functionality of our prototype system. Participants enjoyed the rapid contextual information feedback when they hovered any of the interface elements (e.g., performers, songs or videos).

“I really liked the mouse over feature in the concert programme!”
(Mother of a performer)

“This is really good!”
(Performer about filters and mouse over functionality)

One of the participants mentioned that this mechanism was a bit slow though. Rapid response time is critical to support effective feedback. Providing highly responsive interactive results is important for dynamic browsing interfaces like ours, and fast response time for query reformulation allows the user to try multiple queries rapidly [47].

One aspect that needs further investigation is how to present recommendations for each song. Some users indicated that more recommendations could be showed: they assumed that there were more videos available. Apart from that, they seemed to like the video recommendations (see Q2.8). As mentioned earlier in this chapter, our video recommender takes into account the social bonds between users and performers. In the next chapter we detail the profiling approach used by our video recommender.
“I’m wondering why it (the browsing interface) particularly picked those 2 videos.” (Father of a performer)

While exploring videos displayed in the concert program, users expressed that they were having an engaging experience, but they did not have an option to play a song from begin to end. This feedback suggests the need for supporting more complex narrative alternatives that not only take into account the temporal alignment between videos, but also the preferences and social relationships of each individual user. This subject is the focus of the next chapter, which discusses the

Figure 3.12. Results of the questionnaires from phase 2 evaluation.
balance between automatic generation of video narratives and use of manual processes to reflect personal imprint.

“I think when you got individual songs, or individual pieces, [...] you might well want to, say, see the whole three minutes or something from the beginning.” (Father of a performer)

Our participants appreciated the multi-camera navigation support (Q2.9). This functionality raised the demand of having all the alternative videos playing at the same time and for seamless transition between camera angles. However, users also were aware of the browser and bandwidth limitations in our scenario. In some sessions there were some technical problems when switching from one camera view to another. Instead of starting the new clip from the current time, the playback would start a video from the beginning. This problem clearly led to frustration.

“I liked having a lot of different camera angles, which is something you don’t get with anything else.” (Performer)

“Found the alternative views slightly complicated as regards case of use – couldn’t always tell whereabouts in the performance we were, seemed a bit jumpy. Probably just an issue of getting to grips with the programme though!” (Performer)
Chapter 3. Designing Socially-Aware Video Exploration from Community Assets

Still regarding the multi-camera navigation interface, some participants suggested that it would be nice to have a visual representation of the duration of each clip within the song, as it would help them to situate themselves temporally. This goes in the same direction of the work proposed by Yu et al. [77].

Regarding video annotation, 7 participants declared they would explicitly rate videos while watching (Q2.6).

“I would rate a clip while watching to tell it does not belong to a song or it has poor quality… just to make sure it would not be recommended again!” (Mother of a performer)

“I would tag videos (thumbs up/down) as not being of good quality or in poor position e.g., performers face not visible as obscured by music stand.” (Father of a performer)

A few participants mentioned they normally do not use to rate videos at all.

“I never really use the rating features of YouTube.” (Brother of a performer)

3.4 Discussion

In this chapter we presented our efforts in designing and implementing an interface for browsing a collection of user-generated videos from a shared event. The interface aimed at helping users to easily access contents based on their social interests. This chapter described a two-phased development and experimentation.

First, we discussed the design and development of our initial prototype system. The evaluation of this tool allowed us to identify a number of functional user requirements for interacting with a set of videos from the same concert. These findings guided the development and evaluation of a new video browsing interface. Results from the experiments show that our new prototype satisfied the requirements and led to a clear improvement when compared to the initial system. Using a concert program metaphor (requirement i), participants could search for videos using combined filters (requirement iii) and experience moments of interest from different camera angles (requirement iv). Not to forget that our system provides efficient and informative feedback to help in this process (requirement ii).
Overall, our design decisions have improved the ability to explore videos users care about, among a pool containing the recordings of different parents. Our results clearly indicate that a socially-aware video exploration system like ours (which fulfills the emotional intensity guideline and social connectedness requirement introduced in Chapter 2) provides an improvement over current tools for accessing and navigating a repository of shared media assets. These results directly answer the research question asked in the beginning of this chapter.

Enabling users to explore an event and search for video clips they, and other participants, have recorded is an important step towards making personal media more accessible. But it is just the beginning. Individual video assets most of the times do not provide rewarding narrative experiences that help users remember important events. In the next chapter we discuss the balance between automatic and manual processes for creating personalized stories from community assets.