Introduction

Nearly a decade ago at an ACM SIGMM\(^1\) retreat, one of the grand challenges to the multimedia research community was to develop media authoring tools that would make creating complex media titles as easy as using a WYSIWYG (What You See Is What You Get) word processing system [41]. Since that time, a number of consumer-level video editing tools have been developed that would lead a casual observer to believe that multimedia authoring is a solved problem: using tools like iMovie\(^2\) or Windows Movie Maker (or even more sophisticated tools such as Adobe Premiere or Final Cut Pro), even relatively novice video editors can match their talents with the likes of Sergei Eisenstein (see Figure 1.1).

The process was further simplified by modern content capture tools, such as smartphones, in which recording, (simple) editing and integrated uploading were combined into a single task. In many ways, video editing has been reduced to transferring content taken from a (personal) camera to a computer, throwing out frames that are unwanted, and uploading the resulting production to a video sharing site. While it is indisputable that media capture and sharing is much easier than at any time in the past, we wonder if the resulting products of such authoring interfaces have provided any significant advances for the viewers of media content. It is even questionable if there have been significant advances for content authors.

---

\(^1\) SIGMM is the Association for Computing Machinery’s Special Interest Group on Multimedia, which specializes in the field of multimedia computing, from underlying technologies to applications, theory to practice, and servers to networks to devices.

\(^2\) The services and technologies mentioned in this thesis, if unknown, could very easily be identified via a simple online search, therefore they will not be Web-referenced.
Recent data suggests not. In spite of the ubiquity of video cameras and the growth in video viewing on social networking sites, about 82% of Internet users have never uploaded even a single video [45]. Although most YouTube uploads are amateur content, professional videos are preferred to amateur productions online [38]. From the perspective of personal videos, the problem of creating and sharing content has several dimensions. At a lower level of abstraction, video is not semantically linked; therefore, searching and selecting the desired piece of content to share can become too laborious. At a high-level of abstraction, creating compelling videos – videos that meet the needs and desires of the viewer, not only the producer – is a complex task [69]. Viewers generally expect professionally produced content (in terms of shot selection, story pacing and logical narrative), which most amateur users cannot provide.

Although a number of research efforts have addressed content creation from different perspectives [6][19][26][61], based on user studies [57][63] we observed that traditional authoring tools and current social media services fail to address the interpersonal relationships for sharing media that is personal and important to families and small social groups. Our assumption is validated by other studies [24], which concluded that social media applications like Facebook do not take into account the interpersonal tie strength of the users. Thus, we can conclude that the current media landscape demands a revision of traditional research on multimedia authoring to empower users in recalling and sharing personal media experiences with friends and family. This discussion leads to the following question:

**Main Question**  *Is a new multimedia authoring paradigm required to enable end-users\(^3\) to share more personal media within their social circle?*

During the past years our research work has focused on the study of *socially-aware multimedia authoring*. Working with a group of users at local high schools in two different countries (UK and the Netherlands), the process involved research on different facets related to the creation and sharing of multimedia artifacts composed of personal videos. Apart from the underlying mechanisms for navigating and reusing personal content, this thesis work argues that a new paradigm, *socially-aware multimedia authoring*, is necessary to better fit end-users’ needs. One important aspect of our work is that we decided to follow an interdisciplinary

---

\(^3\) The terms ‘end-user’ and ‘user’ will be used interchangeably in this thesis to describe regular people who operate computer software with minimal technical expertise or previous training.
approach in which both technology and social issues were addressed. As illustrated in Figure 1.2, the core of our methodology integrates knowledge from user-centered design (e.g., need assessment, iterative prototyping and user evaluation) and document engineering.

The remaining of this chapter is organized as follows. In Section 1.1 the main question is split into a number of supportive research questions, which form the main focus of this thesis. Then, the contribution of the thesis is detailed in Section 1.2. Section 1.3 presents the thesis outline and summary of each chapter contribution with respective supportive material. Finally, Section 1.4 overviews the related work, contextualizing the research problem.
Figure 1.2. Timeline of the implementations and evaluations of our system.
1.1 Research Questions

In this thesis, we consider a socially-aware multimedia authoring framework for personalizing video stories from a collection of community assets. The high-level architecture of our framework is sketched in Figure 1.3. The input material includes the video clips that parents agreed to upload, together with a master track recorded by the school. By contributing assets in a shared video repository, each participant gives permission to reuse their own contributions within the community. It is assumed that each participant has the rights to contribute their own material. Privacy and a protected scope for sharing is a key component of our framework. Each media item is automatically associated with the person who uploaded it, and there are mechanisms for participants to restrict sharing of certain clips. Participants can use their credentials for navigating the repository – those parts allowed to them – and for creating and sharing different video compilations intended for different people.

At capturing time, there are no specific filming requirements for users. They can record what they wish using their own camera equipment. The goal is to recreate a realistic situation, in which friends and families are recording at a school concert. This flexibility comes at a cost, however, since most existing solutions that work well in analyzing audiovisual material are not that useful for our use case. As indicated in [59], handling user-generated content is challenging, since it is recorded using a variety of devices (e.g., mobile phones), the quality and lightning are not optimal, and the length of the clips is not standard.

Figure 1.4 enlists four main stages (with the respective application services) that compose the socially-aware multimedia authoring workflow proposed in this thesis: Capture and Processing, Access and Navigation, Creation and Production, and Content Enrichment. We should not forget the importance of looking at the social aspects around personal media in this workflow. The key research questions emerging from each of the four stages are presented below. But first, we take a look into the social requirements.
Chapter 1. Introduction

1.1.1 Social Aspects

Recognizing the importance of looking at personal media as a cornerstone for sharing family experiences [72], the intention of our research is to understand the intersection of social multimedia and social interactions in an asynchronous communication context [37]. We are interested in personal media, and how these can become memory artifacts: the content around which conversation happens. We aim to help small groups of people (such as a family, a school class or a sporting club) viewing, creating, and sharing personalized multimedia. From the technical perspective, a system should combine the benefits of personal focus – knowing whom you are talking with – within the context of temporally asynchronous and spatially separated social meeting spaces.

Sociological theories [24] and user-centered approaches [3][25] have tackled different aspects of the multimedia workflow. For instance, human-centered efforts

![Figure 1.3. High-level architecture of socially-aware multimedia authoring systems.](image)
explore video-mediated communication to share watching videos together over the distance [4]. Similar to us, other studies investigate what people do with media in an asynchronous context, balancing the preponderance of techno-centric work with appropriate user-centric insight [19]. In our work we pay special attention to social theories and human-centered methodologies. Our work, which is predicated on the intrinsic desire to strengthen existing strong ties among people, tackles different aspects of the socially-aware multimedia workflow. That said, the following research question arises:

**Question 1.1** Can a socially-aware multimedia authoring system be defined in terms of existing social science theories and human-centered processes, and if so, which?

**1.1.2 Capture and Processing**

The research question introduced above puts the accent on the social aspects around personal media. While knowledge from online social networks could be mined to determine the strength of ties among people [24], user interest and sentiment analysis also could be used to facilitate media annotation and content

![Multimedia authoring workflow and application services](image-url)
selection [3]. Given the characteristics of end-user content, in this thesis we have chosen to focus on studying the implicit social practices during video capture within groups of people with strong ties. Our assumption, which could be compared to research in the domain of user modeling [29], is that users’ recording behavior can provide useful insights to better understand the interpersonal relationships. The attempt to ‘understand’ end-users and their social practices is just the first step in the socially-aware multimedia authoring workflow. More important are the indications that this new paradigm brings an improvement over the state of the art in multimedia authoring. In this direction, the research question we have is the following:

**Question 1.2** Does the functionality provided by a socially-aware multimedia authoring system provide an identifiable improvement over traditional authoring and sharing solutions? If so, how can these improvements be validated?

### 1.1.3 Access and Navigation

Media selection in socially-aware multimedia is not a case of ‘finding as many essentially equivalent videos of an event as possible’, but ‘finding the relatively few videos within that event that are relevant to me now, and structure them into a story based on my context (and that of the people in the video)’. A key aspect in this process is to support interactive content selection.

While in terms of user experience, user interfaces are important; even more is the underlying interaction design and recommendation mechanisms. In particular, we are interested in technological solutions that can help users accessing and navigating media content with which they have social affinity. Our work acknowledges previous research efforts in video abstraction/summarization [7], content recommendation [40], synchronization and organization of user-generated content from popular music events [44] and home video management and navigation [28]. However, we go a step further by integrating knowledge of the social relationships to improve content searching and selection by individual users of a shared media repository. With this in mind, we ask the research question:

**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?
1.1.4 Creation and Production

Current media authoring is predicated on the notion that content creation is a one-time event. In socially-aware multimedia, content authoring becomes an incremental process of content refinement, sharing and repurposing. ‘Old’ assets remain living entities. This will foster a new generation of create-view-refine-share authoring systems. A key element of this approach is that media gets integrated into some larger narrative story, rather than that the media object is the story itself.

A number of research efforts have addressed this problem by focusing on community video remix [14], automatic generation of video mashups from YouTube content [59], social creation of photo albums [58] and configurable and interactive storytelling [49][52]. The main difference of our work lays on the fact that we do not aim at providing a complete description of an event based on the characteristics of individual media fragments, but personalized video stories (narratives) based on the social bonds between people. Convenience and personal effort are also important factors to consider when generating such narratives. In this context, the research question we have is:

Question 1.4 Where is the balance between automatic and manual processes when authoring personalized narratives users care about?

1.1.5 Content Enrichment

One of the foundations of socially-aware multimedia is that media can take on new meaning based on the insights of downstream viewers. As an example, consider end-user generated comments. They have the potential to enrich and transform the media viewing experience by allowing users to express themselves and interact with others. Currently, media commenting is supported on an overly coarse level. Still, the lack of a embedded ‘media message’ in most personal media content actually presents the viewer of such media with a golden opportunity to superimpose his/her own meaning on top (physically or logically) of the content provided by the media object. We believe that a socially-aware system should enable content enrichment beyond ‘likes’ and out-of-band text comments.

The analysis of user-generated comments around media has resulted in innovative work on the semantic and temporal structure of media events [13], user commenting patterns in video on demand [25] and in live video streaming platforms [34]. Not to forget studies on the aggregated behavior of people and
social media. Examples include motivations behind tagging in *Flickr* [48] and location-aware photo sharing systems [18]. This thesis acknowledges all these efforts for better understanding users and media. But instead of focusing on the aggregation of user interactions around media, we investigate solutions that allow any user – not necessarily the author – to incrementally add personalized comments within multimedia artifacts. By personalized we mean comments that could be used to highlight interesting things for other viewers, e.g., to make a point about a particular event within a video. This discussion leads to the following question:

**Question 1.5** Does the support for timed end-user commenting within pre-authored narratives provide an identifiable improvement over current media commenting approaches?

### 1.2 Our Aim

This thesis investigates mechanisms and principles for togetherness and social connectivity around personal media. The main contribution lays on a new paradigm, *socially-aware multimedia authoring*, which empowers users in telling stories and commenting on personal media artifacts. The work has been evaluated through prototype tools that allow users to explore, create, enrich and share rich multimedia artifacts. Results from our evaluation process provide useful insights into how a socially-aware multimedia authoring and sharing system should be designed and architected, for helping users in recalling personal memories and in nurturing their close circle relationships. Our experimental methodology aims at meeting the requirements needed for social communities that are not addressed by traditional authoring and sharing applications. During this process the intention was not to focus on a specific piece of software, but to take a broader look at the process and its implications. The final goal is to reformulate the research problem of multimedia authoring by emphasizing the importance of the social relationships among casual media authors, featured subjects and recipients of the media.

### 1.3 Thesis Outline and Summary of the Contributions

We summarize below the content and main contributions of each chapter.
Chapter 2 sets the stage by presenting a community video use case in which the social relationships between the people involved plays an essential role. Then, we detail our user-centered methodology, which involved requirements gathering, concert recordings, iterative prototyping and user evaluation. Motivated by social theories, preliminary interviews/focus groups and a survey research about social practices around personal videos, we identify key requirements and specify guidelines for realizing socially-aware multimedia authoring systems. Finally, we report on a long-term evaluation process that validates our approach and shows that socially-aware multimedia authoring is a valid alternative for social interactions when apart. The contributions of this chapter, which directly respond research Question 1.1 and research Question 1.2, include:

- Introduction of a community video use case and motivation of socially-aware multimedia authoring;
- Description of the user-centered methodology followed in this thesis;
- Identification of requirements and specification of general guidelines for realizing socially-aware multimedia authoring systems; and
- Discussion about a 4-year evaluation process that includes the validation of the proposed socially-aware multimedia authoring framework.

This chapter is based on the following papers:

http://doi.acm.org/10.1145/2072298.2072339. (17% acceptance rate)


Chapter 3 considers the development of innovative mechanisms to enable users to browse and navigate a repository of shared media. Context-aware user interfaces and filtering mechanisms are proposed by taking into account
relationships between users of the system and subjects featured in the videos. This chapter also discusses the importance of semantic annotations to describe personal media. Our approach is then compared to traditional (and less individual) media exploration tools. The contributions of this chapter, which directly address research Question 1.3, can be summarized as follows:

- Design and evaluation of an initial interface to facilitate the personalized exploration of a repository of shared media;
- Design and implementation of a new browsing interface based on the requirements elicited in the initial evaluation process; and
- User evaluation of the new interface, demonstrating that, when compared to traditional approaches, we have improved the ability to explore videos users care about, among a pool containing parent-made recordings.

This chapter is based on the following paper:


Chapter 4 compares automatic approaches for generating video stories (or media artifacts) from user-generated content with more manual mechanisms to reflect personal effort and intimacy. Our findings, which directly relates to research Question 1.4, indicate that the balanced combination of manual and automatic processes will be the basis for authoring tools that better fit end-users’ needs. The contributions of this chapter are summarized below:

- Two-phased design, implementation and user evaluation of an authoring system to create personalized video stories from community assets; and
- Discussion about the benefits of a compromise between automatic and manual processes when creating personalized video artifacts.

This chapter contains extracts from the following papers:
Chapter 5 presents mechanisms to support end-user commenting and enrichment of pre-authored video stories. This approach is used as a way to communicate the viewer’s personal view by highlighting a particular event that might be interesting to his/her social circle. This chapter, which directly responds research Question 1.5, brings the following contributions:

- Motivation based on a survey research about media consumption and commenting habits of a group of Internet users;
- Specification and description of temporal document transformations that allow end-users to create and share personalized timed text comments within third-party online videos;
- Design and implementation of a video commenting tool that realizes such document transformations; and
- User evaluation showing that users appreciated the functionalities of our system and would use it to communicate.

This chapter is based on the following papers:
Chapter 1. Introduction


Chapter 6 dedicates to open-ended questions and concluding remarks.

This chapter contains extracts from the following article:


1.4 Related Work

We frame this work within a historical perspective on three areas: conventional authoring systems, interactive storytelling and video mashups/content repurposing.

1.4.1 Conventional Authoring Systems

At the 1993 ACM Multimedia conference, Hardman et al. [43] presented a paper on structured multimedia authoring. Just over a decade later, this study was revised for the initial issue for ACM TOMCCAP [16]. At that time, multimedia authoring was seen by many as a seminal topic within the research community. As described in these publications, several paradigms existed for compositing (or binding) media objects, including:
• **Structure-based composition:** composition where the (often hierarchical) logical structure of the components serves as the basis for generating a particular presentation instance timeline;

• **Timeline-based composition:** composition in which a particular presentation instance determines the content relationships among objects;

• **Graph-based composition:** composition in which the relationships among objects have cause/effect relationships, but limited logical structure; and

• **Script-based composition:** composition where the inherent logical structure of elements is hidden as side effects of a procedural execution model.

All of these methods (of which structure-based remains the most compelling) are examples of relatively formal models in the sense that there is a need for an explicit authoring activity to take place in creating a presentation. This explicit activity was intended to manage the inherent complexity of selecting, editing, combining and positioning media in temporal and spatial dimensions. In many ways, the process was similar to early text processing systems, in which formatting codes and layout directives needed to be directly and overtly inserted into a content stream. In general, formal authoring systems are based on an implicit model in which an editor is assumed to understand the basic aspects of content production. These include understanding:

a) The content alternatives available;

b) The interests (and attention spans) of the intended audience; and

c) The formal or informal narrative and cinematographic principles required to build a compelling story.

While significant steps have been made in better understanding the encoding of narrative structures [32], the management of content and the management of viewer-driven interests provide fruitful areas for new work. We argue that there are two primary reasons that personal content viewers are unresponsive to non-professional content. The first reason is that the opportunity to home-editors represented by b) is largely unexploited by formal authoring systems. In many professional editing situations, all three of these aspects have been well understood, albeit for b) at an aggregate level of detail. For more personal content, home editors would seem to have a tremendous advantage: they typically know the person or persons for whom a particular content object is being created. Sometimes the
intended audience is relatively diffuse (such as one’s 1,000 closest Facebook friends), but other times it can be highly focused: the grandmother of a young high school musician. The second reason that personal content viewers are unresponsive to non-professional content is that conventional formal authoring systems maintain a push model of content rather than a pull model, in which a content viewer is intimately involved in the process of content selection and personalization. This means that the author/editor determines all of the choices, with little infrastructure support for end-user personalization at the detail level.

1.4.2 Interactive Storytelling

During the past decade, various Artificial Intelligence (AI) approaches have been suggested for the creation of configurable and interactive storytelling [49][54]. A main thread of investigation has so far focused on generated content, often involving intelligent animated characters (e.g., Ibanez et al. [35]). Not to forget the use of interactive video as a basis for scenario-driven interactive tours, with additional mini-games for elaborating on specific topics or tasks that arise during exploration process [2]. Another representative example is Vox Populi [68], in which rhetorical documentaries are created from a pool of media fragments, and the Narrative Structure Language (NSL), a production-independent framework for the authoring and delivery of configurable and interactive video narratives [52]. More recently, a system capable of creating different story variants from a baseline video was presented [5].

In general, these systems generate sequencing video shots, while maintaining local video consistency. In order to support the automated generation of the interactive story, extensive use of metadata annotations on individual media objects is made. These systems have been applied to professionally produced media content, using well-defined (and generic) content and story descriptions. Our view on socially-aware multimedia authoring differs from typical interactive storytelling approaches in two important ways. First, the community content that we consider is not professionally produced and annotated. While we provide a reasonable degree of person and object recognition, the poor lighting and overall moderate quality of the content often requires user intervention to classify and locate content fragments. A second difference is that, although we focus on storytelling, we explore this concept in the context of repositories of UGC (User-Generated Content). There is still a structured representation of an overall interactive story space, but there is no control over the way the content is captured. The content
structures that can be made and exploited are only those emerging from the structure of the covered event itself.

1.4.3 Community Video Mashups and Content Repurposing

A second thread of more general story development is represented by work on video mashups and content repurposing. In this respect, it is interesting to note the current shift from local-based home videos management systems [28][65] to global-based video sharing Internet services.

Recent works [39][44] describe frameworks to synchronize and organize user-contributed content from live music events, creating an improved representation of the event that builds on the automatic content match. Shrestha et al. [59] report on an application for creating mashup videos from YouTube recordings of concerts. They present a number of content management mechanisms (e.g., temporal alignment and content quality assessment) that then are used for creating a multi-camera mashup. Saini et al. [53] go a step further by incorporating history-based diversity measurement, state-based video editing rules, and view quality in automated video mashup generations. Naci and Hanjalic [71] report on a video interaction environment for browsing records in music concerts, in which the underlying automatic analyzer extracts the instrumental solos and applause sections in the concert videos, and also the level of excitement along the performances.

Lately, crowdsourcing has been proven to be a good ally for content analysis. For example, fans of a band can be useful for improving content retrieval mechanisms, where a video search engine allows for user-provided feedback to improve, extend, and share, automatically detected results in concepts from video footage recorded during a rock n’ roll festival [11]. Our work builds on previous findings in event modeling [74] and identification [30][31], and video abstraction/summarization [7]. The main difference lays on the fact that we do not aim at providing a complete description of the shared event, but a better understanding of how community media can serve individual needs. Other interesting works propose a community video remixing tool [14], a video repurposing tool [66] and a video enrichment system that enable reciprocity [56]. In this direction, we should mention current practices around news stories, where users can reuse fragments of video clips for expressing opinions [46].

When compared with all these approaches, socially-aware multimedia authoring intends to help end-users generate stories in which social bonds between people play a major role. The previous approaches did not take into consideration
the case in which video authors and the people depicted in the videos are closely related. Similar to us, recent work has proposed a media sharing application that takes into account the interpersonal ties. This tool is capable of producing audio-visual media shows based on events, people, locations, and time [75]. In comparison to our work, this application does not allow for the creation of a narrative-based story based on multi-camera community recordings.