IRIS AperTO

UNIVERSITÀ DEGLI STUDI DI TORINO

AperTO - Archivio Istituzionale Open Access dell'Università di Torino

The art of video MashUp: supporting creative users with an innovative and smart application

This is the author's manuscript

Original Citation:

Availability:

This version is available http://hdl.handle.net/2318/69252 since 2019-12-13T15:33:55Z

Published version:

DOI:10.1007/s11042-009-0449-7

Terms of use:

Open Access

Anyone can freely access the full text of works made available as "Open Access". Works made available under a Creative Commons license can be used according to the terms and conditions of said license. Use of all other works requires consent of the right holder (author or publisher) if not exempted from copyright protection by the applicable law.

(Article begins on next page)

24 April 2024

IRIS AperTO

UNIVERSITÀ DEGLI STUDI DI TORINO

This is the author's final version of the contribution published as:

D. Cardillo; A. Rapp; S. Benini; L. Console; R. Simeoni; E. Guercio; R. Leonardi. The art of video MashUp: supporting creative users with an innovative and smart application. MULTIMEDIA TOOLS AND APPLICATIONS. 53 pp: 1-23.
DOI: 10.1007/s11042-009-0449-7

The publisher's version is available at:
http://www.springerlink.com/index/pdf/10.1007/s11042-009-0449-7

When citing, please refer to the published version.

Link to this full text:
http://hdl.handle.net/2318/69252

This full text was downloaded from iris - AperTO: https://iris.unito.it/

iris - AperTO

University of Turin's Institutional Research Information System and Open Access Institutional Repository

AUTHOR'S PROOF
JrnlID 11042_ArtID 449_Proof# 1 - 07/01/2010

Multimed Tools Appl
DOI 10.1007/s11042-009-0449-7

The art of video MashUp: supporting creative users with an innovative and smart application

Daniela Cardillo · Amon Rapp · Sergio Benini · Luca Console · Rossana Simeoni · Elena Guercio · Riccardo Leonardi

© Springer Science+Business Media, LLC 2010

Abstract In this paper, we describe the development of a new and innovative tool of video mashup. This application is an easy to use tool of video editing integrated in a cross-media platform; it works taking the information from a repository of videos and puts into action a process of semi-automatic editing supporting users in the production of video mashup. Doing so it gives vent to their creative side without them being forced to learn how to use a complicated and unlikely new technology. The users will be further helped in building their own editing by the intelligent system working behind the tool: it combines semantic annotation (tags and comments by users), low level features (gradient of color, texture and movements) and high level features (general data distinguishing a movie: actors, director,

D. Cardillo (✉) · L. Console
Department of Computer Science, University of Turin, Turin, Italy
e-mail: cardillo@di.unito.it

L. Console
e-mail: lconsole@di.unito.it

A. Rapp · R. Simeoni · E. Guercio
Telecom Italia, University of Turin "Progetto Lagrange—Fondazione C.R.T.", Turin, Italy

A. Rapp
e-mail: amon.rapp@guest.telecomitalia.it

R. Simeoni
e-mail: rossana.simeoni@telecomitalia.it

E. Guercio
e-mail: elena.guercio@telecomitalia.it

S. Benini · R. Leonardi
University of Brescia, Brescia, Italy

S. Benini
e-mail: sergio.benini@ing.unibs.it

R. Leonardi
e-mail: riccardo.leonardi@ing.unibs.it

UNCORRECTED PROOF

AUTHOR'S PROOF JrnlID 11042_ArtID 449_Proof# 1 - 07/01/2010

year of production, etc.) to furnish a pre-elaborated editing users can modify in a very simple way.

Keywords Video MashUp · Easy-to-use and intuitive interface · Inexpert users · Intelligent system

1 Introduction

In the last two decades a new era started for internet with the birth of Web 2.0, in the same way a really important revolution is happening for the web services and tools. This can be considered as the equivalent of the birth of Web 2.0 that is to furnish the chance to each user of tailoring the selection of resources, the presentation of contents and the navigation structure to his own specific model and context of usage [10]. At the same time it can be seen a convergence process between media changing the current entertainment landscape. The convergence consists in the fusion of old and new kinds of media into a unique setting of multifaceted digital technologies giving the birth to a cross media framework [27] whose audiovisual contents go more and more through different media. On a parallel path we can see that even users are changing their own entertainment experience: creative users are no longer satisfied with a passive use of the contents offered by the traditional media and try to gain something more interesting from the remixing and mutual exchange of audiovisual material. The contents already existing are taken from many different sources and remixed at different levels—audio, video and both at the same time—in order to create new kinds of cultural and artistic products: these new contents can be defined as video mashup. A video mashup is an audiovisual content obtained through the editing of pre-existing material which has been remixed in a new creative product conveying new meanings. They are the perfect reply to creative demands but, at the same time, to communicative and exchange needs, rather than of social critique, of information and education. The illustrated landscape is the starting point for the development of our application with the aim of easing and induce users to create video mashup.

In this paper we describe an innovative method for the video clip retrieval and the creation of automatic editing supporting a recreational activity like video mashup. Using low level features (LLF), high level features (structured following index terms or descriptors of a specific ontology regarding the cinematic world) and free tags written in by the users, the method has been conceived as the working process of a new tool of video mashup. It is different from all the other tool found in the actual state of art—and it will be deeply explained in the next sections—because it allows:

(a) to explore in an innovative way a repository containing motion-picture clips;
(b) to create automatically editing which simplify the process of creative mash up.

The innovative way mentioned before concerns the knowledge part of the repository, structured following an ontology which mirrors the personal cinematic world of the audience. In this way the user is no more forced to search for contents through other means with traditional searching methods (e. g. Google, Yahoo). Through this behaviour the user avoids both a considerable waste of time and a partial dissatisfaction of the results obtained (or not obtained fully) through criteria different from the cinematic rules. Apart from this advantage, the method provides the user with an automatic editing process which helps him in his creative task giving him the contents with a semantic coherence and a stylistic homogeneity.

UNCORRECTED PROOF

AUTHOR'S PROOF
JrnlID 11042_ArtID 449_Proof# 1 - 07/01/2010

2 Related works

As the video mashup and video sharing become more and more widespread the same happens with the possible applicative solutions in order to support users during the process of video editing. Apart from the professional tools on the market like Adobe Premiere [http://www.adobe.com/products/premiere/], Apple's Final Cut Pro [http://www.apple.com/it/finalcutstudio/finalcutpro/] and simplest tools for amateur users like Apple's iMovie [http://www.apple.com/ilife/imovie/] and Windows MovieMaker [http://www.microsoft.com/italy/windows/products/windowsxp/winxp/using/moviemaker/default.mspx], during the last 2 years lots of online tools proliferated. These were characterized mostly by a peculiar simplification of the functionalities offered (i. e. JayCut [http://jaycut.com], JumpCut [http://www.jumpcut.com]). These are tools with the video editing as a main and general functionality, while examples like Cuts [http://www.cuts.com/] and Sweeney Todd Trailer Editor powered by GorillaSpot [http://www.sweeneytoddmovie.com/mashup/editor.php] have been created specifically for the video mashup and offering the chance to remix already existing contents. These examples give the user the chance to mix clips coming from movies in order to be handled, modified and remixed by users. Nonetheless, all the examples mentioned do offer neither an appropriate quantity of contents in the repository nor functionalities of automatic video editing easing in some way the job of amateur users. Research studies focused mainly on this aspect trying to supply users with an automatic or semi-automatic process of editing in order to help them in realizing their own creation. Most of these researches developed application supporting video editing of amateur contents coming from video recording of the users themselves. A first group of applications selects through the contents recorded by the user only those with a high level of quality through which can be realized a product of video editing. For example, Hitchcock [22] detects automatically suitable clips in the raw video material and presents them to the user for selection and adjustment. Using the automatic analysis of the video combined with standard editing rules, Hitchcock deletes all the shooting[1] judged not suitable to be inserted into the video editing because of the low quality of the movements of the camera, e.g. fast pan, slow zoom. Another method [29] based on automatically extracted metadata divides automatically the raw video material into useful section and useless section (such as "hand shake" and "failure camerawork"), to extract appropriate shots and present them to the users. LazyCut [24] tries to help the user providing a semi-automatic video authoring, based on context aware authoring templates; another series of applications automatizes completely the video editing process adding a soundtrack accompanying the content considered more suitable [25, 34]. While the methods illustrated up to now use exclusively the automatic analysis of the video and the automatic extraction of the features to realize the automatic editing, other researches try to exploit various metadata to ease the task of video editing: using the audio transcript, generated from closed-captioning information and speech recognition, to provide multiple points of view of a video with different semantic content and at different levels of abstraction (including storyboard, editable transcript, etc.) [13]; using the semantic annotation and rhetoric-based method to support the selection and automatic editing of user-requested content from video footage [8]; using the automatic analysis of video and audio content joined with the text analysis of the commentaries in order to divide in a correct way the boundaries of the meaningful scenes (for example scoring a goal in a football match) during sports events in order to insert them in an automatic summary [44]. Differently from the tools illustrated up to now, our tool mixes together automatic extraction of features coming from the automatic

[1] A shooting is an audiovisual segment of changeable duration

UNCORRECTED PROOF

AUTHOR'S PROOF JrnlID 11042_ArtID 449_Proof# 1 - 07/01/2010

analysis and semantic annotations. This process is realized in order to give the chance of an original research through a huge amount of contents to be used for the video mashup and for creating automatically a draft of editing coherent and stylistic homogeneous.

2.1 Preliminary studies

2.1.1 Analysis of current video Mashup

The video mashup is now a phenomenon more and more spread in the web: professional-produced audiovisual contents are remixed to create new contents that enter in the entertainment landscape creating a bidirectional cross-media experience [4]. In this context the author becomes a bricoleur [30], able to remix in a creative way pre-existing contents. Nevertheless the actions linked to the video mashup are numerous and with different demands as far as the communication is concerned. For this reason we can find very differentiated products which can be classified in sub-categories. In order to analyze critically and deeply this phenomenon, earning useful information about the design of a new tool, we analyzed all the different kinds of video mashup available in internet. Studying the classifications already existing in literature [16, 20] which worked as a starting point, a new typology comes out: in this each detected type has been deeply defined in its semantic, syntactic and communicative characteristics. This classification bursts out more than one method already employed in researches like ours [18, 37]: a qualitative analysis of the documents and a two-months observation have been exploited together. Two video portals have been analyzed. The first is TotalRecut [http://www.totalrecut.com/] having as an explicit aim that of collecting video mashup, while the other YouTube [http://www.youtube.com] is opened to accept each kind of audiovisual content. During the observation more than 600 video mashup chosen between the most popular have been analyzed and the comments applied to them monitored. Some of the fixed criterions, drew from the existing literature about the semiotic of the audiovisual and cinema (e. g. [9, 14, 15, 31, 35, 41]), use in the analysis are:

- **Communicative aims**: convey the purpose the user would like to communicate to his audience concerning the video products he used (Parodying, Celebrative, Critical, etc.)
- **Stylistic aspects of the video**:
  - Type and rhythm of the editing: it is useful to explicit the prevailing type of editing used inside the mashup (parallel, etc.)
  - Toning effects and/or titles: it points at the application of a toning effect or at user-edited text frames inside the video
  - Use of photos, still footage and slow-motion: it shows video effects applied to slow down the flow of the video
- **Stylistic aspects of the audio**:
  - Change of soundtrack and/or dubbing: it shows an action which synchronizes a new audio file to the video file already chosen
  - Insertion voice over: it shows the insertion of a narrative voice which works as a comment on the showed images
- **Semantic aspects:**
  - Kuleshov effect: it shows that some of the segments has been reused in a new and different context that changes its original meaning [28, 37]

UNCORRECTED PROOF

AUTHOR'S PROOF

- Change in the role of actors and in the genre: it shows a change in the role of actors and a change in the genre of the original video from which come the clips used for the mashup
- Narrative dimension, coherence and semantic cohesiveness: it shows a narrative or declarative development realized through the characters or discussions focusing on a detailed topic

This classification does not aim to tell about the mutual exclusivity of the partitions, but it has rather been driven by a practice criterion focusing on peculiarities and recurrences in the analyzed videos. The genres identified go:

- from movie recuts (the material coming from movies and serial TV is reused and manipulated with recreational aims)
- to celebratory mashup or tributes (they represent a tribute to a particular character, genre, cultural product)
- to political and critical recuts (they mix together material with the main aim to debunking or strongly criticize the contemporary society)
- to music video recuts (they have a prevailing audio component which goes together with the video one)
- to playlists (they point at letting the user know the author's tastes) and so on.

Each genre has thus been subdivided in sub-genres and for each of these there is a detailed description of semantic, stylistic and communicative characteristics: for example movie recuts have been divided in scene recuts (a single scene of a particular movie is re-edited), trailer recuts (a new trailer from an existing movie is created living a totally different meaning to the original movie), alternate endings (proposing alternative ends for a movie or a serial TV), and fan film recuts (short films created mixing together products coming form different movies/serial TV). The following brief summary illustrates synthetically the emerging characteristics of some genres (Table 1):

From this analysis and from the concerning classification it comes out a high level of differentiation of the actions commonly called mashup. The users express their creativity with lots of different communicative aims focusing their attention on the products they use, carefully chosen for specific communicative aims. Thus a tool which supports this kind of activities should give, other than the functionalities needed for video editing, even the chance to search in an intuitive manner in order to find easily the scenes a user could be willing of using. In order to express his own creativity using already existing products the

UNCORRECTED PROOF

t1.1 **Table 1** Example of classification of mashup

t1.2 Kind of Mashup	Communicative aim	Stylistic aspects	Semantic aspects
t1.3 Movie Recuts	Creative, Narrative, parodying	Voice over, Title, audio change, new editing rhythm	Narrative dimension, change of genre and relationships, Kuleshov effect
t1.4 Tributes	Celebrative	New soundtrack, title, shot and slow-motion	A-temporal representation, no narrative dimension, no cohesion or cohesiveness
t1.5 Political and critical recuts	Critical, parodying, educational	Change of dubbing and soundtrack	Kuleshov, effect, coherence with the treated topic
t1.6 Music Video Recuts	Celebrative, aesthetic	Mix audio, video editing created on the audio rhythm	No narrative dimension, pure rythmic representation

AUTHOR'S PROOF JrnlID 11042_ArtID 449_Proof# 1 - 07/01/2010

user does not need a video editing tool with extremely advanced functionalities (creativity comes out above all from the choice and combination of meaningful clips, rather than applying advanced functionalities to manipulate the images). The main aim is the recreational one directed to a rapid understanding and to the re-elaboration of the initial products. These results were considered in the definition of the guidelines for the concept design drawing our attention on the different and peculiar aims and needs that the users have during the process of creation of a video mashup and on the amusement aspects that this process involves.

2.1.2 Analysis of current tools of video editing

In parallel with this research we developed an heuristic comparative analysis of five video editing tools for non professional users. We took into account MovieMaker and iMovie which are the most popular and some of the most recent tool web based (Jumpcut [http://www.jumpcut.com], Sweeney Todd Trailer Editor powered by GorillaSpot [http://www.sweeneytoddmovie.com/mashup/editor.php], Jaycut [http://jaycut.com]). Through the analysis we tried to find out the standard de facto and the trends in design exploiting an exam of the main aspects concerning the interface and specifics of functionalities. We focused mainly on the standards de facto expressing design solutions which are the most strengthened during this period. Considering that standards are related in some way with the users expectations in the use of video editing tools, they have to be considered having an active part in the design process of a new concept. Based on the results some solutions can be considered as standards because they are applied in almost all of the analyzed systems: for example the Drag & Drop use for the direct manipulation of the video, a TimeLine visualized to edit simple videos (start and end of the clip, title, transitions). New trends come out as well even if they are not completely steady and uniform between the analyzed tools; they show a trend in the design of the concept and interfaces. First of all we detected a lean recourse to textual menus in behalf of a different pervasiveness of graphical components (icons). Secondly, the most recent tools which are the web based characterize themselves with specific functionalities. These make them different from the other tools and give the user instruments useful to satisfy peculiar needs, mainly recreational, of the users. JayCut is simple and rapid in editing and exporting the created video, Sweeney Todd Trailer Editor emphasizes the content to be re-elaborated, JumpCut gives the chance to remix contents produced by other users. According to the results coming from the analysis of the video editing tools available today, some other issues and guidelines for designing new future tool interface were found out. In particular the results of this comparative analysis drawn our attention on the essential features that have to be included in our application and emphasized the prominent role that has the visual and graphic components in the current design trends.

In the following table we illustrated briefly the components present in each of the analyzed tools of video editing. After the analysis of the functionalities available for the users the table allows to compare at a glance the modules working at the back-end of the different tools (Table 2).

2.2 Guidelines and concept design

Preliminary works allow us to point out some guidelines which drove us during the elaboration and the definition of the concept. Some of them follow:

Simplification in the way of retrieval material to be re-elaborated it should be easy for the user to find simply and rapidly what he needs mixing together the research with the editing phase in order to create only a unique flow of experience.

UNCORRECTED PROOF

AUTHOR'S PROOF
JrnlID 11042_ArtID 449_Proof# 1 - 07/01/2010

t2.1 **Table 2** Comparison of the modules in each of the analyzed tools

t2.2		Database	Ontology	Automatic Editing	Rule Engine
t2.3	Video Mashup Tool	Huge and structured DB	Ontology of the cinematic world	Automatic editing with semantic and stylistic homogeneity	Rule Engine for retrieval based on LLF and HLF
t2.4	MuveeMix	No Db	No ontology	Automatic editing with stylistic homogeneity	No Rule Engine
t2.5	JumpCut	No Db	No ontology	No automatic editing	No Rule Engine
t2.6	SweeneyTodd	Small and not structured Db	No ontology	No automatic editing	No Rule Engine
t2.7	Hitchcock	No Db	No Ontology	Automatic editing with stylistic homogeneity	Rule engine based on LLF

Emphasis on the amusement and the serendipity it should be favoured the retrieval of unexpected videos so that the creativity of the user can be stimulated through new ways of finding the videos. These manners should be able to take advantage from the passions of the users and to characterize the search as a recreational activity.

Versatility and characterization it is strongly recommended to back users' creativity in its numerous appearances and for the various communicative aims exploiting. at the same time, peculiar characteristics and functionalities exclusive of the tool compared to others already available. The new tool has to satisfy the different user needs emerged from the analysis of the video mashups, giving at each user the possibility to express his peculiar aims in order to create a specific kind of video mashup. Besides the tool has to develop a set of functionalities that are immediately recognizable by the users as original and innovative.

Reduction of control (set filters) favouring easiness of use a strongest easiness in managing the editing functionalities allows users to focus mainly on communicating different and personal meanings rather than on the technical perfection of the editing work. Filters set in advance to automate editing operations allows even non expert users to create videos with new meanings without being discouraged by the huge number of functionalities and rules to be settled.

Improvement of visual and graphic components to the detriment of textual menus and direct manipulation of the elements (drag and drop) the design of the interface must strongly reduce textual controls preferring a direct and intuitive management of the videos which should be re-elaborated.

Starting from these guidelines, we defined a design concept for an online tool of video mashup. The tool is a web application that is conceived as a new feature for an innovative paradigm of cross-media Inter-tainment experience [39]. Exploiting video editing with already existing contents the new paradigm will favour the improvement of available contents and their exchange between different kinds of media. The application we illustrated is not created to work as a stand alone but rather as a web application linked to a project for a new interactive Tv (DynamicTv [38]). This integration has been thought as a part of a virtuous circle giving from one side the chance to access to a repository with a large variety of contents and from the other side the chance to make public the mashup on TV: in this way users should be encouraged to realize their creative contents. Offering the

UNCORRECTED PROOF

AUTHOR'S PROOF JrnlID 11042_ArtID 449_Proof# 1 - 07/01/2010

users the chance to participate at a unique entertainment experience of creation should increase the quantity of contents available to the TV audience, improving the exchange of audiovisual contents through different media. User can navigate and interrogate the repository choosing three different variables: the celebrity, the topos and the "stilema". Celebrities represent the motion-picture or TV-serial actors particularly the most famous and recognized as important figures of the cinematic world. The topos (*antique Greek* 'topos') represent the narrative places that is cyclic themes universally recognized as belonging to a well defined genre (for more details see *Ontology of the cinematic world* section), like for example the gunfight, the attack to the stage coach and the robbery in the western genre. The "stilema" represent predetermined visual styles established analyzing the different styles occurring in the history of cinema or directly linked to the genre culture. Users will be allowed to choose up to two celebrities, a topos and a "stilema" in the same query in order to receive as a result a composite collection of clips representing celebrities inserted into a typical narrative context having an homogeneous visual style. The tool is not only a means to have a retrieval of the requested clips as a traditional search engine, but it returns as a result of the query an automatic editing exploiting metadata (high and low level features) and tags. At the end of the process the user will have a collection of clips provided with coherence and homogeneity in its style. Once the requested clips have been found users can decide how to change them in accordance with their preferences (they can change the order the clips are inserted in the editing, change the start and end points, modify the audio properties of each clip, etc.). Moreover we specifically designed a concept suited for the cinematic and television fans because they are particularly predisposed to become participatory audience [3, 21].

2.3 Architecture of the tool

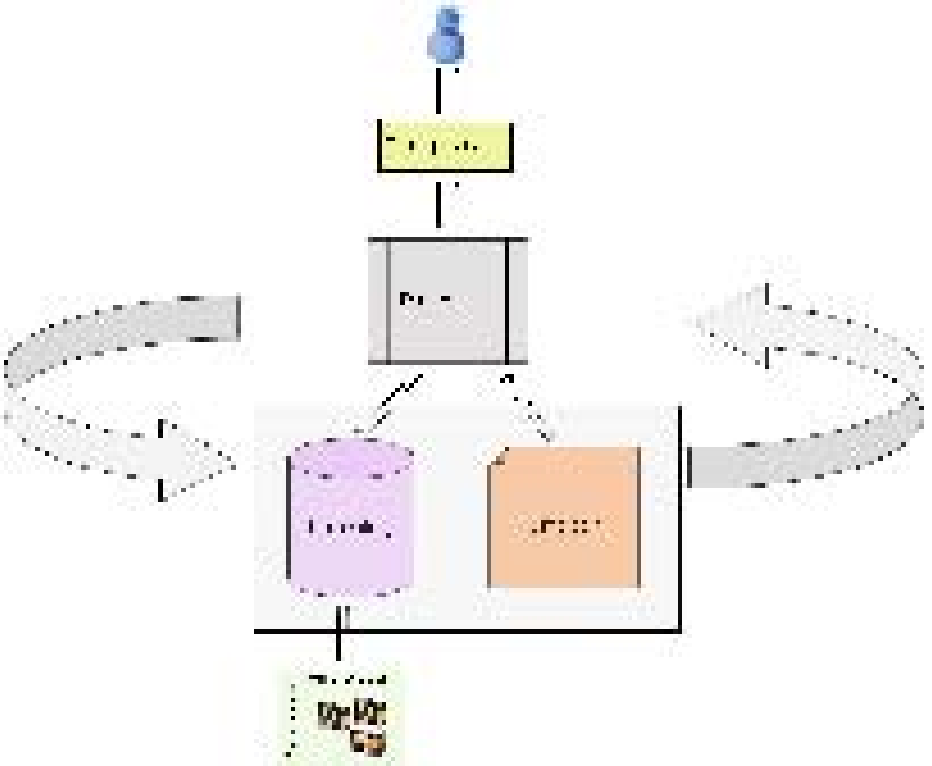
In this section we present an architecture for supporting the share and remix of video content, discussed in the previous section. This architecture (Fig. 1) represents a conceptual model which describes the main modules cooperating within the MashUp framework, together with the communication channels between each module. In the following sections it will be first provided a brief description of the phase of pre-processing the clips before they are placed into the repository and then it will be depicted an overview about the functionalities and the main implementation details of the system modules.

2.3.1 Video database and metadata generation

In order to allow a correct working of each module illustrated in the next section, it is strongly necessary a pre-processing phase on the videos inserted into the repository. Before we go into details of the discussion, it will be beneficial to introduce some important terms related to videos. Motion picture specialists perceptually segment a video into a hierarchy of partitions, as the one shown in Fig. 2, where the level of semantic increases from bottom to top. Accordingly to this decomposition, on a first level a video can be completely and disjointly segmented into a sequence of *scenes*, where each scene depicts and conveys a high-level concept or a story (for this reason scenes are also referred to as Logical Story Units, LSU [23]). Actually, the concept of scene is much older than motion pictures, because it ultimately originates in Greek theater. However, while a scene is traditionally a continuous sequence that is temporally and spatially cohesive in the real world, on the contrary (as noted in [17]) it is not necessarily cohesive in the projection of the real world on videos.

UNCORRECTED PROOF

AUTHOR'S PROOF
JrnlID 11042_ArtID 449_Proof# 1 - 07/01/2010



Q5

Fig. 1 Abstract model of the architecture of the MashUp Tool

On a lower level, scenes can subsequently be segmented into a sequence of basic video segments named *shots*. While scene did exist before video, on the other hand shots originate with the invention of motion cameras and are considered to be the longest continuous frame sequences that come from a single camera take (i.e., what the camera images in an uninterrupted run [17]). Shots sharing common perceptual low-level

UNCORR

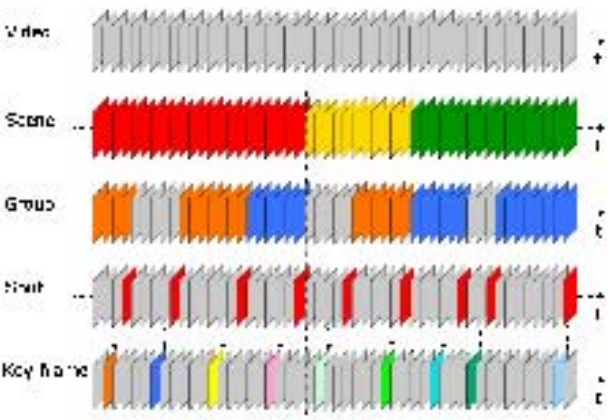


Fig. 2 A hierarchical decomposition for video

AUTHOR'S PROOF JrnlID 11042_ArtID 449_Proof# 1 - 07/01/2010

characteristics can then be clustered together into higher entities called *groups* (or *clusters)* of shots (see Fig. 2). Examples of groups are visually similar shots, or video segments sharing the same camera motion. Finally, on the bottom level of the hierarchy, one or more *key-frames* can be extracted from shots as static significant paradigms of the shot visual-content. The video repository used for video mash-up contains video material segmented at different levels of the hierarchy and the related metadata (that are HLF = High-level Features, MLF = Mid-level features, LLF = Low-level features, MA = Manual annotations). In particular, in order to create new content, the following entities can be retrieved from the repository:

- Full length videos;
- Clips of any length uploaded by users;
- Logical Story Units;
- Shots and groups of similar shots.

Full length videos When full length videos[2] are inserted in the repository, they are enriched by HLF that are metadata [32, 33] concerning title, genre, director, actors, year of production, origin country, etc. This kind of data are the typical ones used to describe complete movies in the movie databases like for example IMDB [http://www.imdb.com/ last access April 2nd, 2009]. They allow a structured knowledge of the domain because these data contribute in building the knowledge base offering a complete overview of the cinematic world. Q6

Clips of any length uploaded by users Since clips[3] may be directly uploaded by each user, these can be enriched by Manual Annotation (MA) in the form of semantic tags added by the user at the moment of the upload.

These two kinds of contents are both involved in the pre-processing phase through which can be segmented into LSU, shots and groups of shots.

Logical story units Logical Story Units are automatically segmented as sequences of contiguous and interconnected shots sharing a common semantic thread. These units constitute narrative scenes autonomous in their meaning and they are extracted as explained in [7]. Besides HLF derived (automatically inherited) from the full length videos metadata (of which are part), or from metadata structured on the basis of the ontology (see Ontology of the cinematic world section) and inserted in a subsequent period of time by the editing office, LSU can be enriched by some MLF such as duration, scene pace (e.g. to distinguish between fast scenes in action movies and slow scenes in dramas), shot-transition pattern and scene entropy for fast dialogue indexing [6].

Shots and groups of shots In order to extract the shots from the whole video it is necessary to detect all possible types of transition between two adjacent shots. These can be either cuts, dissolves, fades or wipes. The algorithm used to detect shot transitions is based on the classical twin comparison method, where the error signal used to detect transitions is based

[2] A full length video is a full motion picture inserted in the repository by an editing office.

[3] Clip can be considered as video units that is audiovisual segments of different duration extracted from a full length video, automatically or manually. Consequently, a clip can be a content uploaded by users, but even Logical Story Units or single shots automatically extracted.

UNCORRECTED PROOF

AUTHOR'S PROOF
JrnlID 11042_ArtID 449_Proof# 1 - 07/01/2010

on statistical modeling. The distance between two frame is estimated by evaluating the difference of the corresponding color intensity distributions. In order preserve also spatial color information, each frame is partitioned into rectangular regions and from each of them a color histogram is then extracted. The actual number of regions which provides the best trade-off between miss-detection (caused by the use of global histograms) and the false-detection (caused by the global and local motion when small regions are used), generally depends from the video spatial resolution. In general, in order to ensure an adequate training of the histogram we adopted rectangular regions with a number of pixels at least ten times bigger than the possible values of the histogram. Assuming that the contents of two consecutive shots can be represented by two independent random variables, an abrupt transition is modeled as a drastic variation in the color density function of adjacent frames while dissolves are detected evaluating the difference between the color density function of the actual frame and the one predicted by the dissolve model described in [1].

After being extracted, each shot is then described in terms of Low-Level Features (LLF) which are physical characteristics related to color, motion and audio that can be directly extracted from the video stream. In particular the investigated features are:

- Color features
- Vector Quantization codebooks extracted as described in [5] which constitutes an efficient index of the color structure of shot key-frames;
- Dominant Color descriptor, as defined in MPEG-7 standard, which specifies the most representative set of colors in the shot key-frames;
- Color Layout descriptor, as defined in MPEG-7 standard, which specifies the spatial distribution of colors in the shot key-frames.
- Motion features
- Motion Activity descriptor, as defined in MPEG-7 standard, which describes the standard deviation of the module of motion vectors in the compressed stream (Jeannin 2001);
- α-ratio descriptor, which measures the intensity of human-perceived motion by computing the ratio between the non-motion compensated macro-blocks against the total number of macro-blocks;
- Motion Vector Directions, which measures the amount of motion along the principal directions of motion (up, down, left, right);
- Motion Activity Map, which measures the amount of motion and its spatial distribution along the duration of the shot [58].
- Audio features
- Time-domain features (Zero Crossing Rate (ZCR), etc.);
- Frequency-domain features (spectrogram, etc.).

When all the LLF have been extracted they are returned in an XML file, which is stored and related to the video segment. Based of the LLF characterization of shots, it is possible to group similar shots by employing traditional clustering algorithms, to detect LSU boundaries, to produce video summaries as in [5] and to compare video segment similarities within the database. This helps in retrieving similar shots up to specific user's requests, expressed in the form of query-by-example, thus providing the user with similar material for the composition of new video content. For example, motion descriptors have been already proved of being successful in video browsing applications by filtering out all high or low action shots, as in [26]. In conclusion, rather than having only high level metadata

UNCORRECTED PROOF

AUTHOR'S PROOF
JrnlID 11042_ArtID 449_Proof# 1 - 07/01/2010

the contents of the repository can be enriched with useful mid and low level metadata able to substantially contribute to the retrieval process described in the next section.

Increasing the quantity of metadata in the repository As described above clips may be gained either from the long video and so inherit their high level metadata or they can be directly uploaded by each user. In the first case we are sure that clips (LSU, shots or group of shots), derived by full length videos, bring with them all the high level metadata characterizing the long video; most of these segments are manually annotated too by the editing office with metadata structured as the ontology shows in the next section (see *Ontology of the cinematic world* section). In the second we cannot be sure about the data describing clips, so it could be possible that a clip has not information of high level but it is plainly described through the tag the user wrote uploading it.

In order to enrich the information about each single clip it has been created a Java module working as described in the next paragraphs. Each one of the topos is characterized for being further articulated into a subdivision in places, figures and themes of the topos itself (see *Ontology of the cinematic world* section); these are ordered in a table of the repository so that each clip can be referred specifically to one of them.

We need to use this table in order to find all the words correlated with the terms saved into it. With the term 'words correlated' we mean synonyms, hyperonyms, hyponyms that is all the words standing on the same hierarchical level, one level above and one below in a hypothetical hierarchical tree having as a component the key term we take in consideration that is not the root of the tree itself.

The result of this process is another table containing all the terms divided per topos and attribute of the topos, i.e. place, figure and theme, which once created will be a permanent component of the repository. When the user upload his own clip, this module starts a sort of exploration through the data of the clip; if these are complete no actions will be done, on the opposite if the field of topos is empty the module starts its work. It explores the tag of the clip, reads them and compares them with the correlated terms saved into the other table, in case there is a direct relation between two terms or more, the corresponding topos term will be learned and uploaded into the clip table so that an information is added and the information regarding the clip is complete. Let us explain the whole process with an example: the user upload a clip tagging it with the words 'impenetrable equatorial forest' and 'secret watcher' but without adding any other metadata regarding the topos and its sub properties. Once the Java module starts to work it finds empty cells in the topos table and goes to read the tag box comparing the terms found in it with the correlated words mentioned before. The two sentences the user inserted can be respectively read as paraphrases of 'jungle' and 'spy'. Once an analogy is found the root word (that is the topos to which belong the correlated words taken into consideration) this is added to the metadata: in our case the 'place = jungle' and the 'figure = spy' are included in the classification of the topos "Adventure".

This module has been created through java and the use of an multi-language dictionary, MultiWordNet [http://multiwordnet.itc.it/english/home.php]. The cited website gives the opportunity to the user to process words in different languages while the Java Api passed through a license are created only for the Italian language. For saving time in processing and storing the information it has been thought that it would be better to create once this process of enriching the metadata of the clip. For this reason the module works creating at the beginning and only once the new table with the topos and their correlated terms which will be saved definitely as a table of the repository and re-used each time it will be necessary.

UNCORRECTED PROOF

AUTHOR'S PROOF
JrnlID 11042_ArtID 449_Proof# 1 - 07/01/2010

Q2 *2.3.2 System modules*

In this section we describe the function of the system modules stressing in particular the content adaptation which represents the most relevant "intelligent" behavior of the system.

Dialog manager The dialog manager is the module in charge of establishing an interaction between the system and the user. It is implemented as a Java Servlet receiving users' requests and replies to them. It grants a complete and secure monitoring of data exchange due to its being the unique I/O point. All he actions made by each single user are saved by the dialog manager into a log file accessed by the user modeling manager. Besides this primary task this module is able to assign different requests to the appropriate module; the requests are tailored on the basis of the user's demands and the output is forwarded to the user through the dialog manager.

Repository All the clips and full length videos are contained in this repository. Each one of the clips is identified with a unique ID apart from metadata both of low and high level. This kind of data are assigned to the clip during a phase of pre-processing which are illustrated in the previous section.

Clip can be considered as the minimum video unit that is a brief audiovisual segment extracted from a longer movie sequence. This latter contains a series of metadata which will be inherited by all the clips coming out from it [32, 33].

User modeling module A user model is a knowledge structure that represents the profile of a single (registered) user. The dimensions in which the user description is structured refer to concepts defined in the User Model taxonomy. This taxonomy defines the concepts needed to describe user profiles. The user model in our application maintains personal information and a multi-dimensional representation of the user interests in each category of the domain ontology, by associating to each category an expressed value of interest. Notice that the model is initialized to a uniform set of values representing the same interest value for each domain category. Such values will be modified in the further update phases.

The user modeling manager initializes the user model and updates it.

The data collected in the UM taxonomy are features derived from the registration form (e.g., age, job, gender, interest in the different genres, etc.). The User Modeling module represents the interface between the system and the knowledge about the users. This module is in charge of managing both the update of the UM (in case the user adds new useful information) as well as any request of user data supplied by other modules. Each time a module needs to get any information about the user, it asks the user modeling manager.

Ontology of the cinematic world In order to manage in a better way the concepts of heredity and the relationship between the attributes of the clips it is appropriate to describe in a semantic way the cinematic world [36]. It has been formally structured in a conceptual model, an ontology which allows reasoning and expanding the knowledge through rules of inference. The main concepts of the ontology are the celebrities, the topos and the editing styles or "stilema" with their properties and the related restrictions. The class containing the celebrities has been conceived to include those considered the most famous actors on the Italian and foreign stage. The ontology of the "stilema" identifies visual styles repeating through the history of the cinema and identified through three specific visual descriptors, that are the texture and the color composition, and the dynamism of the movements. The topos have been specified for each genre and characterized by three properties: places,

UNCORRECTED PROOF

AUTHOR'S PROOF
JrnlID 11042_ArtID 449_Proof# 1 - 07/01/2010

figures and themes. Places are the typical environments of a particular genre, figures are the recurring characters and themes are the narrative situations distinctive of a specific genre. The western genre has places—the saloon, the bank, the hotel, etc.—figures—the bandit, the Indian, the foreigner, etc.—and themes—the attack to the stage coach, the robbery to the bank, the fight, etc. The knowledge base has been structured through the analysis of the reference literature but also through the specific know-how of experts of cinema and strong cinema users. The ontology of the topos has been structured firstly following the reference literature (Brodwell 1994–1997–2003; [14, 19, 31, 35, 41, 42]) by a group of specialists. Starting from this research topos come out divided into ten macro genres of the cinematic field. In the next stage two focus groups took place to validate them: nine users specialists in the domain per each; they were selected through a screening questionnaire that investigated the frequency of movies' fruition and their knowledge of the history of the cinema. The two groups were separated by age (1st group from 18 to 35 y.o., the 2nd group from 36 to 65 y.o.). Some stimulus were presented to the group's participants: overall a series of representative images of every specific genre, after which, the users were invited to try to identify by themselves, characterizing places, figures and themes. After this "creative" session their topos were compared with the specialist's ones defining all together those to be taken as valid and those to be erased. These two phases have been repeated for each of the six macro genres that were analyzed. Basically, the main result is a substantial coincidence of the topos identified by the specialists and those emerged by the users that was true for both groups. Other specific sessions of focus group were managed to individuate the privileged celebrities to be included in the tool and the ontology of "stilema". Besides that other qualitative techniques with end users were used in combination, i.e. questionnaires to evaluate liking and stardom of celebrities, and vis à vis interviews with specialists of the domain. The results of these researches allowed to define the ontology structuring high level metadata for the clips in the repository. The ontology allows a sort of retrieval in the repository approaching categorization and division of contents distinctive of the human mind. In Fig. 3 a screenshot of a part of the ontology is shown: it goes down to the deepest level for two of the ten topos; it is not exhaustive of the whole but it gives a real idea of how the ontology is developed.

Q8

Rule engine The rule engine is the most innovative component of the system which works implementing the knowledge base through a set of rules that specify how to act on the assertion set [40]. The knowledge of a domain expert is encoded into the rule set consisting of little more than a set of *if-then* statements providing the basis for the so-called "expert-systems". It works applying specified rules which follow this schema

$$\begin{array}{c} \textit{if}\,(\textit{logical combination of user features/value pairs}) \\ \textit{then}\,(\textit{inferred movies features/value pairs}) \end{array}$$

to predetermined facts which have been arranged in advance and will be followed by the rules. The rules we planned to create are for example:

- *rules of selection*, processing the user's criteria of selection in an intelligent way in order to return only the clip really interesting for him following the preferences saved in the user's profile;
- *rules of priority*, returning the references to those clips belonging to the categories the user elected as his favorite.

A simple example explaining the way these rules work could use the first kind of rules; the situation could be as follows: the user chooses the values "actor = Johnny Depp", while

UNCORRECTED PROOF

AUTHOR'S PROOF
JrnlID 11042_ArtID 449_Proof# 1 - 07/01/2010

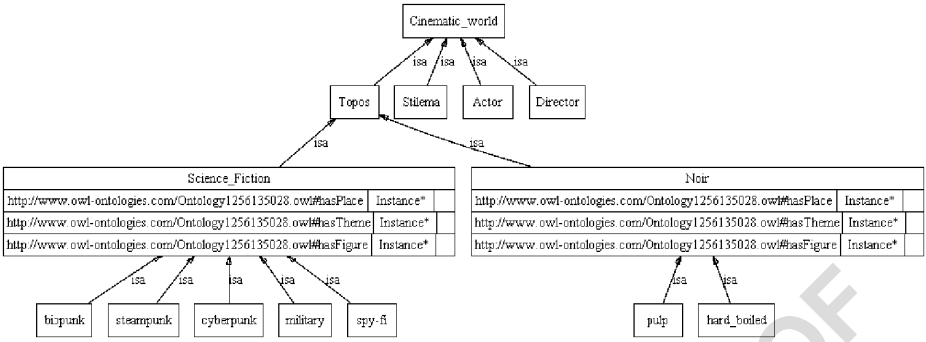


Fig. 3 Screenshot of a partition of the ontology

Q5

from her user profile it can be argued that she is a woman aged 20 or more. The applied rule would be structured as

$$\text{R1)} \qquad \begin{array}{l} \textit{if}\ (\text{gender} = \text{“female”} \wedge \textit{age} \geq 20) \\ \textit{then}(\text{genre} = \text{“sentimental comedy”}) \end{array}$$

returning the reference to a specific genre responding to the requisite inserted into the user profile.

Exploiting the preferences stored in the user profile we can propose a way in which the rule engine could work managing the knowledge base contained in the ontology.

Suppose S is the rule-based system supporting the user to deal with the choice of clips. Assume that knowledge is represented as a set of rules and that the knowledge base of S contains very simple rules such as:

$$\text{R2)} \qquad \begin{array}{l} \textit{if}\ (\text{age} \leq 18 \wedge \text{age} \geq 14 \wedge \text{gender} = \text{“}\textit{female}\text{”}) \\ \textit{then}(\text{genre} = \text{“}\textit{comedy}\text{”} \wedge \text{theme} = \text{“kiss”}) \end{array}$$

The user *Anne* formulates her query requesting clips giving exclusively the name of her favourite actor. Basing on the information stored in the user profile (i.e. Anne is a girl aged 16) and applying the rule *R2* the expert system returns the reference to those clips with the explicitly requested actor and the inferred features (i.e. *then* condition in the rule).

A simple and brief use case could explicitly explain the full process going from the query to the process of clips and from the cluster of similar clips to the automatic editing offered to the final user. Once the query is inserted through the interface it is conveyed from the Dialog Manager to the other components of the architecture. The query owns detailed requisites that is one or more celebrities, a topos and a “stilema”; in most of the cases in the repository there will be numerous clips responding to the requirements of the query. So it will be the rule engine to apply determined rules as illustrated above in order to select the most suitable clips for each case and user. In case the user is not logged in another process work behind the system: different algorithms work together apart from the rule engine and they establish different weights for each clip calculating this measure basing on the adherence of each clip to the query inserted by the user in order to give a reasoned answer to his query. During this process firstly the HLF are taken into consideration, after these are completely coincident with the requirements inserted through the query the process goes on at a lower level that is that of the LLF trying to find a full similarity in these metadata. ‘Full similarity’ here means not a complete sameness in the LLF of each clip but rather clips having high weighted LLF. The weight to which we refer can be based

UNCORRECTED PROOF

AUTHOR'S PROOF JrnlID 11042_ArtID 449_Proof# 1 - 07/01/2010

on each one of the features mentioned in the previous section, that is color, dynamism of the movement and kind of texture. Taking into high consideration these low level metadata clips are ordered following weights satisfying the requirements of the query.

This has been described as being the most original and worthy part of the whole tool, that is why it is in a full phase of development and deepening with the aim of creating complete and exhaustive algorithms covering all the possible exception of each query.

User interface The phase of designing the user interface had the difficult aim to create a first version of the MashUp Tool focusing above all on the extreme easiness of using it. The main point to take into primary consideration has been the segment of public who should become everyday users of the tool that is not expert users. The ideal person who approaches himself to the mashup is the one who likes or loves cinema and movies but considers the mashup a playing activity not requesting specific competences about more professional tools on the market and on the web.

So the main aim followed in designing the user interface has been that of respecting some simple rule of simplicity and easiness of use. It has been designed in flash but its back-end is managed in java through the JMF [http://java.sun.com/javase/technologies/desktop/media/ jmf/ 2.1.1/download.html]. The Java Media Framework is a Java API which allows developers processing media file and to handle them with operations like cut, copy, paste and so on.

Through the interface the user will be able to formulate his requests through the various panels and functionalities. The interface is mainly divided into three windows. The main window contains all the choice icons and areas through which the user gives his query to start the search and having a movie editing; this is a part of the tool which can be used by non-expert user and able to return a complete movie product without going too deeply into processing video and audio clips (Fig. 3). This window favors the visual communication rather than the textual one [43] identifying the possible choices of the user through icons. The user can insert his request selecting first one or two actors and then a topos, finally even a "stilema". The result will be seen in the big central panel of the interface through which other changes are possible. Clips can be modified and moved through the whole video tank to the Drag & Drop function. Selecting a clip pop up windows will offer the user the chance to apply advanced functionalities to modify audio and video of the clips (Fig. 4).

The second and the third windows contain respectively advanced audio and video controls to modify and handle the files. These have been created to allow expert user to employ the tool as well as the majority of professional software for processing video and audio files. The functionalities are, as mentioned above, cut, copy and paste both in the video and audio sections, the application of a determined editing style or effects of transition between different video clips, the change of the volume of numerous audio clips making it homogenous and so, more professional.

2.4 User study

In order to test the acceptability of the concept at the basis of the service, rather than the usefulness and the complete series of functionalities, two focus groups with 8 people per each have been organized. The 1st group was made of specialists in the use of video editing tools (professional, like Adobe Premiere and Apple's Final Cut Pro, and simplified, like Windows MovieMaker and iMovie). The 2nd group was made by experts of the cinematic world without any previous experience with video editing tools, but possibly interested in using video editing applications in the future. The collected data and the data analysis were exclusively qualitative.

The focus groups have been 2 h long and have been divided into three main sections. During the first phase, each one of the participants introduced himself and told about his habits as far as

INCORRECTED PROOF

AUTHOR'S PROOF
JrnlID 11042_ArtID 449_Proof# 1 - 07/01/2010

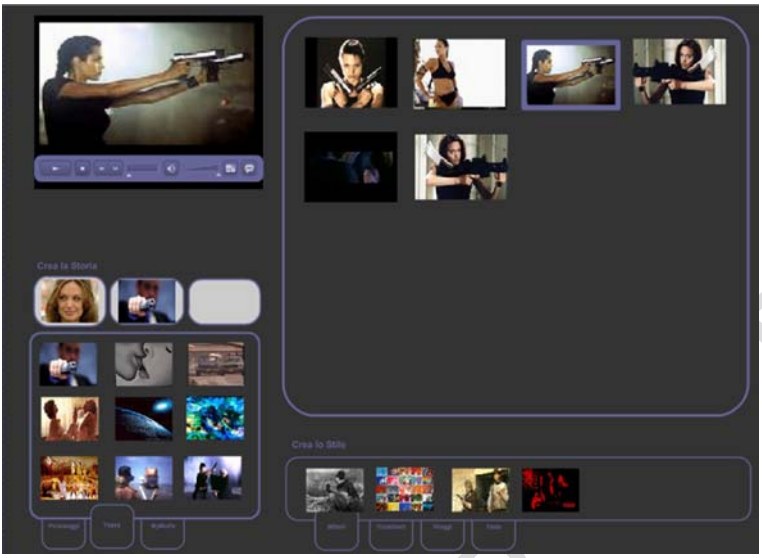


Fig. 4 Screenshot of the main window

Q5

the fruition and/or creation of the various kinds of audiovisual communication (trailer, commercial, clip, short film). The primary aims were both highlighting possible lacks of functionalities in tools at disposal of each user today and listening to the needs and demands of the users (and eventually the motivations at the base of the lack of interest in using video editing tools).

During the second section three different tools have been introduced to the users—MuveeMix [http://www.muveemix.com], JumpCut [http://www.jumpcut.com], Sweeney Todd Trailer Editor [http://www.sweeneytoddmovie.com/mashup/editor.php]: they are web based with specific characterization (i.e. the first focusing on its easiness of use, the second on its social aspects, the third on the contents to be remixed). An evaluation of their functionalities has been asked to users with the aim of highlighting the most interesting and innovative functionalities.

In the last section our concept application has been introduced through two use cases drawn in storyboards in order to evaluate its general acceptability and the offered functionalities plus having new ideas for a future development of the project. At the end of the focus group users answered to a questionnaire: users had to point on the main five functionalities emerged from the discussion that should enrich the functionalities of our new application.

As far as the application is concerned the results have been quite homogeneous between the two groups. It emerges the need to create and compose videos in a very simple manner, without a long training on the applications. The main motivations in the lack of use of the actual video editing tools on the market seem to be the excessive amount of time requested for learning and using these applications and the difficulty to find the raw material to starting from in the composition and creation of a personal work. But, in comparison of the tools already known by the user and the applications presented during the discussion, our tool is perceived more immediate to use and with bigger potentialities in order to express the creativity of the users. First of all, the results highlighted the high satisfaction in the concept. The application is perceived as simple and intuitive to use. The emerged most liking aspects are:

- the concept allowing to re-elaborate contents of the motion-picture (Paola, a user involved in the second group, said "*The cinematic imaginary is so huge and various!*

UNCORRECTED

AUTHOR'S PROOF JrnlID 11042_ArtID 449_Proof# 1 - 07/01/2010

It's wonderful that I can find the clips that I want and combine them in a personal way to express my creativity all in a single application")

- the organization in topos, celebrities and "stilema" allowing an original, amusing and innovative search (Giorgio, a user involved in the second group, said "*This search modality is very simple and original; it is really made for people who loves cinema*");
- the visual predominance rather than the textual one in the settings of the application (Elisa, a user involved in the first group, said: "I really like the idea to limit the textual command and label in the interface. In this manner the interface is more intuitive"

Some other suggestions that came from the users will be considered in the future steps of design. The most unexpected findings were:

- Users suggest to use the application not only to create new videos, but also to compose "visual playlists", collecting the best scenes of the favorite movies and letting them know and exchange with other users. From the analysis of data it emerges strikingly the demand of a place where fruition and creation can be carried out together: the users involved express their enthusiasm about using the tool even as an organized archive where finding particular scenes to be watched once again (Elisa, a user involved in the first group, said "*I would like to see again in sequence all the kiss scenes that I liked most in the history of cinema*").
- Users suggest to limit the action of the rule engine, not in his selection activity but mainly in his post-production activity, because they were afraid of losing control on their work, thinking that an excessive work of the artificial intelligence would reduce their creativity and their possibilities of expression (Marco, a user involved in second group said: "*I really like that the application could find for me all the material that I want, even the clips that I didn't expect; but I don't think that it should modify them; this is my role*".

While the first suggestion would be strongly considered in the next phases of the design, we consider the second one less important. We think that the attitude of the users against the intelligence of the engine is having reference to a misunderstanding of his real potentiality. In fact, a major explanation of the work of the rule engine carried on during the last phase of the focus groups, in order to explain his action to support the users, has partially reduced the negative perception. Besides in the final version of the tool this functionality will be hidden to the user that will receive only the results already post-produced. Then, to test the liking of this functionality, will be necessary not to ask directly to the user the acceptance of the philosophy of the automatic support, but to compare the liking of the post produced results with the simply selected and raw results.

Finally, the functionalities emerged as basic from the questionnaire are those of establishing the start and the end of each clip, the separated management of the audio and video parts and, finally, the chance to insert private and personal contents: these features are considered essential from the 90% of the users, while other functionalities (for example the social aspects, like sharing contents and work in collaboration with other users) gain a smaller consent.

3 Conclusions

In this paper we presented a new application of video mashup working with a peculiar process and allowing the editing of pre-existing contents apart from favoring the collection of contents and their flow through different media. The application described in fact is not a stand alone means but rather a web companion for a new interactive TV. The main aim of this new

UNCORRECTED PROOF

AUTHOR'S PROOF
JrnlID 11042_ArtID 449_Proof# 1 - 07/01/2010

application is that of allowing easy and immediate editing through a recreational process of handling and editing the videos. Nevertheless the work presented is not to be considered a finished and completed work, but just the first step of a huge cyclic project going from the building of the application to its testing and back to a revision process up to the construction of a tool satisfying the requests of the users.

At the moment as we described before the application has already gone twice through this process; now it is going through a major process of redesign in order to be perceived by the user as more easy but more complete in functionalities. After having completed this phase the application will be once again submitted to a testing phase to establish if major changes are still needed.

References

1. Adami N, Leonardi R (1999) Identification of editing effect in image sequences by statistical modeling, in Proceedings of Picture Coding Symposium PCS '99, Portland, Oregon, USA, pp 157–160
2. Aimeri L, Frasca G (2002) Manuale dei Generi cinematografici. Utet, Torino
3. Anderson C (ed.) (2005) Remix planet. Wired Issue 13.07.05.
4. Askwith ID (2007) Television 2.0: Reconceptualizing TV as an Engagement Medium. Masters thesis of Science in Comparative Media Studies at MIT
5. Benini S, Bianchetti A, Leonardi R, Migliorati P (2006) Extraction of significant video summaries by Dendrogram Analysis. In Proceedings of International Conference on Image Processing ICIP'06, Atlanta, GA, USA
6. Benini S, Migliorati P, Leonardi R (2008) Fast dialogue indexing based on structure information. In Proceedings of ICIP'08, San Diego, CA, USA
7. Benini S, Xu L-Q, Leonardi R (2005) Identifying video content consistency by vector quantization. In Proceedings of WIAMIS'05, Montreux, Switzerland
8. Bocconi S (2004) Semantic-aware automatic video editing. In Proceedings of the 12th Annual ACM International Conference on Multimedia, ACM, New York, NY, USA
9. Bordwell D (1997) On the history of film style. Harvard University Press, Cambridge
10. Bordwell D, Thompson K (1994) Film history: an introduction. McGraw-Hill, New York
11. Bordwell D, Thompson K (2003) Film art: an introduction, 7th edn. McGraw-Hill, New York
12. Brusilovsky P (1996) Methods and techniques of adaptive hypermedia. User Model User-Adapt Interact 6(2–3):87–129
13. Casares J, Yocum D, Stevens S, Myers BA, Long C, Dabbish L, Corbett A, Bhatnagar R (2002) Simplifying Video Editing Using Metadata. In Proceedings of ACM Conference on Designing Interactive Systems: Processes, Practices, Methods and Techniques (DIS 2002)
14. Casetti F, Di Chio F (1990) Analisi del film. Bompiani, Roma
15. Cassani D (2006) Manuale del montaggio. Utet, Torino
16. Center of Social Media (2008) Recut, reframe, recycle. Quoting Copyrighted material in User-Generated Video. http://www.totalrecut.com/CSM_Recut_Reframe_Recycle_report.pdf
17. Cotsaces C, Nikolaidis N, Pitas I (2006) Video shot detection and condensed representation. IEEE Signal Process Mag 23:28–27
18. Diakopoulos N, Luther K, Medynskiy Y, Essa I (2007) The evolution of authorship in a remix society. In Proceedings of Hypertext and Hypermedia. Manchester, UK, September 2007
19. Forlai L, Bruni A (1998) Archetipi mitici e generi cinematografici. Audino, Roma
20. Gallagher O (2008) "Video recuts and the remix revolution: whose rights are being infringed?" The Development of an Online Remix Community known as Total Recut http://www.totalrecut.com/masters.doc
21. Giest M (2007) The rise of clip culture online. BBC News
22. Girgensohn A, Boreczky JS, Chiu P, Doherty J, Foote J, Golovchinsky G, Uchihashi, S, Wilcox L (2000) A semi-automatic approach to home video editing. In Proceedings of UIST 2000
23. Hanjalic A, Lagendijk RL, Biemond J (1999) Automated high-level movie segmentation for advanced video retrieval systems. IEEE Trans Circuits Syst on Video Technology 9
24. Hua X, Wang Z, Li S (2005) LazyCut: content-aware template-based video authoring. In proceedings of ACM Multimedia 2005
25. Hua X, Lu L, Zhang H (2004) Optimization-based automated home video editing system. IEEE Trans Circuits Syst Video Technol 14(5):572–583
26. Jeannin S, Divakaran A (2001) "MPEG -7 visual motion descriptors," in IEEE Trans Circuits Syst Video Technol, June 2001

UNCORRECTED PROOF

AUTHOR'S PROOF JrnlID 11042_ArtID 449_Proof# 1 - 07/01/2010

27. Jenkins H (2006) Convergence culture: where old and new media collide. New York University Press, New York, 2006a
28. Kuleshov L (1974) Kuleshov on film: writings by Lev Kuleshov. University of California Press, Berkeley
29. Kumano M, Ariki Y (2002) Automatic useful shot extraction for a video editing support system. In Proceedings of MVA 2002:310–313
30. Lévi-Strauss C ([1962] 1974) The savage mind. London: Weidenfeld & Nicolson
31. Moine R (2005) I generi del cinema. Lindau, Torino
32. Mu X, Marchionini G (2003a) Enriched video semantic metadata: authorization, integration and presentation. In Proceedings of the 66th Annual Meeting of the American Society for Information Science and Technology 40:316–322
33. Mu X, Marchionini G (2003b) Statistical visual feature indexes in video retrieval. In Proceedings of SIGIR 2003
34. Peng WT, Chiang YH, Chu WT, Huang WJ, Chang WL, Huang WJ, Hung YP (2008) Aesthetics-based automatic home video skimming system. In Proceedings of International Multimedia Modeling Conference, 2008
35. Rondolino G, Tomasi D (1995) Manuale del film. Linguaggio, racconto, analisi. Utet, Torino
36. Schwarz K, Kouwenhoven T, Dignum V, van Ossenbruggen J (2005) Supporting the decision process for the choice of a domain modeling scheme. In Formal Ontologies Meet Industry
37. Shaw R, Schmitz P (2006) Community annotation and remix: a research platform and pilot deployment In ACM Multimedia 2006, Oct. 23–27, Santa Barbara, CA
38. Simeoni R, Etzler L, Guercio E, Perrero M, Rapp A, Montanari R, Tesauri F (2007) InnovativeTV: from an old standard to a new concept of Interactive TV—an Italian job. In: Jacko JA (ed) HCI 2007. LNCS, vol 4552. Springer, Heidelberg, pp 971–980
39. Simeoni R, Geymonat G, Guercio E, Perrero M, Rapp A, Tesauri F, Montanari R (2008) Where have you ended up today?—Dynamic TV and the Inter-tainment Paradigm, Proc. EuroITV 2008 Conference, Springer Verlag LNCS 5066, pp 238–247
40. Stefik M (1995) Introduction to knowledge system. Morgan Kaufmann, San Francisco
41. Tomasi D (2004) Lezioni di regia. Modelli e forme della messinscena cinematografica. Utet, Torino
42. Tomasino R (1998) Figure dell'immaginario. Cento anni di mitologie del cinema. L'epos, Palermo
43. Vellar A, Simeoni R, Montanari R, Rapp A (2008) A parasocial navigation concept for movie discovery. In Proceeding of IADIS International. Conference Interfaces and Human Computer Interaction 2008
44. Wang J, Xu C, Siong C, Duan L, Wan K, Tian Q (2005) Automatic generation of personalized music sports video. In Proceedings of ACM Multimedia 2005
Q10 45. Zeng W, Gao W, Zeng DZ (2002) Video indexing by motion activity maps In Proceedings of ICIP 2002, I-912-I-915 vol. 1, Rochester, NY, USA

UNCORRECTED PROOF

Q5

Q1/Q11 **Daniela Cardillo** Current position: Computer Science Department—University of Turin: (PhD Student Education 2006). Degree in Communication Sciences specialization in Communication in the Society of Information Technology at the University of Turin (Italy); (2002). MA in Eighteenth Century Studies: Blake and the Age of Revolution at the University of York (UK); (1996). Degree in Foreign Languages and Literature specialization in German Philology at the University of Messina (Italy)

AUTHOR'S PROOF
JrnlID 11042_ArtID 449_Proof# 1 - 07/01/2010

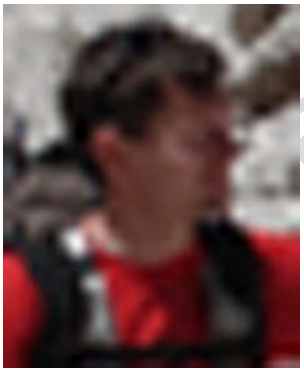
Q5

Q11

Amon Rapp Degree in Communication Sciences from the University of Turin, Italy, 2006. Since 2006 he's been working in the research and trends area in Telecomitalia Lab. In 2007 he received a research scholarship "Progetto Lagrange—Fondazione C.R.T." He is with Telecomitalia Lab and the department of Computer Science, University of Turin. At present he works in the Service Design activities within the Continuous Cross Ambient Communication Research Project and the Dynamic Tv Research Project of Telecomitalia.

CORRECTED PROOF

Q5



Q11

Sergio Benini was born in August 1975. He is originally from Verona, Italy and has been living and studying in Brescia for his academic studies. He's received his Electronic Engineering Master's degree (cum laude) at the University of Brescia in 2000 with a thesis entitled "3D Medical Images Segmentation and Visualization", which won prize "D.B. Grazioli" by Italian Academy of Science. Between May 2001 and May 2003 he's been working in Siemens Mobile Communication R&D in Milan, Italy, on mobile network management projects. From November 2002 he's started in Brescia his PhD studies in Information Engineering under the supervision of professor Riccardo Leonardi, focusing on Video and Multimedia Analysis. During his PhD studies, between September 2003 and September 2004 he has conducted a placement in British Telecom Research (Ipswich, U.K.) working in the "Content & Coding Lab" on a clustering procedure for grouping video shots into story units under the supervision of Dr. Li-Qun Xu. Since March 2005 he has become Assistant Professor in the Telecommunications group of DEA at the University of Brescia.

AUTHOR'S PROOF JrnlID 11042_ArtID 449_Proof# 1 - 07/01/2010

Q5

Q1/Q11

Luca Console Position: Full Professor of Computer Science, Università di Torino, Dipartimento di Informatica; Facoltà di Lettere e Filosofia, Corso di laurea in Scienze della comunicazione, Past positions 1989-1992: Researcher (Assistant Professor), Universtà di Udine 1992-1993: Associate Professor, Universtà di Udine 1993-2000: Associate Professor, Università di Torino Education Laurea Degree in Computer Science, University of Torino 1985 PhD in Computer Science, 1990

CORRECTED PROOF

Q5

Q11

Rossana Simeoni si è laureata in scienze dell'informazione nel 1991 e nel 2005 consegue il diploma di maturità artistica. Dal 1992 lavora in TILab. Ha collaborato ed ha assunto ruoli di responsabilità in progetti focalizzati su: Gestione dei Servizi, Gestione dei Processi e CRM. È stata referente per Telecom Italia nel progetto europeo IST-CAWICOMS sulla personalizzazione dei servizi e prodotti. Attualmente è impegnata nella progettazione e sviluppo di un nuovo paradigma di TV interattiva nel contesto del progetto *DynamicTV.* All'interno del progetto si occupa principalmente di *Human-Computer Interaction* e di strategie di correlazione semantica rivolte a coniugare Media e Comunicazione.

AUTHOR'S PROOF
JrnlID 11042_ArtID 449_Proof# 1 - 07/01/2010

Q5

Q11

Elena Guercio was born in Turin, Italy, on 9th September 1969. In 1994 she graduated in Padua in Psychology. In 1997 she attended the Corep Master in Ergonomics and now she's one of the 34 certified european ergonomists in Italy (EurErg). From 1994 to 1995 she had a post lauream experience in a consulting office for training on the job. From 1995 to 1998 she worked in Telecom Italia and in Fiat Auto, as a consultant in usability field. In 1998 she was employed in Telecom Italia. For 7 years she's worked in the usability group where she has been involved in a lot of telco and ICT projects both internal and european. The projects have concerned: planning and evaluating new services (i.e. VAS for fixed and mobile network), products (i.e. mobile and fixed phones, web sites and portals), systems (i.e. videocall systems) and also processes (communication between mobile office and a control centre: WFM management) from the users' point of view. The applied methodology has been, in particular, the user centred design and all its tools (focus and creative groups, usability test, cognitive walkthrough, questionnairres, etc…). Since 2005 she's been working in the research and trend area in Telecom Italia, where the service design and the user centred planning is focused on blue sky projects and not only on the short term ones.

CORRECTED PROOF

Q5

Q11

Prof. Riccardo Leonardi has obtained his Diploma (1984) and Ph.D. (1987) degrees in Electical Engineering from the Swiss Federal Institute of Technology in Lausanne. He spent 1 year (1987-88) as a post-doctoral fellow with the Information Research Laboratory at the University of California, Santa Barbara (USA). From 1988 to 1991, he was a Member of Technical Staff at AT&T Bell Laboratories, performing research activities on image communication systems. In 1991, he returned briefly to the Swiss Federal Institute of Technology in Lausanne to coordinate the research activities of the Signal Processing Laboratory. Since March 1992, he has been appointed at the University of Brescia to lead research and teaching in the field of Telecommunication. His main research interests cover the field of Digital Signal Processing applications, with a specific expertise on visual communications, and content-based analysis of audio-visual information. He has published more than 50 papers on these topics. Since 1997, he acts as an evaluator and auditor to the European Union ACTS programme.