# Being a self-director: enhance user creativity with a video mash up tool

Amon Rapp[1], Daniela Cardillo[2], Rossana Simeoni[3], Luca Console[2]

[1]Telecom Italia, Technology, Research and Trends department Via G. Reiss Romoli, 274, - University of Turin – "Progetto Lagrange – Fondazione C.R.T." 10148 Turin, Italy

amon.rapp@guest.telecomitalia.it

[2]Department of Computer Science, University of Turin, corso Svizzera 185, 10149, Torino

{cardillo, lconsole}@di.unito.it

[3] Telecom Italia, Technology, Research and Trends department Via G. Reiss Romoli, 274, 10148 Turin, Italy

rossana.simeoni@telecomitalia.it

## ABSTRACT

In this paper, we describe the development of an innovative tool of video mash up. This application is immediate and intuitive to be used by non professional users for creative and (re)creational moments; it works taking the information from a repository of videos and putting into action an intelligent system that combines low level features and high level metadata to provide a semi-automatic editing supporting users in the production of video mash up.

## Categories and Subject Descriptors

D.2.2 [**Software**]: Design Tools and Techniques – *modules and interfaces, object-oriented design methods, user interfaces.*

## General Terms

Design, Reliability, Experimentation, Human Factors.

## Keywords

Video mash up, intuitive interface, intelligent system.

## 1. INTRODUCTION

The actual world of entertainment culture is evolving and having a deep change because of the convergence of media and technological platforms and, at the same time, thanks to the spreading of technologies and applications accessible to different kinds of users. There are two different processes going on: the first one is the unification of traditional media with new media starting a cross-media framework [9]. In this environment different contents' flows can spread around in different devices and different communication network. The second process finds its main reason in the growing dissatisfaction coming from the extreme passiveness of the user using traditional media. It grows consequently a new active culture especially in young people bringing them both to use contents different from those proposed by traditional media and to create expressive works of their own [16]. The growing spreading of application belonging to Web 2.0 allowed common users to express their own creativity and to share it quickly with other users. One of the key-points of this change is made by ***video mash up***, an audiovisual product made by users, in which professional-produced audiovisual contents are remixed to create new amateur contents that express new meanings. Contents already existing in the mass culture are driven from different sources out of their contexts and remixed by non professional users at different levels - audio, video and both at the same time – with the aim to create new kinds of cultural and artistic products. In this paper we describe a new application with the aim to improve the user experience during the different steps of the composition of the video mash up. The final target of this new tool is both supplying a new instrument of effective, original and amusing research of raw material for the mash up and supporting the recreational task of the user giving him back as a result a semi automatic editing with a stylistic and semantic homogeneity.


*Ace 2009*, Oct 29-Oct 31, 2009, Athens, Greece


## 2. DESIGNING THE CONCEPT

Some of the existing research studies focused on the chance of offering automatic or semi automatic editing supporting the video editing work to the users; some of them used exclusively methods of automatic analysis and automatic extraction of video features (e.g., [6, 8, 12]). Others focused on the semantic annotation [4], or on the mix of video, audio and written texts analysis [22]. None of these works have been realized specifically for easing the video mash up work. Even the existing tools do not seem to be realized for a specific use of them in the video mash up context; we are speaking both about professional tools (e.g. Apple's Final Cut Pro [2]) and tools for non expert people (e.g. Apple's iMovie [3]). Nevertheless in the last two years few tools online saw a simplification of their functionalities (e.g. Jumpcut [11]) and seem to be directly correlated with the video mash up creation [19]. Nevertheless none of these applications seems to be helpful in the research and in the composition phase; often greater simplicity is meant only by a minor number of functionalities and options.

To fully understand the context in which to insert the new application, an analysis of the ecosystem has been held through a qualitative investigation of documents and a two-months long

participant observation. We examined two web portals for the collection and the distribution of audiovisual material produced by users: the first dedicated to the collection of video mash up (TotalRecut [20]), the second opened to any audiovisual content (Youtube [24]). The analysis focused on the communicational aims, on the stylistic aspects of video (i.e. the kind of editing, the use of photos, etc.) and audio (e.g. change of soundtrack, insertion of voice over) and on the semantic aspects (e.g. a narrative dimension, the change in the meaning compared to the source, etc.) of video mash up. As a result of the analysis we had a detailed classification of different genres of the most diffused mash up in the Web. This classification tries to go deeply into the description of the found genres listing their semantic, syntactic and communicational characteristics apart from their sub-categories whether they exist (Table 1 is a brief summary in which some emerging characteristics of some genres are illustrated synthetically). Moreover, in order to evaluate the most common solutions on the actual market of tools of video editing for non professional users, an heuristic and comparative investigation has been held on five of the most popular tools [3, 10, 11, 19, 23] with the aim of finding some standard de facto, best and worst practices in the design and in the functionalities offered by these services.

**Table 1. Example of classification of video mash up**

| Kind of Mash up | Communicative aim | Stylistic aspects | Semantic aspects |
|---|---|---|---|
| **Movie Recuts** | Creative, recreational, parodying | Voice over, audio and editing rhythm change | Narrative dimension, cohesion, etc. |
| **Tributes** | Celebrative | New soundtrack, title, slow-motion | A-temporal representation, no narrative dimension |

The work carried on in this phase allowed us to derive some guidelines for the design of a new video mash up tool: ***reduction of control (set filters) favouring easiness of use:*** a strongest easiness in managing the editing functionalities allows users to focus on communicating different and personal meanings rather than on the technical perfection of the editing work; ***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; ***simplification in the way of retrieval material to be re-elaborated:*** finding and mixing the searched clips will be for the user a unique and easy flow of experience; ***emphasis on the amusement and the serendipity:*** users' creativity can be incited through the retrieval of unexpected videos; ***emphasis on the cinematic world:*** exploiting users' affection towards icons and typical moments of the cinema to increase the participation and stimulate the creativity.

Starting from these guidelines the initial concept of the tool of video mash up has been defined. The tool is a web-based application conceived for non professional users, having recreational and amusement goals. It has been thought as an integrated component of a cross-media framework as a web companion of DynamicTv [18]. The final aim is that of giving birth to a virtuous circle in which the increase of users participation and quantity of contents coming from making public the produced mash up are directly linked. The tool mixes a navigation interface to explore contents with a video editing interface favouring the visual communication rather than the textual one [21]. Users can query the repository of clips through three different variables that are the topos, the celebrities and the "stilema". ***Topos*** represent the narrative places, that is cyclic themes universally recognized as belonging to a well-defined genre, like for example the gunfight, the countdown, the robbery and the explosion in the action genre. ***Celebrities*** represent the motion-picture actors, the most famous and most recognizable in the cinematic world. The "***stilema***" represent predetermined visual styles established analyzing the different styles occurring in the history of cinema or directly linked to the genre culture (e. g. the "fear stilema" is characterized by nocturnal colours and a nervous editing style and camera movements).

**Figure 1:** mash up tool main window

The tool has a graphical interface easy and intuitive to be used (Fig.1) that allows the users to select the material for the video mash up and to manipulate this material for their creative aims. In particular, in the bottom left section of the interface users will be allowed to choose up to two celebrities, a topos and a "stilema" in the same query: as a result they will not receive simply a list of clips having the parameters of the given query, as in a traditional search engine, but they will have a set of clips representing the celebrities taken in a typical narrative context and having semantic coherence and stylistic homogeneity as high as possible. This result will be visualized in the central area of the interface. This result is a pre-elaborated editing produced automatically by the tool which can be modified directly by the user (he/she can change the order the clips are inserted in the editing or the style associated at the clips can be changed, etc.). Clips can be modified and moved through the whole video thanks to the Drag & Drop function. In the top left section of the interface, users can visualize the content of each clip in the video player.

# 3. A WORKFLOW OF INFORMATION: THE PROCESS OF CREATION

This section concerns how the whole process of creative entertainment will take place allowing users to interact with a simple tool and, in few steps, to create a video mash up. To better understand the creation process, it will be described as a mainstream of actions happening in a unique flow. This will be shortly introduced by a description of the user's actions followed by a deep analysis of the corresponding events in the back-end of the tool. Each of the modules will be briefly analyzed, and the reader can refer to Figure 2 as an illustrated version of the architecture. Analyzing the different steps it will be clear that the unique task of the user will be that of inserting his query without worrying at all about the complexity of the process going on in

the back-end and having as a result a final editing using given clips.

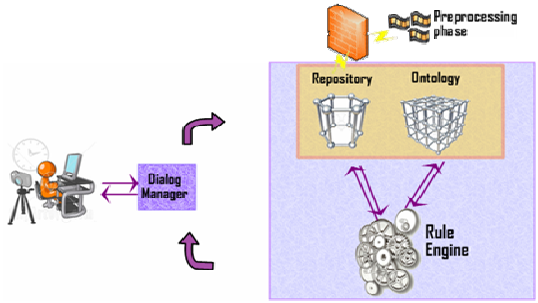


**Figure 2:** abstract model of the architecture of the mash up tool

The user starts his experience with the mash up tool as a non expert subject with the goal of creating a short movie to test his abilities as a novice director. Once the user approaches the tool the unique knowledge he needs to have is that of his own idea, that is the characters he would like to be in his video and the appeal he wishes for his creation. The user can carry out his choices selecting in the interface the celebrities (a maximum of two units), a topos and a stilema to form and to be applied to the final product. Once the user selects his choices the ***dialog manager*** is charged of establishing an interaction between the interface and the back-end of the tool. The dialog manager grants a complete and secure monitoring of data exchange.

At this point the combined query is passed to another module that is the ***repository*** containing all the clips and the full length videos. Here each one of the clips is identified with a unique ID plus a series of metadata both of low and high level assigned to the clips during a ***phase of pre-processing***. Each video has a hierarchical partition decomposing in five levels having on top full videos. The second level is made by scenes conveying a high level concept or a short and simple story; that is why they are also defined LSU, or Logical Story Units [7]. At a lower level scenes can be segmented in shots that are the longest continuous frame sequences coming from a single camera uninterrupted run. The lower level is made by key frames which can be extracted from shots as a static representative visual content. The repository content is constituted by videos segmented at different levels of the illustrated hierarchy; these videos bring with them a series of metadata [14, 15]. During the pre-processing phase both full videos and clips (LSU, shots) are enriched by high level features as title, directors, actors, year of production (typically used to describe complete movies in the movie databases), and low level features as characteristics related to colour, movement and audio which can be automatically extracted identifying in a weighted way each single shot. The clips (LSU and shots) are enriched even by metadata structured on the basis of a specific ontology and manually inserted by an editorial staff. This ***ontology*** is a structure of the cinematic knowledge managing the relationship between the attributes of each clip in order to rule at a semantic level the categories of the domain [17]; this ontology is a formal conceptual model finding his base in the analysis of the reference literature and, at the same time, in the specific know-how of experts of cinema and strong cinema users (see below par. 4). The main categories structured in the ontology are the celebrities, the topos and the editing styles or "stilema" with their properties and the related restrictions. For example, the celebrities can be actors like 'John Wayne' and 'Anthony Hopkins', the topos can be 'fight' and 'kiss', the stilema can be 'fear' and 'silent cinema'.

A further step, hidden to the user but coming as a consequence of his query, is the process through which goes the provisional result of the query in the ***rule engine***; this works on the knowledge base in order to produce an editing of (some of the) retrieved videos. So, once the query of the user is processed through the repository, it is further submitted to a process of reasoning in order to expand the knowledge through rules of inference. This process transforms the tool from a simple means to have a retrieval of the requested clips, as a traditional search engine, into a more sophisticated instrument, returning as a result of the query an automatic editing exploiting high and low level features metadata besides of tags. To fully understand this last phase of the process it is necessary a quick but detailed survey of the reasoning. The rule engine works using multiple sets of rules: *rules of selection*, processing the user's criteria in order to return only the clip really interesting for him; *rules of priority*, selecting the clips and giving them a specified weight remarking an order of importance between all the clips selected; *rules of editing*, structuring the clips in a pleasant way by applying some simple effect like that of erasing the jump cut; *rules of atmosphere*, giving a visual homogeneity to the resulting editing.

The query initially formulated by the user is characterized by one or more actors, a topos and a "stilema"; going through the process described above, it goes without saying that the resulting clips are more and more reduced passing through the various steps and the different components. The rule engine applies its rules, like for example:

- rules of selection: choose the clip basing on how many elements satisfy the request of the user (actors, stilema, topos);
- rules of editing: measure the possible aggregation of couples of clips basing on the LLF (similar texture, direction of movements, similar colours) and then propose a first editing.

At this point the result is in its final form and the Dialog Manager can return it to the user who can modify or enrich it following his own preferences. Watching at the editing proposed by the tool, the user can decide that the result is satisfactory or proceed with a further query to enrich the editing with other clips, in order to give an additional logical meaning to the final result. In case the clips are enough for the aim of the user he can go on modifying the audio and video characteristics of the sequence proposed by the tool.

## 4. USER STUDIES

During the service design steps in which the development of the tool's concept took part users have been involved in the process more than once and for different aims.

In order to define the main categories of the ontology, that is the topos, two focus groups have been organized. They had the objective to validate and integrate a classification of topos and "stilema" defined by a group of specialists based on the references literature [e.g. 1, 5, 13]. Focus were composed by two groups of nine people each with an high movie knowledge but separated for age (i.e. 18-35 e 36-65). The focus groups lasted two hours each and were separated into two main steps: during the first the users were put into a brainstorming session in which they proposed autonomously recurring topos characterizing a specific genre.

During the second step the results of the previous phase have been compared with the classification made by the experts. The results of these works with users allowed us to define a specific ontology which tries to give a structure to the cinematic world of Italian users.

In order to test the acceptability of the concept and of the offered functionalities in the tool of video mash up, two focus groups, with eight people each, have been organized. The first group was composed by experts of video editing. On the opposite the second group was made of experts of cinema potentially interested in the use of video mash up editing tools. Focus groups lasted about two hours each and were divided into three parts. The first phase had the objective to collect habits and uses of users while watching videos on the web and their use of video editing applications. The second phase had the aim of allowing users to find a set of essential functionalities for a tool of video mash up, through the discussion of the results of the heuristic analysis carried out during the initial phases of the concept design (see above par. 2). The last drove its centre on the presentation of the service concept and of its functionalities, introduced to users through specific use cases. As far as the application is concerned, the results have been quite homogeneous between the two groups. Users appreciated the proposed application: in particular the possibility to re-elaborate the material coming from the cinematic world and the organization of the repository of clips in topos, actor and "stilema" (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*") were considered original, amusing and innovative. The application is perceived as simple and intuitive to use, 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 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*"). This demand places itself at the basis of the change happening in the entertainment today, where the vision of contents is more and more linked to the (re)creational activity.

## 5. CONCLUSIONS & FUTURE WORKS

The illustrated work has its major quality in its being a tool for non professional users immediate to use and allowing creative users to give vent to their propensity as directors without encountering the difficult problems of using professional tools with hundreds of functionalities. Natural and future prosecution of the work will be the re-design of the tool in order to add new functionalities, giving in such a way an effective response to the users' demands coming out from the focus groups.

## 6. REFERENCES


[1] Aimeri, L. and Frasca, G. 2002 Manuale dei Generi cinematografici. Utet, Torino

[2] Apple. "Final Cut Pro". http://www.apple.com/it/finalcutstudio/finalcutpro/

[3] Apple. "iMovie." http://www.apple.com/ilife/imovie/

[4] Bocconi, S. 2004. Semantic-aware automatic video editing. In Proc. ACM MM 2004.

[5] Cassani, D. 2006 Manuale del montaggio. Utet, Torino.

[6] Girgensohn, A., Boreczky, J. S., Chiu, P., Doherty, J., Foote, J., Golovchinsky, G., Uchihashi, S. and Wilcox, L. 2000. A semi-automatic approach to home video editing. In Proceedings of UIST 2000.

[7] Hanjalic, A. Lagendijk, R. L. and Biemond, J. 1999. Automated high-level movie segmentation for advanced video retrieval systems. IEEE Transactions on Circuits and Systems on Video Technology, vol. 9, June 1999.

[8] Hua X., Wang Z. and Li, S. 2005. LazyCut: content-aware template-based video authoring. In Proc. ACM MM 2005.

[9] Jenkins, H. 2006. Convergence Culture: Where Old and New Media Collide. New York University Press, New York

[10] JayCut, http://jaycut.com/

[11] JumpCut, http://www.jumpcut.com

[12] Kumano M., Ariki, Y. 2002. Automatic Useful Shot Extraction for a Video Editing Support System. In Proc. MVA 2002: 310-313.

[13] Moine R. 2005. I generi del cinema. Lindau, Torino

[14] Mu, X. and Marchionini G. 2003. Enriched video semantic metadata: authorization, integration and presentation. In Proc. 66th Annual Meeting of the American Society for Information Science and Technology (Vol.40, 316-322).

[15] Mu, X. and Marchionini, G. 2003 Statistical visual feature indexes in video retrieval. In Proceedings of SIGIR 2003.

[16] Obrist, M., Geerts, D., Brandtzæg, P. B. and Tscheligi, M. 2008. Design for creating, uploading and sharing user generated content. In Proc. ACM CHI 2008.

[17] Schwarz, K., Kouwenhoven, T., Dignum, V. and van Ossenbruggen., J. 2005. Supporting the decision process for the choice of a domain modeling scheme. In Proc. FOIS 2005

[18] Simeoni, R., Geymonat, G., Guercio, E., Perrero, M., Rapp, A., Tesauri, F., Montanari, R. Where have you ended up today? - Dynamic TV and the Inter-tainment Paradigm. In Proc. EuroITV 2008

[19] Sweeney Todd Trailer Editor powered by GorillaSpot http://www.sweeneytoddmovie.com/mashup/editor.php

[20] Total Recut http://www.totalrecut.com/

[21] Vellar, A., Simeoni, R., Montanari, R., Rapp, A. 2008 A parasocial navigation concept for movie discovery. In Proc. IADIS 2008.

[22] Wang, J., Xu, C. Siong, C., Duan, L., Wan, K. and Tian, Q. 2005. Automatic generation of personalized music sports video. In Proc. ACM MM 2005

[23] Windows Movie Maker http://www.microsoft.com/italy/windows/products/windowsxp/winxp/using/moviemaker/default.mspx

[24] Youtube, http://www.youtube.com.