INTERACTIVE VIDEO MASHUP BASED ON EMOTIONAL IDENTITY

Luca Canini, Sergio Benini and Riccardo Leonardi

Department of Information Engineering, Signal & Communication Lab
University of Brescia, via Branze 38, 25123, Brescia, Italy
phone: + (39) 030 3715528, fax: + (39) 030 380014, email: {*firstname.lastname*}@ing.unibs.it

ABSTRACT

The growth of new multimedia technologies has provided the user with the ability to become a videomaker, instead of being merely part of a passive audience. In such a scenario, a new generation of audiovisual content, referred to as *video mashup*, is gaining consideration and popularity. A mashup is created by editing and remixing pre-existing material to obtain a product which has its own identity and, in some cases, an artistic value itself. In this work we propose an emotional-driven interactive framework for the creation of video mashup. Given a set of feature movies as primary material, during the mixing task the user is supported by a selection of sequences belonging to different movies which share a similar emotional identity, defined through the investigation of cinematographic techniques used by directors to convey emotions.

1. INTRODUCTION

Recent advances in technology have brought basic and semi-professional tools for the production of multimedia content within the reach of common users. As a result, new forms of creativity are emerging and acquiring an ever-growing visibility, such as the practice of combining multiple audiovisual sources into a derivative work (known as video mashup) whose semantics could be very different compared to the one of the original videos. In this context, a key point is the use of pre-existing material, which could be protected by copyright law [6]. This issue has been widely discussed and, although no coordinated action has yet been done to understand such a phenomenon from a juridical point of view, a substantial amount of user-generated content uses copyrighted material in ways that are eligible for fair use consideration, *i.e.,* their value to society is considered greater than the value to the copyright owner [1].

So far, several semi-automated tools for video editing and mashup have been proposed. In [7] and [3] authors present systems for the creation of custom videos which are able to omit those part of the source streams characterized by a low quality in terms of camera motion (too fast pan, slow zoom, etc.). While this approach is suitable for home-made videos which are likely to contain useless parts, the method proposed in [10] focuses on professional content and guides the user in the editing process according to well known rules of the film grammar, combining a low-level analysis of the video with high-level metadata.

The *LazyCut* system [9], based on both content analysis techniques and content-aware authoring templates, suggests the user the temporal structure of the video to be authored and different composition methods. In [12] instead, authors propose a system based on media aesthetics: given a video and an incidental background music, a new content is generated by skimming the video stream so that the visual rhythm caused by shot changes in the final product is synchronous with the music tempo.

In this paper we propose a framework for the interactive creation of new multimedia content using a set of feature movies as primary material. Our approach differs from those adopted by most of the existing tools since in the mixing task we concentrate on the emotive sphere of the video semantics [8]. In such a way, rather than trying to achieve a cognitive coherence of the final mashup, we give preference to its continuity at the emotional level. To do this, we rely on the *emotional space* built in [4] whose dimensions are related to filming techniques adopted by directors to convey emotions. Here each movie is represented as a trajectory, whose evolution over time provides a strong characterization of the film itself, since emotionally different scenes occupy different regions of the emotional space. The principle is that suitable mixing points between movies are to be found where their emotional trajectories get very close and share a common behaviour, thus ensuring a good continuity in the evoked mood of the generated audiovisual product.

Our framework offers two ways to create a mashup: in the first one, on a specific user command the best switch candidate between one movie and another is submitted for user approval. In the second one, the user gets more control on the application being able to constrain the mixing points to a suitable portion of the emotional space as well as to choose among a pool of candidate scenes at each mixing step.

This paper is organized as follows: in Section 2 the emotional space is presented, together with the characterization of its axes. Section 3 illustrates the proposed framework for video mashup, while Section 4 describe a user study and examples of generated mashups. Conclusions are finally drawn in Section 5.

2. SPACE OF EMOTIONAL IDENTITY

The creation of a blockbuster movie is a creative process that undergoes complex phases of product design. One among the several aspects investigated during the movie conception is its emotional impact (or 'identity') *i.e.,* the mixture of feelings the director intends to communicate while shaping such an art product.

In [5] the author presents a tool to describe the emotional identity of any design object, by placing it in a 3D space according to its shape, efficiency and social context. In this space the three axes refer to the so called *natural, temporal* and *energetic* dimensions, respectively. The natural one splits the space into a passional hemisphere, referred to warm emotions, and a reflective hemisphere, that represents offish and cold feelings. The temporal axis divides the space into two zones, one related to the slowness of the past and expe-

 ISSN 2076-1465

rience and another describing an identity quickly projected towards the future. Finally, the energetic axis identifies objects with high emotivity and, in opposition, minimal ones.

Considering a movie as a piece of art and design, in [4] the emotional identity is transposed in cinematographic terms. At first, each axis is associated to a couple of adjectives in a dichotomic relationship. The couple *warm/cold* is linked to the natural axis. The temporal axis is described in terms of *dynamic/slow*, while the dichotomy *energetic/minimal* is associated to the third axis. Then, filming and editing techniques used by directors to convey emotions and give scenes a particular mood [2] are investigated and linked to the selected dichotomies.

2.1 Movies as space trajectories

The dichotomies warm/cold, dynamic/slow and energetic/minimal characterise the axes of the emotional space described in [4] (where the reader can refer to for an in-depth description). For the natural axis, the colour of the light illuminating the scene is considered, since the spectral composition of the light source is very important in the definition of a scene atmosphere. By using a *white patch* algorithm and a non-uniform quantization law, the illuminant is mapped on the $\mathscr{N}$-th interval of a one-dimensional warm/cold scale.

For characterising the temporal axis, the motion sensation evoked in the movie is selected, since motion dynamics are often employed by directors to stress the emotional identity of a scene. To capture the sensation of speed and dynamism or a feeling of calm and tranquility, shot pace, camera and object motions are combined in a single index $\mathscr{T}$, which is bound to the dichotomy dynamic/slow on the temporal axis.

The energy $\mathscr{E}$ of the audio track is linked to the energetic axis, as important scenes in movies are usually bound to a particular choice of the soundtrack, *e.g.,* gentle and pleasant music for romantic moments, loud and aggressive audio for action sequences, silences and reprises in dialogues.

In the defined space, a movie is drawn as a cloud of points, where each point represents a shot, and it is defined by a triplet $\bar{\mathscr{S}} = (\mathscr{N}, \mathscr{T}, \mathscr{E})$. During the movie playback, these points are connected in temporal order by a cubic spline, creating a trajectory which describes the evolution of the movie's emotional identity, as shown in Figure 1. This trajectory, free to move over the entire space, gives an accurate characterization of the movie, since it is not restricted to a fixed set of emotions or to a pleasure-arousal scheme [11].

3. MASHUP FRAMEWORK

In order to provide the user with a video mashup tool, a framework has been built, as depicted in Figure 2. The core of the system is the emotional space, where a shot is rendered as a point whose coordinates give a strong characterization of its emotional identity.

The upper stage, the *application level*, is composed by two modules that implement the mashup functionalities: the first one, called *Assisted mashup* (AM), guides the user almost completely through the mixing task and is suitable for those people who want to be introduced to the mashup practice or just to see at a glance what the system is able to offer. The second one, named *Creative mashup* (CM), has been designed to give the user an in-depth control of the system, accessing and tuning several parameters, in order to obtain

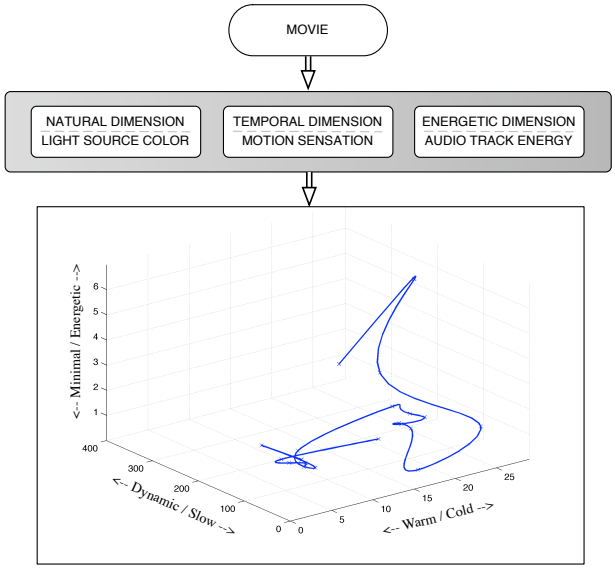

Figure 1: Space of emotional identity: general framework (top) and trajectory from an excerpt of "Matrix" (bottom).

a final multimedia content characterized by a high level of customization.

At the top level, the *graphical user interface* (GUI), providing visual feedback, makes the communication between the user and the system easier. Here follows an explanation of how suitable mixing points are selected as well as a description of the modules and of the user interface.

3.1 Mixing point candidates

Suitable mixing points in the emotional space are selected according to minimum distance criteria on both shot coordinates and movie trajectories. This means that when the user decides to swich movie, the system first looks for a restricted number of shots belonging to other movies which are at low Euclidean distance from the current shot. By using this distance d, defined as

$$d\left(\bar{\mathscr{S}}_i, \bar{\mathscr{S}}_j\right) = \left[(\mathscr{N}_i - \mathscr{N}_j)^2 + (\mathscr{T}_i - \mathscr{T}_j)^2 + (\mathscr{E}_i - \mathscr{E}_j)^2\right]^{\frac{1}{2}},$$

the system is able to retrieve the best shot candidates based on low-level feature characterisation. However, in order to preserve the continuity of the evoked mood over a longer interval and reduce the risk of outliers, the analysis of emotional similarity is extended to movie trajectories. To do this, the pool of the suitable mixing shots is then ranked according to the distance between trajectories over a neighbourhood of shots. In particular, the distance D between the trajectories ξ_A and ξ_B of the movies A and B over a neighbourhood N_h of shots around $\bar{\mathscr{S}}_i \in \xi_A$ and $\bar{\mathscr{S}}_j \in \xi_B$ is given by:

$$D\left(\xi_A, \xi_B\right) = \sum_{k \in N_h} d\left(\bar{\mathscr{S}}_{i+k}, \bar{\mathscr{S}}_{j+k}\right).$$

The ranked list of the best shot candidates is then provided as an input to the mashup modules which are described in the following.

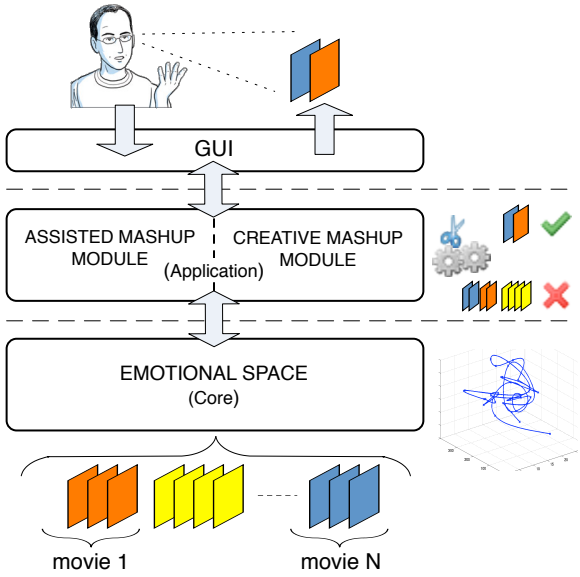

Figure 2: Block scheme of the proposed mashup framework.

3.2 Assisted mashup

The AM module is responsible for the creation of the mashup, without the necessity for the user to set or tune any parameter of the system. This makes the composition task very quick, intuitive and suitable for beginners.

The principle is the following: the user selects a movie in the database and one of its scenes, and the playback starts. On a specific user command, the movie reproduction stops and the user can browse through already played shots. Meanwhile the system looks for a restricted number of shots belonging to different movies which are closer to the one currently browsed in terms of representation in the emotional space, and orders them according to the similarity of the trajectory shapes over a neighbourhood of shots. *At any time, the highest shot in the rank is proposed as the best candidate for the continuation of the mashup.* In case the user approves it as a suitable mixing point, the playback is resumed starting from the new inserted shot and continues with the movie it belongs to, until the next user interaction. Otherwise, the next shot from the ranked list is submitted for user approval.

3.3 Creative mashup

With respect to the AM module, the Creative mashup gives the user more in-depth control on the mixing activities. Before the choice of a starting scene, *the user can focus on a particular sub-volume of the emotional space*, so that only shots within this "reachable" portion will participate in the final mashup. Space constraints can be set independently on the three dimensions, so as to obtain a product with a very specific emotional identity (*e.g.*, fast rhythm, cold lighting and loud music). In case the playback tries to access a shot which does not belong to the reachable volume, the movie reproduction jumps to the following shot placed inside the allowed area.

Differently from the AM module, where only one shot is proposed at a time, here *the whole pool of browsable candidates for selection is shown at once to the user*, who can finally perform his choice regarding the new mixing segment.

3.4 Graphical user interface

Thanks to the two modules described in Sections 3.2 and 3.3, our framework is suitable for both beginners and experts in the mashup practise, who are supposed not to be interested in being aware of the technology behind the application, but to deal only with the editing task at a high level.

The mockup of the interface related to the CM module has been designed (Figure 3) following these guidelines and aiming at obtaining an intuitive and easy-to-use product. This interface is composed of four main objects:

- the window for the playback in the right upper corner;
- the big panel in the bottom where the user can browse through the key-frames of the played shots;
- the panel where the trajectory is depicted synchronously with the playback and where the boundaries of the reachable volume of the emotional space can be tuned by using the three bars on the left;
- the window in the left bottom corner containing the key-frames of the candidate shots for the mashup task, which is activated by pressing the red "mashup" button placed in the middle of the interface.

The interface associated with the AM module can be easily derived from the one presented, by hiding those elements which are not intended to be available in this modality (*e.g.*, the three bars for the selection of the sub-volume in the emotional space).

4. USER TEST AND GENERATED EXAMPLES

In order to verify how our framework is perceived by users and to have a feedback about the quality of the achievable results, a user study has been carried out by using the video material coming from 90 movies of different genres. A group of 8 people of both genders, with a different degree of experience in the usage of multimedia editing tools and interested in the practice of video mashup have been asked to generate new mashups with the AM module.

At first, users have been taught about the system, *i.e.*, the principles it is based on and how it works. Contextually, they have been freed to use the module for a session of thirty minutes, in order to become more confident with the application. After the initial training, each member of the group has been assigned the task to create a mashup (from 2 to 5 minutes long) trying to convey a particular mood which is up to him/her to decide.

In Figure 4 the mashup generated by user #3 is given. The user chose one scene of *The Matrix* as a starting segment and interrupted its playing after 50 *s*. Among the proposed list of shots he selected the ninth in the rank, belonging to *Riders of the Lost Ark*, to continue his mashup. After around 40 *s* the user decided to switch to a *War of the Worlds* shot, which was the fourth one in the selection list. Beyond showing a strong coherence in the scene lighting, motion and sound (corresponding to a *cold/dynamic/energetic* sub-volume of the emotional identity space), all the three segments of the final mashup share, up to a certain extent, "chasing" as a main theme: Trinity hunted by an Agent in the first segment, Indiana Jones escaping from a cave in the second one, and people running away from aliens in the third one.

Figure 5 illustrates a second mashup example, created by user #7. She focused on movies belonging to the drama genre, mainly selecting dialogue scenes with a positive re-

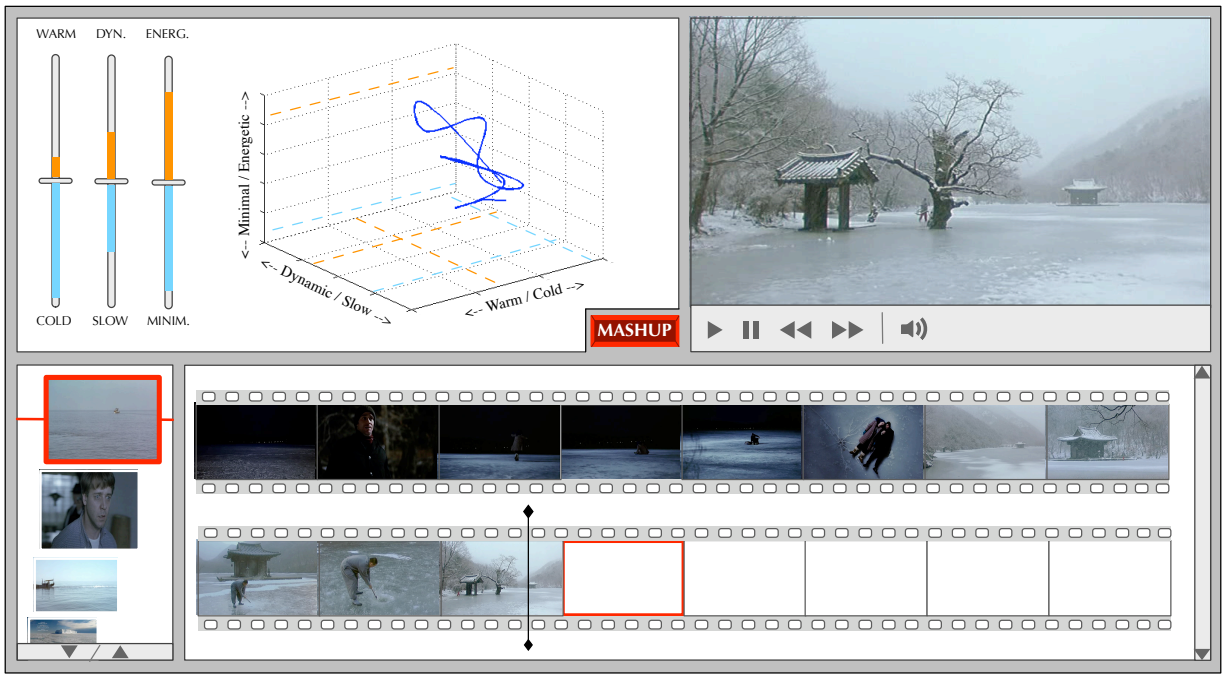

Figure 3: Mockup of the graphical user interface for the Creative mashup module.

Figure 4: Mashup created by user #3 by mixing scenes from *The Matrix*, *Riders of the Lost Ark* and *War of the Worlds*.

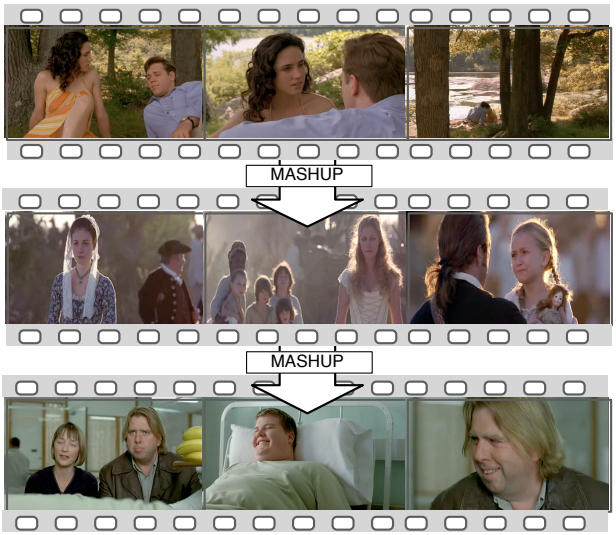

Figure 5: Mashup created by user #7 by mixing scenes from *A Beautiful Mind*, *The Patriot* and *All or Nothing*.

inforcement, but excerpted from prevailing dour movies. To start the mashup with, she chose a segment from *A Beautiful Mind* (102 *s*), when the two main characters fall in love. Then she selected an excerpt from *The Patriot*, the third proposed by the system: a touching scene when the main characters are leaving their families heading for war. After 45 *s* she decided to switch to a segment from *All or Nothing*, the third in the ranking, which depicts a family gathered around the son who is recovering after a heart attack, where the family members are laughing and trying to get relaxed after the strain of the previous scenes. Beyond the evident continuity of the low-level features which characterise their emotional identity, all three segments involve "dialogues among lovers or relatives" as a main theme and play an essential narrative role in the respective film plots.

As a final step, users filled in a questionnaire (reported in Table 1) for rating their satisfaction with both the quality of the generated mashups and the respondence to their own expectations. Marks were given using a five-level Likert scale (*1=min/fully disagree*, *5=max/fully agree*) and the obtained evaluation results are presented in Figure 6 as a diagram of the average mark per each question.

Table 1: Selected statements from the user questionnaire.

No.	Statement
1.	The colour atmosphere does not change across the transitions between mashup segments
2.	The sensation of motion does not change across the transitions between mashup segments
3.	There is no brisk change in audio between different mashup segments
4.	The general mood of the whole mashup is coherent
5.	Rate the whole experience of creating the mashup
6.	Using such a tool stimulated my creativity
7.	The final product responds to my expectations

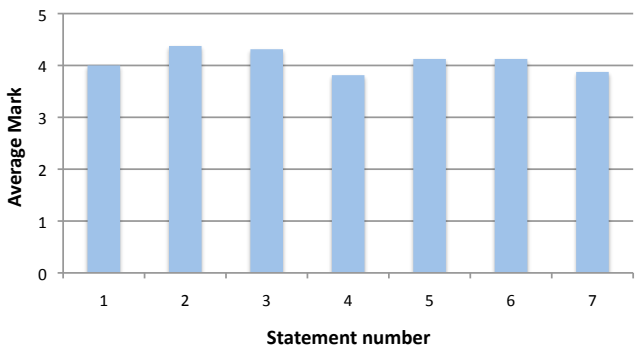

Figure 6: Average mark per question for evaluating user satisfaction with the mashup tool and experience.

By evaluating the answers to the user test, we observed a substantial agreement among users regarding the ability of the system to create mashups which are in general continuos in terms of colour atmosphere and sound. On the other hand, we noticed that when the pace of the selected segments is high, the users perceive a continuity in the sensation of motion which is stronger if compared to that of mashups with a slower rhythm.

As a concluding remark, all users judged the whole experience of creating a mashup as a positive recreational activity which stimulates their creativity. Additional positive and negative aspects about the tool, together with personal comments, were also gathered from users and will be soon used as potential improvements to be integrated in the next release of the mashup application.

5. CONCLUSIONS

In this paper we propose a framework for the automatic creation of new multimedia content by editing and remixing pre-existing material taken from a set of feature movies. The system guides the user through the mixing task by providing a selection of emotionally similar scenes, which have been previously characterized by using a space of emotional identity. The proposed approach, instead of trying to achieve a cognitive coherence of the final product, focuses more on the continuity of the evoked mood on the emotive level. The performed user test shows that the mashup application successfully assists both professional and non-expert users in the editing of new video products, stimulating their creativity along the different phases of the recreational process.

REFERENCES

[1] P. Aufderheide and P. Jaszi. Recut, reframe, recycle: Quoting copyrighted material in user-generated video. Technical report, Center for Social media, American University, 2007.

[2] B. Brown. *Cinematography, Theory and Practice*. Elsevier, 2002.

[3] M. Campanella. *Balancing automation and user control in a home video editing system*. PhD thesis, Eindhoven University of Technology, 2009.

[4] L. Canini, S. Benini, P. Migliorati, and R. Leonardi. Emotional identity of movies. In *Proc. of ICIP*, Cairo, Egypt, 7-11 Nov. 2009.

[5] C. T. Castelli. Trini diagram: imaging emotional identity 3D positioning tool. In *Proc. of SPIE*, Dec. 1999.

[6] O. Gallagher. Video recuts and the remix revolution: Whose rights are being infringed? - the development of an online remix community known as total recut. Available at www.totalrecut.com/masters.doc, 2008.

[7] A. Girgensohn, J. Boreczky, P. Chiu, J. Doherty, J. Foote, G. Golovchinsky, S. Uchihashi, and L. Wilcox. A semi-automatic approach to home video editing. In *Proc. of ACM UIST*, pages 81–82, San Diego, California, USA, 2000.

[8] A. Hanjalic and L.-Q. Xu. Affective video content representation and modeling. *IEEE Transactions on Multimedia*, 7(1):143–154, Feb. 2005.

[9] X.-S. Hua, Z. Wang, and S. Li. Lazycut: content-aware template-based video authoring. In *Proc. of ACM MM*, Singapore, 6-11 Nov. 2005.

[10] M. Kumano and Y. Ariki. Automatic useful shot extraction for a video editing support system. In *Proc. of the MVA*, Nara, Japan, 11-13 Dec. 2002.

[11] A. Mehrabian. Pleasure-arousal-dominance: a general framework for describing and measuring individual differences in temperament. *Current Psychology: Developmental, Learning, Personality, Social*, 14:261–292, 1996.

[12] W.-T. Peng, Y.-H. Chiang, W.-T. Chu, W.-J. Huang, W.-L. Chang, P.-C. Huang, and Y.-P. Hung. Aesthetics-based automatic home video skimming system. In *Proc. of MMM*, Kyoto, Japan, 9-11 Jan. 2008.