Browsing Multilingual Making-Ofs

Carlos Teixeira[1], *Ana Respício*[2] *and Catarina Ribeiro*[1]

[1]LASIGE
[2]OR Center University of Lisbon, DI
1749 - 016 Lisbon, Portugal
{cjct, respicio}@di.fc.ul.pt, catarina.ribeiro@gmail.com

Abstract

The present work describes a new film player that enriches cinematographic experience and boost film-viewer interaction. A multilingual subtitle time alignment algorithm provides natural browsing links between film and the corresponding making-of. Results are presented for a set of known English spoken films, also using the corresponding Portuguese and Spanish subtitles.

Index Terms: multilingual processing, subtitle alignment, making-of alignment, interactive cinematography.

Introduction

In the past few years, films and related add-ons have become an important resource for the SLT (Speech and Language Technology) area. The production and use of subtitles provides a clear bridge between the video signal processing, including audio and speech, and textual annotation suitable for NLP (Natural Language Processing) and related applications: automatic subtitle translation [1], parallel corpora construction based on subtitles [2][3] and information retrieval [4].

Specialized techniques have been proposed for automatically extracting film's images [5] and shots [6]. These can be respectively be used in DVDs menus and trailers. The menus provide an intuitive interactive tool for browsing the film from the "outside" of the film. However, no attempt is found so far in the opposite direction: using the film itself to browse additional information. To our knowledge this paper presents the first prototype for evaluation of this type of interaction.

The development of applications for film content analysis [7][8], browsing [9], and skimming [5] has been a challenge in the past few years. These applications can be useful for film production, for someone who wants to study the film from different perspectives or even just to provide a faster and better understanding of the film for a common user [10]. Technological developments, the increasing availability of movie related data, as well as the deployment of new standards for multimedia navigation have lead to the development of specialized techniques to approach movie content analysis and skimming [4]. New research lines have been focused on innovative approaches to integrate additional multimodal media, such as the ingredients (previous inspirational writings, screenplay) and sub-products from film production (subtitles, making-of, interviews, and writings about the film) [10]. This paper describes a prototype implementing one of those possible integrations: the making-of merged into the correspondent film. A making-of is a documentary film about the film production. Making-ofs and other similar film add-ons are having increasing importance in the DVD market: "DVD add-ons have become much more expansive, even reflexive, with consumers/fans willing to pay much higher prices for deluxe versions of their favorite films. Films are now frequently shot specifically to include material for add-ons, with the foreknowledge that studios will probably earn an increasing percentage of their profits from a growing catalogue of DVDs" [11].

The implementation of the proposed prototype encloses two main problems: the synchronization of the film with the related parts in the making-of; and the availability of an interactive film player, able to integrate two video streams, alternatively presented according to viewer decisions. The solution found for the first problem assumes that both film and making-of are subtitled in the same language. This allows the use of text alignment techniques for video synchronization. A similar approach was recently presented for film enrichment [12]. Additional decisions should be made in order to find: first, if there is a relevant making-of scene to be presented; secondly, find the most relevant making-of scene to be presented. At the present time, our research has pursuit a multilingual approach, only requiring a few specific language lexica, such as characters and locations names. The quality and the consistency of the subtitles have revealed to be more important than the used language. The second problem also reutilizes some of the strategy used in previous author's work [12]. However the need for interaction and the visioning of a second film raises additional problems, namely while maintaining the accessibility using the web.

This paper is organized in five sections. Next section describes the approach adopted for the first problem: finding time anchors for synchronization of the film with the related parts in the making-of. Section 3 describes the solution found for the second problem: the interactive film player integrating the making-of visioning. Section 4 presents results obtained in two known English spoken films. Finally, the last section presents conclusions and future work.

Time alignment

The making-of often includes small shots from the film using the same subtitles in both videos. These are considered here as reliable anchors for the required alignment, since such subtitle matching can be robustly detected. However, the order of these shots in the making-of often depends on a different time structure. A making-of usually contains interviews from film cast, having a specific time structure, sometimes according to the specific agenda of the interviewer or even the answers from the cast. Accordingly, our approach considers similarity between film and making–of subtitles independently of their original order in any of those contents. However, film shots fill a small percentage of the making-of duration. Some parts of the making-of contents may not even be related to any specific scene in the film. This is the case when, for instance, the cast is asked about personal issues about themselves, their family or friends.

This section briefly describes the proposed approach to obtain the alignments between film and making-of subtitles according to the main blocks in Figure 1.

Before being coded into a vector token, the subtitles from

both film and making-of are both normalized in a preprocessing stage. Every sequence of alphanumeric characters (letters or digits) occurring between white spaces or punctuation marks is considered a word. All letters are converted to lowercases. Each word has an associated multiplying weight according to its importance for the alignment. Stop words receive lower weights. Proper names, such as the character names, receive higher weights. The current system does not consider multiword expressions, very frequent in languages, and whose correspondence between languages is not always obvious, and their inclusion has not yet been tested.

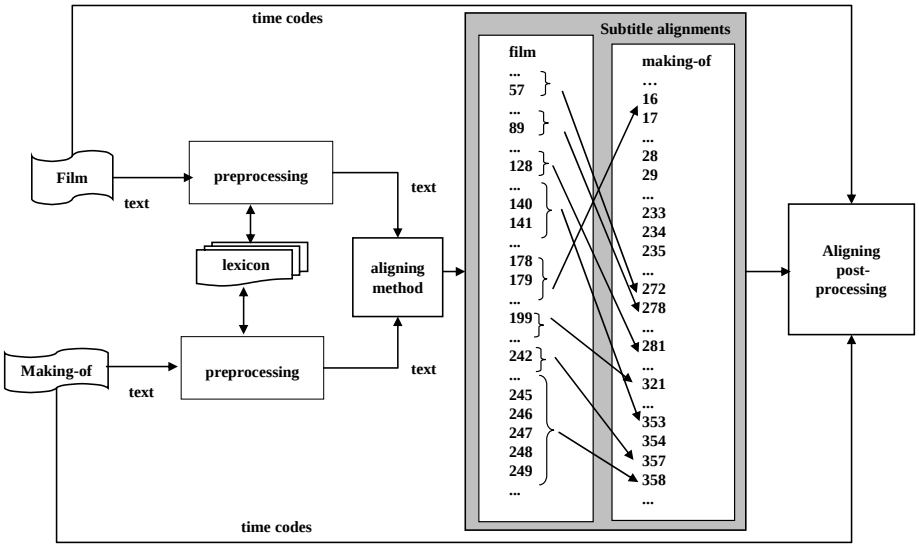

Figure 1: *Aligning system*.

The output of the aligning method is represented in the "subtitle alignments" block of Figure 1. The arrows provide the time anchors that will allow switching the viewing between the film and the correspondent making-of. These time anchors are built exclusively between the subtitles of each video stream. In the future, we plan to consider additional information such as content-based features [7][8], and other related contents such as the screenplay.

The subtitle alignment is done, one by one, evaluating the similarity measure of the well-known vector space model. In this model, each subtitle is represented by a token vector representing the word occurrences in the lexical space. Similarities between subtitles are ranked by evaluating the deviation of angles between the correspondent token vectors. This is equivalent to compute the cosine of the angle between the corresponding vectors. Consequently, the similarity between the *i*-th film subtitle and the *j*-th making-of subtitle is given by equation (1).

$$S(i,j) = \frac{film_i \bullet mkof_j}{\|film_i\| \cdot \|mkof_{ji}\|} \qquad (1)$$

where $film_i$ is the vector representing the *i*-th subtitle of the film and $mkof_j$ is the vector representing the *j*-th subtitle from the making-of. The same measure was successfully used in a related study for screenplay alignment [12].

The alignment method will first compute a similarity matrix for all the subtitles. Considering that each row represents a film subtitle and each column a making-of subtitle, a maximum value is found for each row and the corresponding subtitle index memorized in a list. At this stage every subtitle from the film will have an assignment to a making-of subtitle. However, most of these links will show a very low similarity. It was found empirically that a threshold for similarity above 0.80 will include a significant number of links, with a very low rate of wrong assignments (less than 2 per film).

Each subtitle assignment between the film and the making-of is considered here as an anchor. The subtitle time codes are the links between the results from the alignment algorithm and the video signal. More than this, once asked by the film viewer, the player should be able to decide if there is no relevant scene in the making-of or, in the opposite case, where to start in the making-of timeline.

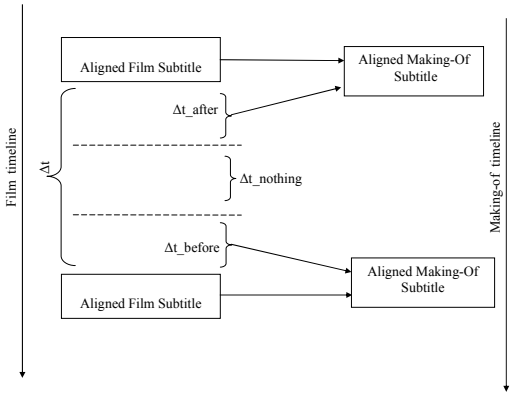

Figure 2: *Improving the making-of film coverage.*

Our post-processing module implements a simple strategy depicted in Figure 2. The time duration between two film aligned subtitles (Δt) is divided into two time segments (2).

$$\Delta t_{after} = \Delta t_{before} = \Delta t / 2. \qquad (2)$$

Within the first segment (Δt_after), any making-of request will drive the viewer to the making-of instant of the previous aligned subtitle. Within the second segment (Δt_before), the video stream will be switched to the making-off time instant of the next aligned subtitle. However, time between two consecutive aligned subtitles in the film is sometimes very long including several different scenes (more than 20 sec.). This will confuse the viewer that can be switched to some making-of scenes which were only related to film scenes already forgotten (Δt_after is excessively long) or not yet seen (Δt_before is excessively long). In order to overcome this problem, both Δt_after and Δt_before were limited to a maximum of 10 seconds. The remaining film between these two segments (Δt_nothing) will not be allowed to redirect to the making-of video.

This simple strategy is expected to be improved in the near future, integrating more data into the decisions about the boundaries for the above mentioned segments.

A system like the one described above can hardly detect every semantic relation between the film and the making-of. Some of the aimed relations can even only be explained by aesthetic reasons which can be very difficult to be automatically detected. In the near future, we will improve our system to include advanced knowledge models. Nevertheless, we expect that manual corrections will still be necessary. However, in order to demonstrate the performance of our automatic methods, all the results presented in this paper were directly collected from these methods without any direct human intervention.

User Interface

This section describes a prototype implemented for the demonstration of our new multimedia object, for testing the new developed methods and for usability tests. Figure 3 describes the components of the user interface.

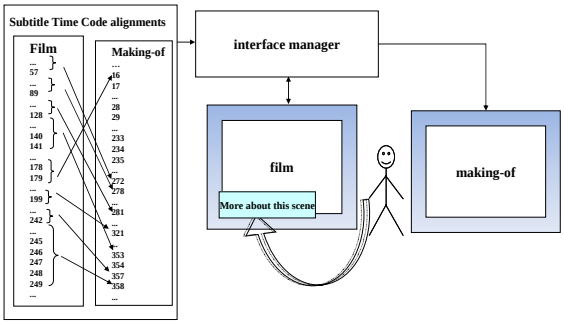

Figure 3: *User interface: playing the film and making-of video streams.*

The user scenario encompasses a film session using our prototype as a common video player. However, this film viewer has an additional button named "more about this scene". This button, once pressed, will pause the current film and start a new video: the making-of, in a second video player. However, instead of starting from the making-of beginning, the video will start in an instant related to the scene just previously seen in the film. The user can later close the making-of player and return to the paused film for continuing the previous film session. The state transition network of these events can be found in Figure 4.

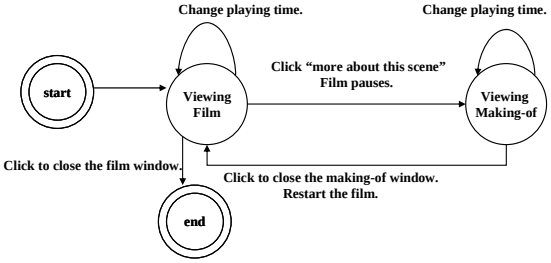

Figure 4: *State transition network.*

Both players and the single input button are controlled by the interface manager module. This module reads a pre-computed table of time assignments, every time the user requests to see more about the current scene of the film. A snapshot of the transition from the film to the making-of using the presented prototype is shown in Figure 5.

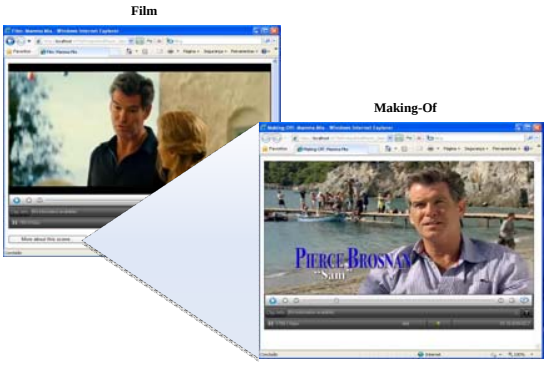

Figure 5: S*napshot of the transition from the film to the making-of.*

Results

Figures 6 and 7 display the distribution of aligned subtitles, along normalized time, for two English films, considering subtitles in three languages: English (EN), Portuguese (PT), and Spanish (SP). These results were obtained with the similarity threshold set to 0.80.

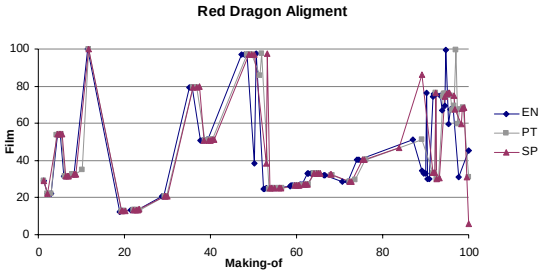

Figure 6: *Aligned subtitles for "Red Dragon".*

The alignments for "Red Dragon" in Figure 6 are almost coincident for the three languages. These indicates that the main dialogue lines, that are similar in film and making-of, are well identified and aligned. The exceptions found reflect one of two situations. In the first one, English subtitles are, in this case, mainly aimed at the hearing impaired thus displaying all sounds as subtitles (e.g. "*[Suspenseful instrumental music]"*).

This fact adds around 200 pseudo-subtitles to the English version which reflects on the chart as alignment "delays" when comparing to the two Iberian languages. In the second situation, other non coincident alignments reflect small differences in translation or dialogue lines that do not appear in one of the languages.

The alignment of "Red Dragon" reveals the references in the making-of interviews to scenes of high tension and action, mainly concentrated in the following times of the film: 13-15% – Jack Crawford talks with Will Graham about Hannibal Lecter, 30-35% and 70-76% – Will visits Lecter at prison, 60-68% – Francis Dolarhyde kills Freddy Lounds, and 95-100% – climax scene where Francis tries to kill Will's family.

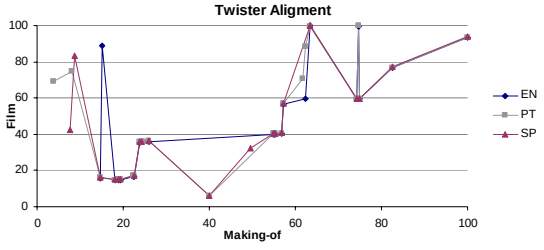

Figure 7: *Aligned subtitles for "Twister"*.

For the film "Twister", Figure 7, there are less alignment points than for "Red Dragon", since the making-of of "Twister" is mainly based on interviews and backstage images, containing few references to film scenes. Nevertheless, similarly to the "Red Dragon" results, the alignments are almost coincident for the different languages. The exceptions are mainly explained by: small translation discrepancies between the film and the making-of; differences in the subtitles due to different amounts of text - number of words and letters which has to be divided across a different number of subtitles; and textual signs in the film that were also translated into the Iberian subtitles, thus increasing the number of non-native language subtitles. The references to film scenes, in the of making-of time, are mainly concerned with setting Dorothy, the machine for twister analysis – 18%, in Helen Hunt interview; the divorce of the main characters – 20-22%, in Bill Paxton interview; approaching the storm – 58-62%, in comments about Jan De Bont; and comments about Helen Hunt and Bill Paxton entering the storm – around 78%.

The discussed results show clearly that the proposed system is almost language independent and can be used for films subtitled in different languages.

Conclusions

We presented results for two known English spoken films. Additional films have been tested. Preliminary results reveal several interesting findings: how film shots are added into the making-of; an informal test, with a set of five users, found the prototype accurate and useful. Our research has pursuit a multilingual approach, only requiring a few specific language lexica. We tested English, Portuguese and Spanish subtitles versions for the same films. The quality and the consistency of the subtitles have revealed to be more important than the used language. For instance, using subtitles from different writers, having different regional origins, can cause severe performance degradation. These findings reveal that the proposed system is almost language independent.

A prototype was developed implementing a new film player for those who want to know more about the stories behind the scenes. In the short term, we will run formal usability tests with questionnaires through the internet. Other future developments consider further linguistic refinements and the use of knowledge models for establishing additional anchors between the film and the making-of.

Acknowledgments

This work was partially supported by FCT through the Large-Scale Informatics Laboratory and Operations Research Center – POCTI/ISFL/152 – from the University of Lisbon. The authors acknowledge the comments of three anonymous reviewers about a previous version of the paper.

References

[1] Armstrong, S.; Way, A.; Caffrey, C.; Flanagan, M.; Kenny, D. and O'Hagan, M.; "Improving the quality of automated DVD subtitles via example-based machine translation", Proc. of Translating and the Computer 28, London, 2006.

[2] Tiedemann. J., "Building a multilingual parallel subtitle corpus", Proceedings of 17th CLIN, Leuven, Belgium, 2006

[3] Tiedemann, J., "Synchronizing Translated Movie Subtitles", Proceedings of the Sixth International Conference on Language Resources and Evaluation, LREC 2008, Marrakech, Morocco, May 28-30, 2008.

[4] Lee, Jia-Hong; Cheng, Jia-Shin and Wu, Mei-Yi, "A Multimedia Tool for English Learning Using Text Retrieval Techniques on DVD Subtitles", Intelligent Tutoring Systems 2006: 701-703, 2006.

[5] Li, Y., Lee, S.H., Yeh, C.-H., Kuo, C.-C.J.: Techniques for Movie Content Analysis and Skimming: Tutorial and Overview on Video Abstraction Techniques. IEEE Signal Processing Magazine. 23, 79-89, 2006.

[6] Smeaton A., Lehane, B., O'Conor N., Brady, C., Craig, G.,:Automatically Selecting Shots for Action Movie Trailers, MIR'06, 231-238, 2006.

[7] Guo, G., Li, S.Z.: Content-Based Audio Classification and Retrieval by Support Vector Machines, In IEEE Transactions on Neural Networks, 14/1, 209-215, 2003.

[8] Smeulders, A., Worring, M., Santini, S., Gupta, A., Jain, R.: Content-based Image Retrieval: the End of the Early Days, In IEEE Transactions Pattern Analysis and Machine Intelligence, 1349-1380, 2000.

[9] Hammoud, R., Mohr, R.: Interactive Tools for Constructing and Browsing Structures for Movie Films, In: MULTIMEDIA '00 - eighth ACM International Conference on Multimedia, 497-498, ACM, New York, 2000.

[10] Teixeira, C., "Multimedia Enriched Digital Books", ACM 17th Conference on Information and Knowledge Management proceedings, ACM, BooksOnline'08 Workshop on "Setting the Roadmap for Research Advances in Large Digital Book Repositories", Napa Valley, CA, USA, 2008.

[11] Brereton, P., Editorial: The consumption and use of DVDs and their add-ons. Convergence: The International Journal of Research into New Media Technologies, 13 (2). ISSN 1748-7382, 2007.

[12] Teixeira, C., Respício, A.: See, Hear or Read the Film. In Ma L., Rauterberg M. and Nakatsu R. (eds.) Entertainment Computing – ICEC2007, LNCS, vol. 4740, 271-281. Springer, Berlin/Heidelberg, 2007.