

StoViz: Story Visualisation of TV Series

Philippe Ercolessi
IRIT - Université Paul Sabatier
118 route de Narbonne
31062 Toulouse, France
ercolessi@irit.fr

Hervé Bredin
LIMSI-CNRS
B.P. 133
91403 ORSAY CEDEX,
France
bredin@limsi.fr

Christine Sénac
IRIT - Université Paul Sabatier
118 route de Narbonne
31062 Toulouse, France
senac@irit.fr

ABSTRACT

Recent TV series tend to have more and more complex plot. They follow the lives of numerous characters and are made of multiple intertwined stories. In this paper, we introduce STOVI_Z, a web-based interface allowing a fast overview of this kind of episode structure, based on our plot de-interlacing system. STOVI_Z has two main goals. First, it provides the user with a useful overview of the episode by displaying each story separately and a short abstract extracted from them. Then, it allows an efficient visual comparison of the output of any automatic plot de-interlacing algorithm with the manual annotation in terms of stories and is therefore very helpful for evaluation purposes. STOVI_Z is available online at <http://stoviz.niderb.fr>.

Categories and Subject Descriptors

H.3.1 [Information Systems]: Information storage and Retrieval—*Content Analysis and Indexing*

General Terms

Experimentation

Keywords

plot de-interlacing, visualization, evaluation, non-linear browsing

1. INTRODUCTION

The following paper presents our contribution to the technical demo session of ACM Multimedia conference. Based on the idea that TV series – which tend to have more and more complex plot, with numerous characters and multiple intertwined stories – are already segmented into narrative themes in post-production, we present a system able to discover and to show through a video browser the structure of an episode.

The system consists of two main elements: an offline pre-processing step and an online video browser.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ACM Multimedia 2012 Nara, Japan

Copyright 20XX ACM X-XXXXX-XX-X/XX/XX ...\$10.00.

The pre-processing step proceeds by segmenting the episode into shots, then by grouping related consecutive shots into semantic *scenes* and finally by clustering those scenes into *stories* (a story is a group of scenes, not necessarily contiguous, showing a strong semantic relation). In addition, an abstract of the episode is generated. The online video browser then relies on the output of the pre-processing step to present a video navigation interface to the user.

We have called *plot de-interlacing* the process of clustering scenes as it aims at extracting interlaced stories from a single video, each of them focusing on different semantic information of the whole video plot. As this structuring approach is mostly relevant for fictional videos, our work focuses on recent TV series which have multiple threads of narration. We chose to illustrate the behavior of our system on a set of 14 episodes from 2 different TV series (*Ally McBeal* and *Malcolm in the Middle*).

2. MOTIVATIONS AND NOVELTY

We designed this web-based interface to illustrate the results of our plot de-interlacing approach, and to provide a tool for users willing to have a quick look at the narrative structure of an episode, or willing to follow the evolution of a particular character or a particular story.

Moreover, since both stories and scenes are subjective concepts, the evaluation of systems for automatic segmentation into scenes or stories is not straightforward. It is why STOVI_Z also aims at helping this evaluation process by showing a graphical representation of the video, its manual scene or story segmentation, and the stories detected automatically. A numerical error rate (similar to the *Diarization Error Rate* used in speaker diarization) is thereby provided in order to quickly gain insight into the quality of the automatic approach.

As another goal of our video structuring system is to improve the quality of abstraction techniques, STOVI_Z also presents the resulting abstract and allows the user to browse it efficiently.

From all video browsers available, STOVI_Z can be compared to JOKE-O-MAT [3], as JOKE-O-MAT tends to emphasize particular events in an episode, by allowing a browsing punchline by punchline, and letting a user to follow desired features (such as characters) in the punchlines. However, the main difference is that JOKE-O-MAT is mostly based on a segmentation system, whereas STOVI_Z clusters segments into homogeneous stories.

3. SCIENTIFIC CONCEPTS

3.1 Multimodal descriptors

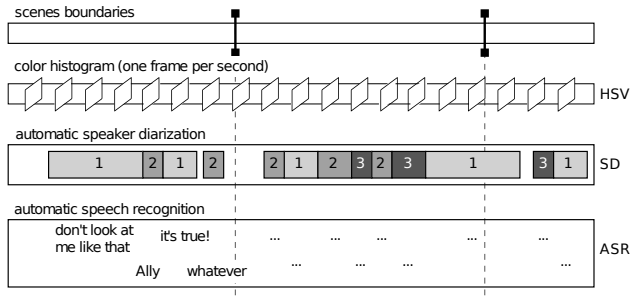


Figure 1: Set of available modalities

Looking for semantic information in the structure of a video, video segments are described using three different modalities: color histogram, speaker presence and automatic speech recognition, illustrated in Figure 1.

Based on these modalities, distances between video segments are computed to perform a clustering both for scene segmentation [1] and plot de-interlacing [2].

3.2 Plot de-interlacing

Plot de-interlacing relies on the clustering of scenes into semantically homogeneous stories [2].

First, episodes are automatically classified into two groups by looking into communities of detected speakers: *Characters-driven Episodes* (CdE) where stories follow a specific set of characters, and *Semantic-driven Episodes* (SdE) where stories are determined by a specific topic rather than a set of characters. Then, the applied clustering method depends on the type of episode: *simple agglomerative clustering* for CdE with distance between scenes based solely on the *speaker* modality, or *multimodal spectral clustering* for SdE.

4. VIDEO BROWSER

The technical demo proposed here is a web-based interface allowing users to browse a video and to visualize the video structure extracted by the plot de-interlacing and abstraction systems.

4.1 Overview

The video browser is composed of three areas, as shown in Figure 2. Area **A** is the control panel. It allows the user to select the desired episode, to modify the content of area **C**, by selecting the annotation to be displayed (manual or automatic) and to play the abstract. Area **B** is the video player. Area **C** is the timeline zone. In this area, each story is displayed as a single timeline, composed of several little rectangles, each one corresponding to a scene of the video.

4.2 Story display area

This area shows the selected timelines. A timeline corresponds to a subset of the original video: one story, the abstract or the whole video.

Figure 3 shows how a timeline is displayed. The sequence in the timeline is composed of a set of scenes of the video. Each scene is described by a keyframe which is enlarged

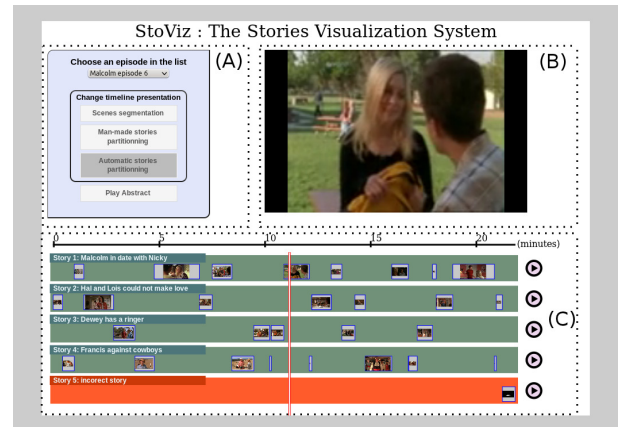


Figure 2: StoViz web-based interface



Figure 3: Timelines

when the mouse goes over it, and a colored rectangle whose width depends on the scene duration.

When the presentation mode is set to *Automatic story partitioning*, the color of each scene changes based on its status: gray if the scene is correctly clustered, red if the scene is part of the wrong story, or yellow if the scene has been clustered correctly but belongs to more than one story and has been assigned to only one.

Each timeline shows a short text describing the story it contains. A full description, made by annotators, is also accessible by moving the mouse over this text.

5. CONCLUSION

STOVIZ is a web-application that allows a user to browse a video by following a specific story which takes place into its narration. It is also a very helpful piece of software for researchers interested in the plot de-interlacing domain, as it allows a very effective comparison between manual and automatic plot de-interlacing.

As the plot de-interlacing system is generic and does not rely on models, it is easy to imagine that STOVIZ could be used to browse other types of video than TV series such as TV news, movies, or any types of video providing a structure that can be dissociated into several stories as, for example, a news video archive in order to summarize topics.

6. REFERENCES

- [1] H. Bredin. Segmentation of TV Shows into Scenes using Speaker Diarization and Speech Recognition. In *ICASSP 2012, IEEE International Conference on Acoustics, Speech, and Signal Processing*, Kyoto, Japan, March 2012.
- [2] P. Ercolessi, H. Bredin, and C. Sénac. Toward Plot De-Interlacing in TV Series using Scenes Clustering. In *IEEE CBMI 2012*, 2012.
- [3] G. Friedland, L. R. Gottlieb, and A. Janin. Joke-o-mat: browsing sitcoms punchline by punchline. In W. Gao, Y. Rui, A. Hanjalic, C. Xu, E. G. Steinbach, A. El-Saddik, and M. X. Zhou, editors, *ACM Multimedia*, pages 1115–1116. ACM, 2009.