1 Introduction

Semi-automated solutions can transform the images of a film into a comic strip. Whereas the conversion of movies [Hwang et al. 2006] and 3D games [Shamir et al. 2006] to comics has been addressed before, we propose novel ways of user interaction and—most prominently—leverage much information from a movie’s screenplay. Thanks to standard screenplays’ highly structured “studio format,” we can employ textual analysis at several stages:

- The line counts of the scenes in the screenplay are used to estimate every scene’s temporal placement in the film.
- The scenes’ dialogue is extracted from the screenplay and turned into speech balloons placed sequentially.
- Verbal directions such as (whispers) are recognized and mirrored by corresponding type and speech balloon styles.
- If a character’s dialogue is interrupted by actions, its different parts are used to form a double-burger style speech balloon.
- Offscreen speech is put into a speech balloon emanating from a panel’s side; a voice-over is placed in a rectangle.
- Verbs in action lines such as The VASE crashes onto the floor. are recognized and turned into noise balloons.

2 Features

The first step is to segment the film into scenes to coarsely align the dialogue of the screenplay with the images. Since most scenes comprise lots of cuts, their boundaries cannot be found directly by looking for pixel-wise discontinuities. Thus, our software detects scenes in the screenplay’s text to build a first guess on the timing. An incremental process that clusters HSL color histograms adjusts the scene boundaries in an iterative fashion.

The software extracts a user-defined number of frames per minute from the film and places each of them in its own “panel” on the comics’ pages. The frame’s portion that is to be used can be zoomed and panned interactively. Initially, the panel reveals as much of the frame as possible without showing blank areas. The user can adjust how many rows and columns each panel spans on the page’s grid. The layout is updated automatically, preserving temporal order.

Dialogue and action from the screenplay are used to create speech and noise balloons; the directions given in the screenplay control the font and the balloon’s shape. Since a screenplay only describes what is visible, the user may add thought bubbles by hand. Every bubble is a parametric vector shape that can be edited interactively.

Since the software cannot determine the speaker’s location, the user has to specify the endpoint of each speech balloon’s prong. The screenplay reveals which dialogue lines belong to one character, so all prongs for a character can be set with one mouse click per panel.

As an aid in placing the speech balloons, the software attempts to detect faces by standard image processing. The layout algorithm leaves faces unoccluded and works in reading order, never allowing a balloon to extend past the upper edge of an earlier one in the same panel. Non-fitting dialogue spills over to the next panel. Any balloons remaining at the end of the scene are placed in an extra window, ready to be dragged onto the panels.

The frames can be cartoonized analogously to [Winnemöller et al. 2006]. The result is exported as a PDF file.

3 Results

To load and process a DVD-quality short movie of twelve minutes’ length our software takes less than a minute on a standard PC. This includes scene detection, face detection, and the generation and placement of balloons. Depending on the personal preferences, the user may spend a quarter of an hour to a full hour on fine-tuning the automatic result. The scene boundary detection requires only little user intervention. For instance, in the movie “Ocean’s Eleven,” the method finds 46 of the 80 scene boundaries among the extracted frames precisely; it is off by one frame for another 11 boundaries.

References

