Names and faces linking in video

Overview

We have developed software for automated video annotation where we link names found in the subtitles or fans transcripts to faces or objects in the video.

In depth description

In a first case we name faces in soap series by using the weak supervision of narrative texts that describe the events in the video and that are drafted by fans. Several unsupervised methods that operate without any manual labelling of exemplar faces, and methods that use a limited number of labelled exemplars are presented and evaluated. All methods exploit the multiple co-occurrences between faces shown in the video and names mentioned in the texts to compute the strength of the linking and reinforce this coupling by means of an Expectation Maximization algorithm. We show that the unsupervised methods attain competitive results without any prior human supervision. The results show F1 values between 80% and 86% for the recognition of the face-name pairs without any human supervision. These figures rise only slightly when a number of faces were manually labelled beforehand. This technology is developed in collaboration with the Computer Vision group of KU Leuven.

In a second setting we have recognized animals in videos using subtitles. In this framework, the alignment between animals and their names is performed again using an Expectation Maximization algorithm which is adapted to two very different circumstances- 1) when the bounding boxes are available and 2) when the frame as a whole is used instead of bounding boxes. With the goal of maximizing precision, recall and F-measure, the experiments compare a multitude of natural language processing approaches and visual features when associating animal names in the subtitles with visual patterns. The proposed unsupervised methods obtain 83.1% F1 using bounding boxes and 65.7% F1 without bounding boxes in a fully automated pipeline.

Potential fields of application

This technology can be included as a part of a media search engine, where it provides better automated indexing of the video data.

Possibilities for exploitation

The technology is further developed in the national project PARIS (IWT-SBO-110067) where we link textual and visual data in user generated Web content and Web shops.

Further Information

Further technical information is available in Pham, P.T., Deschacht, K., Tuytelaars, T. & Moens M.-F. (2013). Naming Persons in Video: Using the Weak Supervision of Textual Stories. In *Journal of Visual Communication and Image Representation*, 24 (7), 944-955.

and Dusart, T., Nurani Ventikasubramanian, A. & Moens, M.-F. (2013). Cross-Modal Alignment for Wildlife Recognition. In *Proceedings of the 2013 ACM International Workshop on Multimedia Analysis for Ecological Data* (pp. 9-14). ACM.

Contact person

Prof. Marie-Francine Moens
Department of Computer Science
Celestijnenlaan 200A
B-3001 Heverlee, BELGIUM
sien.moens@cs.kuleuven.be