

Evaluation of a Video Annotation Tool Based on the LSCOM Ontology

Emilie Garnaud, Alan F. Smeaton and Markus Koskela

Abstract—In this paper we present a video annotation tool based on the LSCOM ontology [1] which contains more than 800 semantic concepts. The tool provides four different ways for the user to locate appropriate concepts to use, namely basic search, search by theme, tree traversal and one which uses pre-computed concept similarities to recommend concepts for the annotator to use. A set of user experiments is reported demonstrating the relative effectiveness of the different approaches.

Index Terms—Video annotation, ontology, LSCOM, semantic concept distances.

I. INTRODUCTION

In visual media processing, a lot of progress has been made in automatically analysing low level visual features in order to obtain a description of the content. However, annotations by humans are still often needed to extract accurate deep semantic information from within. Indeed manual tagging of visual content has become widespread on the internet through what is known as “folksonomy” in which human annotators provide descriptive content tags.

One of the challenges in the area of human annotation is generating consistency across annotations in terms of both the vocabulary used and the way it is used. The common approach here is to provide users with an ontology, or an organisation of allowable semantic tags or concepts. This is popular in enterprises such as photo and video stock archives where only a small number of people actually perform the annotation and thus they are familiar with the ontology and the way it is used. In more open-ended applications such as social tagging or tagging by untrained users then ontologies are regarded as too restrictive and too hard to learn in a short period of time and so such applications favour free form tagging at the expense of the consistency the use of an ontology brings.

Here, we address the issue of how an untrained user could use a pre-defined ontology to index video in the domain of broadcast TV news. Specifically, we use the LSCOM ontology [1], of about 850 concepts to help index media by semantics.

II. VIDEO ANNOTATION TOOL

Traditional annotation tools based on a lexicon or ontology usually provide a full list of concepts with no, or very poor ways to navigate it. This works quite well for a small lexicon or for users who are trained to use it, but this is not scalable to a larger ontology or the case where the users are untrained. Thus in order to use the LSCOM or any other large ontology

to index video, we need to support different ways for the user to navigate it in order to complete the annotation process.

In our annotation tool there are four distinct ways to annotate content, described as follows.

A. Basic search

An alphabetically-ordered list of the ontology and a search box to find matching concepts is provided which is simple but effective when users have a good knowledge of the ontology.

B. Search by themes

More than 700 concepts of the ontology have been arranged into 19 different themes such as *Arts & Entertainment, Business & Commerce, News, Politics, Wars & Conflicts* ... so an annotator can search for a concept by first selecting a theme that seems to fit with the shot.

C. Recommended concepts

In previous work introduced in [2] we computed similarity among all pairs of concepts in the LSCOM ontology using a combination of usage co-occurrence as the ontology was used to index a corpus of 80 hours of video, combined with visual shot-shot (and by implication, annotation-annotation) similarities. We used these concept-concept co-occurrences to generate “recommended concepts” at any point after annotation by at least 1 concept. This worked by determining the 15 concepts most similar to the *set* of concepts already used to annotate a shot, and this top-15 was refreshed every time an additional concept was used in annotating a shot.

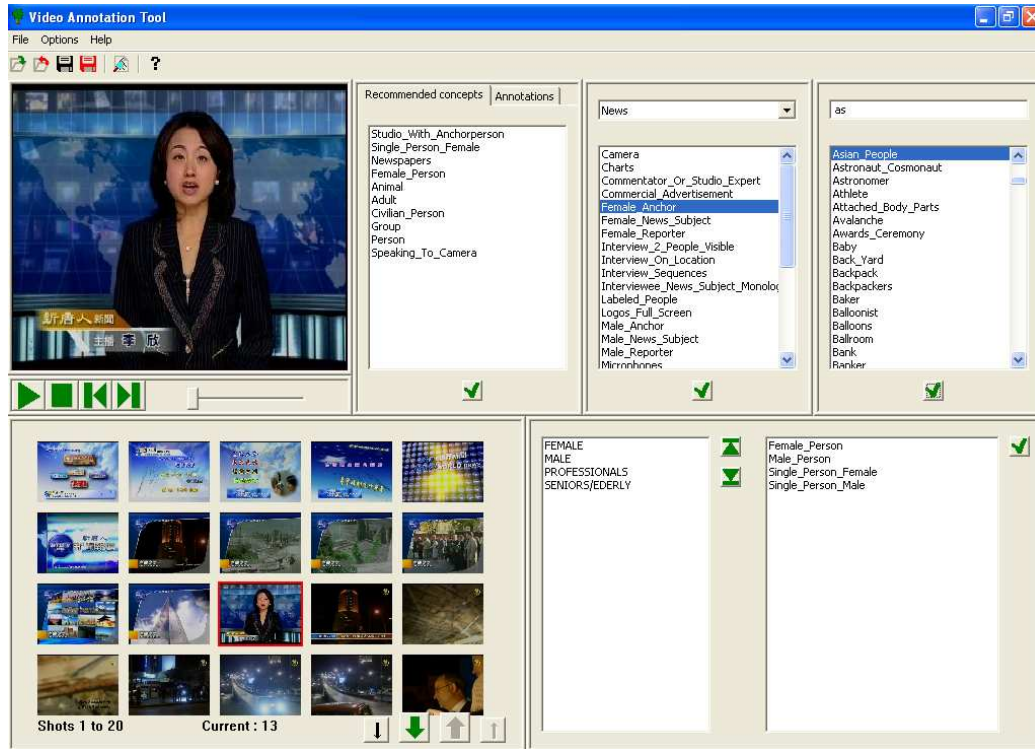
D. Tree organization

An hierarchical version of the ontology has recently been completed so we introduced some of its elements in our tool by creating an area where a user can navigate among different trees of the ontology.

III. EXPERIMENTS AND ANALYSIS

We performed preliminary experiments involving 10 native English-speaking users who each annotated 40 shots using different functionalities of the tool, either in a restricted timeframe or with unlimited time to complete. To replicate the scenario of an untrained user annotating material on the internet, our users did not receive any special training in using the annotation tool. Shots to be annotated were selected randomly and people used functionalities in a Latin squares protocol so as not to bias the results. We analyzed four

E. Garnaud is with Institut EURECOM, 2229, Route des Cretés, BP 193 – 06904 Sophia Antipolis Cedex, France and A. Smeaton and M. Koskela are with the Centre for Digital Video Processing, Dublin City University, Glasnevin, Dublin 9, IRELAND. email: Alan.Smeaton@dcu.ie



different aspects of the annotation process namely the overall time spent on annotating, the number of annotations per shot, the shot annotation rate, and the number of annotations during the first minute. Results are shown below. The best annotation

	Search Only	Search + Themes	Search + Recmd.	Entire Tool
Average time per shot	1m 53	2m 06s	1m 53s	1m 59s
# annotations per shot (Avg)	6.9	7.2	11.3	10.9
Annotation rate	6.1	5.8	10.1	9.2
Avg annotations in 1 st minute	6.3	5.2	7.7	7.7

performance is obtained using the “recommended concepts” feature because the time spent in free annotation is the same as the “search only” version (representing the traditional approach) but the number of annotations is greater when recommendations are used. Using the “themes” feature seems to slow down the annotation process without increasing the number of annotations, probably due to a lack of knowledge of the ontology and the way concepts had been organised into different themes. Also, some shots are really good for annotation by themes but others are not, which is why they are a good complement to searching for concepts to annotate.

We also found an unexpected result from the “entire tool” experiment which surprisingly doesn’t seem to be the most effective ! Once more, this seems to be due to a lack of knowledge of the tool by users. Our whole point of using untrained users is to replicate the common situation of untrained users annotating resources on the internet. If we examine the number of annotations done during the first minute then “recommended concepts” and “entire tool” have

the same performance but after the first minute people lost time searching the ontology for additional concepts as they did not have enough knowledge to know when to stop as searching the ontology does not provide any kind of closure to the process.

IV. CONCLUSIONS AND FUTURE WORK

The approach of using recommended concepts as a way of annotating seems to be promising though the size of our experiment is small. The “recommended concepts” could be improved by collecting more data to link associated concepts. Indeed, some associated concepts are really good (like “store”, “landlines”, “bank”, “office” and “female_person” for “administrative_assistant”) but some others are not, such as (“harbors”, “boat_ship”, “business_people”, “canal” and “lakes” for “house_of_worship”).

The tool seems to be powerful for various user profiles. For beginners, it helps them to learn the ontology and for experts it provides a way to annotate concepts that they are not used to annotating which improve their knowledge of the ontology.

ACKNOWLEDGMENT

This work was supported by Science Foundation Ireland under grant 03/IN.3/I361, by the EC under contract FP6-027026 (K-Space) and by IRCSET.

REFERENCES

- [1] M. Naphade, J.R. Smith, J. Tesic, S.-F. Chang, W. Hsu, L. Kennedy, A. Hauptmann and J. Curtis. *Large-Scale Concept Ontology for Multimedia*, IEEE Multimedia, 13(3) July-Sept, 2006, pp.86–91.
- [2] M. Koskela and A.F. Smeaton. *Clustering-Based Analysis of Semantic Concept Models for Video Shots* In Proc. IEEE International Conference on Multimedia & Expo (ICME 2006). Toronto, Canada. July 2006.