Evaluation of a Video Annotation Tool Based on the LSCOM Ontology

Emilie Garnaud
Institut EURECOM, 2229, Route des Crêtes
BP 193 – 06904 Sophia Antipolis Cedex France

Alan F. Smeaton
Center for Digital Video Processing & Adaptive Information Cluster
Dublin City University, Glasnevin, Dublin 9, Ireland

***Abstract*— In this paper we present a video annotation tool based on the LSCOM ontology [2] 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.**

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 [1].

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, and the common approach here would be 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 can be 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 would bring.

In this paper we address the issue of how an untrained user could use a pre-defined ontology to index video content in the domain of broadcast TV news. Specifically, we use the LSCOM ontology [2], which contains about 850 concepts.

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 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

The tool provides a full list of the ontology and an edit box where the user can type a word to search for a matching concept. This is very simple but really 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 selecting a theme that seems to fit with the shot.

C. Recommended concepts

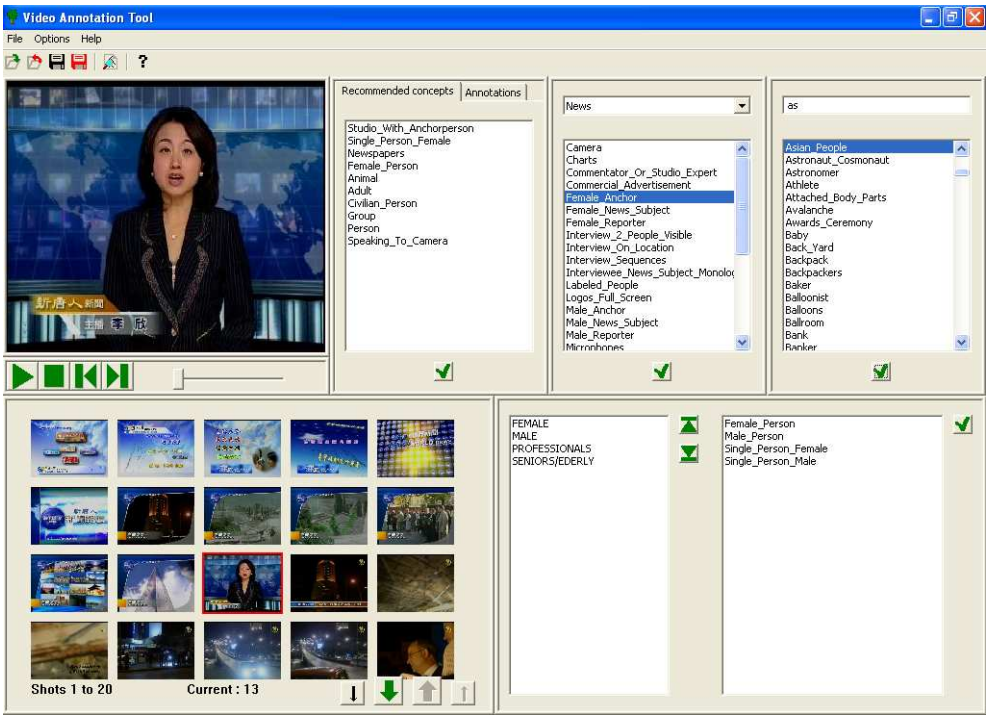
In previous work [3] we computed similarity among all pairs of concepts in the LSCOM ontology using a combination of usage co-occurrence as the ontology was used to index a corpus of 80 hours of video, combined with actual shot-shot (and by implication, annotation-annotation) similarities. We used these concept-concept similarities 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 hierachical version of the ontology has recently been completed so we tried to introduce 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 experiments involving 10 native English-speaking users who each annotated 80 shots using different functionalities of the tool, either in a restricted timeframe or with unlimited time to complete, and after a short training period. 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 different aspects of the



annotation process namely the time spent on annotating, the number of annotations per shot, the ouput and the number of annotations during the first minute. Results are shown below.

	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
Output	6.1	5.8	10.1	9.2
Annotaions per minute (Avg)	6.3	5.2	7.7	7.7

The best annotation 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 we believe these 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 whole ontology for additional concepts as they did not have enough knowledge to know when to stop because the 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. 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 and by the EC under contract FP6-027026 (K-Space).

References

[1] Folksonomy. `http://en.wikipedia.org/wiki/Folksonomy`, Last visited August 2006.

[2] 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.

[3] M. Koskela, A.F. Smeaton and G. Gaughan. *Semantic Analysis of Concept Models for News Videos*, Proceedings of VCIMS - Workshop on Visual Categorisation and Image Management Systems, Sunderland, U.K. 2006.