AUTOMATIC EXTRACTION OF SEMANTIC PREFERENCES FROM MULTIMEDIA DOCUMENTS[1]

Manolis Wallace, Phivos Mylonas and Stefanos Kollias

Department of Computer Science
School of Electrical and Computer Engineering
National Technical University of Athens, Greece
fmylonas@image.ntua.gr

ABSTRACT

One of the most important topics in modern multimedia research is the treatment of documents and users at a semantic level. In this framework, the automated extraction of semantic preferences from multimedia content is an important problem. This paper is part of our ongoing work in the field of semantic multimedia analysis and retrieval; it extends on previous work on scene and shot detection, contour extraction and object tracking, descriptor extraction and matching, and semantic document analysis, in the direction of automated extraction of semantic user preferences. Such preferences can then be utilized towards the personalization of the multimedia retrieval process. The methodology of the paper is based on the utilization of a fuzzy relational knowledge representation model and a novel definition of semantic context.

1. INTRODUCTION

It is a common fact that multimedia retrieval is by far more difficult to tackle than text retrieval, as in this case it is more difficult to match user requests to available documents. This is why the role of user profiles in multimedia retrieval is much more important [3]. The focus of technological attempts in the field of combining user interests with audiovisual archives and multimedia documents can be divided into two major areas. The analysis of a multimedia document for the extraction of the topics related to it and the extraction of both preferences from the analyzed documents.

The problem of analyzing the content of a multimedia document is quite different than that of analyzing a textual document and a lot more complicated. Firstly, the entities to be indexed are not directly encountered in the document; recognizable features must be extracted and matched to respective ones found in the knowledge base. Secondly, a multimedia document contains objects and events, whose relations are spatiotemporal, rather than purely grammatical. Finally, abstract concepts, such as "sports" and "arts" are not directly encountered in multimedia documents, and they must be inferred from the concrete objects and events, as well as features (such as light) which are not attributed to a particular object or event.

In this paper we build upon previous and ongoing research work performed in the fields of object detection [4] and semantic document analysis [8]. Based on these we attempt to cluster documents in usage history in a meaningful manner and thus extract the user's semantic preferences.

2. SEMANTIC INDEX CONSTRUCTION

The required semantic index needs to be constructed in an automated manner from the archived multimedia documents. This task may be achieved using techniques such as the ones initiated in [4]. Several algorithms have been implemented for detecting semantic entities using the mpeg signal of a video document. They consist mainly of the following parts: shot detection, characteristic shot analysis, moving object detection, object feature detection and description extraction, matching with AV description of semantic entities. Scene detection can be considered as the first stage of a non-sequential (non-linear) video representation. For this reason, scene cut detection algorithms are first applied by video indexing and retrieval systems to extract characteristic frames and shots on which video queries can be applied.

Regarding the object localization and tracking problem, an active contour model is used, trying to improve the procedure's performance in terms of both accuracy and computational cost. These initial results are used as input in a contour estimation method, in order to extract the objects' accurate contours. After extracting the bounding polygons and/or the contours of the detected moving objects in a video segment, some visual descriptors can be utilized to characterize the captured objects or regions, according to the MPEG-7 framework [5]. Following this stage, matching of these descriptors to ones stored in a semantic knowledge base will be used as a means for automatic detection of events and objects.

As a result, the semantic index could be generated via automatically recognizing objects and events in a multimedia document and mapping them to semantic entities [1]. This issue is a complicated, still open problem. Similar input, though, can be acquired via textual analysis of the structured textual information contained in the metadata that accompany annotated multimedia documents. So, in order to provide input feed to our algorithm, we still use semi-automatic index generation for multimedia documents and automatic index generation for textual documents, as the primary step towards the document analysis.

Figure 1. Contour detection in a soccer sequence

3. DOCUMENT ANALYSIS

This section of the paper refers to the analysis of the above automatically generated semantic index, with the aim of extracting a document's semantics. More formally, we accept as input the semantic indexing of available documents, i.e. the semantic index I. In the latter, each document is represented as a normal fuzzy set d on the set of semantic entities. Based on this set, and the knowledge available to the system in the form of semantic relations [6], we aim to detect the degree to which a given document is indeed in related to a thematic category t. We will refer to this degree as $R_T(t,d)$. In designing an algorithm that is able to calculate this relation in a meaningful manner, a series of issues need to be tackled

1. A semantic entity may be related to multiple topics.
2. A document may be related to multiple topics.
3. The semantic index may contain incorrectly recognized entities.

In the following, keeping these issues in mind the proposed approach may be decomposed into the following steps [7]:

- Create a single semantic relation that is suitable for use by the thematic categorization module.
- Determine the count of distinct topics that a document is related to, by performing a partitioning of semantic entities, using their common meaning as clustering criterion.
- Fuzzify the partitioning, in order to allow for overlapping of clusters and fuzzy membership degrees.

- Identify the topic that is related to each cluster.
- Aggregate the topics for distinct clusters in order to acquire an overall result for the document.

In the process of content analysis we will have to use the common meaning of semantic entities. We will refer to this as their *context*. A document d is represented by its mapping to semantic entities, via the semantic index. Therefore, the context of a document is again defined via the semantic entities that are related to it [9].

Using the height of the context as a similarity metric and applying an agglomerative clustering procedure, each resulting cluster is described by the crisp set of semantic entities $c \subseteq {}^{0+}d$ that belong to it, where ${}^{0+}d$ is the support of d, i.e. the crisp set of entities contained in it. A semantic entity s should be considered correlated with cluster c, if it is related to the common meaning of the semantic entities in it. Therefore, the aforementioned correlation is measured as:

$$C_c(s) = \frac{h(K(c \cup \{s\}))}{h(K(c))}$$

where $K(\cdot)$ is the context of a set of semantic entities, as defined in [9]. Using this as a classifier, we may expand the detected crisp partitions, as to include more semantic entities. Partition c is replaced by cluster c^{fuzzy}:

$$c^{fuzzy} = \sum_{s \in {}^{0+}d} s / C_c(s)$$

The process of fuzzy hierarchical clustering has been based on the crisp set ${}^{0+}d$, thus ignoring fuzziness in the semantic index. In order to incorporate this information when calculating the "final" clusters that describe a document's content, we adjust the degrees of membership for them as follows:

$$c^{final}(s) = t(c^{fuzzy}(s), I(s,d)) \ , \forall s \in {}^{0+}d$$

where t is a t-norm [2]. The semantic nature of this operation demands that t is an Archimedean norm. Each one of the resulting clusters corresponds to one of the distinct topics of the document.

In order to determine the topics that are related to a cluster c^{final}, two things need to be considered: the scalar cardinality of the cluster $\left| c^{final} \right|$ and its context. Since context has been defined only for normal fuzzy sets, we need to first normalize the cluster as follows:

$$c^{normal}(s) = \frac{c^{final}(s)}{h(c^{final}(s))}, \forall s \in {}^{0+}d$$

Obviously, semantic entities that are not contained in the context of c^{normal} cannot be considered as being related to the topic of the cluster. Therefore

$$R_T^*(c^{normal}) = w(K(c^{normal}))$$

where w is a weak modifier. Clusters of extremely low cardinality probably only contain misleading entities, and therefore need to be ignored in the estimation of $R_T(d)$. On the contrary, clusters of high cardinality almost certainly correspond to the distinct topics that d is related to, and need to be considered in the estimation of $R_T(d)$. The notion of "high cardinality" is modeled with the use of a "large" fuzzy number $L(\cdot)$. $L(a)$ is the truth value of the preposition "a is high", and, consequently, $L(|b|)$ is the truth value of the preposition "the cardinality of cluster b is high".

The set of topics that correspond to a document is the set of topics that correspond to each one the detected clusters of semantic entities that index the given document.

$$R_T(d) = \bigcup_{c^{final} \in G} R_T(c^{final})$$

where $\cup$ is a fuzzy co-norm and G is the set of fuzzy clusters that have been detected in d. The topics that are related to each cluster are computed, after adjusting membership degrees according to scalar cardinalities, as follows:

$$R_T(c^{final}) = R_T^*(c^{normal}) \cdot L(|c^{final}|)$$

4. SEMANTIC PREFERENCES EXTRACTION

As far as the main guidelines are concerned, the extraction of semantic preferences from a set of documents, given their topics, is quite similar to the extraction of topics from a document, given its semantic indexing. Specifically, the main points to consider may be summarized in the following:

▪ A user may be interested in multiple topics.
▪ Not all topics that are related to a document in the usage history are necessarily of interest to the user.

These issues are tackled using similar tools and principles, as the ones used to tackle the corresponding problems in content analysis. Thus, once more, the basis on which the extraction of preferences is built is the context. The common topics of documents are used in order to determine which of them are of interest to the user and which exist in the usage history coincidentally.

Moreover, since a user may have multiple interests, we should not expect all documents of the usage history to be related to the same topics. Quite the contrary, similarly to semantic entities that index a document, we should expect most documents to be related to just one of the user's preferences. Therefore, a clustering of documents, based on their common topics, needs to be applied. In this process, documents that are misleading (e.g. documents that the user chose to view once, just to find out that they do not contain anything of interest to him) will probably

not be found similar with other documents in the usage history. Therefore, the cardinality of the clusters may again be used to filter out misleading documents.

What is common among two documents $d_1, d_2 \in D$, i.e. their common topics, can be referred to as their common context. This can be defined as

$$K(d_1, d_2) = R_T(d_1) \cap R_T(d_2)$$

A metric that can indicate the degree to which two documents are related is, of course, the height of their common context. This can be extended to the case of more than two documents, in order to provide a metric that measures the similarity between clusters of documents:

$$Sim(c_1, c_2) = h(K(c_1, c_2))$$
$$K(c_1, c_2) = \bigcap_{d \in c_1 \cup c_2} R_T(d)$$

Sim is the compatibility indicator for the clustering of documents in H^+ ($H = \{H^+, H^-\}$ is the usage history, comprised of documents H^+ that the user has indicated interest for and documents H^- for which the user has indicated some kind of dislike). Using this criterion a hierarchical clustering process is applied on documents in H^+, using a threshold on the similarity as the termination criterion.

The topics that interest the user and should be classified as preferences are the ones that characterize the detected clusters. Degrees of preference can be determined based on the following parameters:
▪ The cardinality of the clusters. Clusters of low cardinality should be ignored as misleading.
▪ The weights of topics in the context of the clusters. High weights indicate intense interest.

Therefore, each of the detected clusters c_i is mapped to a positive interest as follows, where the notion of "high cardinality" is modeled with the use of a "large" fuzzy number $L(\cdot)$:

$$U_i^+ = L(c_i) \cdot K(c_i)$$
$$K(c_i) = \bigcap_{d \in c} R_T(d)$$

The information extracted so far can be used to enrich user requests with references to topics that are of interest to the user, thus giving priority to related documents. What it fails to support, on the other hand, is the specification of topics that are known to be uninteresting for the user, as to filter out, or down-rank, related documents. In order to extract such information, a different approach is required and negative interests should be verified by the repeated appearance of topics in documents of H^-.

$$U^- = \sum s_i / u_i^-$$

$$u_i^- = L(\sum_{d \in H^-} R_T(s_i, d))$$

The representation and handling of preferences using fuzzy sets is developed to a greater extent in [10].

5. RESULTS

A sample knowledge base has been created, containing more than 1000 entities, few of them accompanied by their low level MPEG-7 descriptors, and seven distinct semantic relations. Some of the semantic entities in the knowledge base have been characterized as thematic categories. Document indexing based on textual annotation has been automatic, whereas detection of events and objects has been performed in a semiautomatic manner.

	d_1	d_2	d_3	d_4	d_5	d_6	d_7
arts	0.84	0.85			0.75		
cinema	0.36	0.86			0.76		
theater	0.89	0.27			0.20		
football		0.77	0.84	0.67			
war		0.77					0.55
medicine				0.82	0.64	0.91	0.91

Table 1. Thematic categorization of documents. Values below 0.1 have been omitted

Document d_1 contains a shot from a theater hall. The play is war-related. On the other hand, document d_2, is more interesting, as it contains a sequence of shots from a news broadcast. Due to the diversity of stories presented in it, the semantic entities that are detected and included in the index are quite unrelated to each other:

d_2= (sitting person)/0.9 + (army or police uniform) /0.8 + lawn/0.5 + goal/0.9 + tier/0.7 + speak/0.9 + goalkeeper/0.8 + shoot/0.5 + performer/0.7 + seat/0.9 + curtain/0.7 + scene/0.8 + tank/0.9 + missile/0.8 + explosion/0.9 + river/1

As can be seen in Table 1, the algorithm successfully identifies the existence of more than one distinct topics in the document.

As far as the last step of our method is concerned, given the set of documents that we already presented, our algorithm successfully identifies following preferences:

$$U_1^+ = arts / 0.75 + cinema / 0.36 + theater\, 0.20$$

$$U_2^+ = football / 0.67$$

$$U^- = war / 0.64 + medicine / 0.25$$

assuming that $U^+ = \{d_1, ..., d_5\}$ and $U^- = \{d_6, d_7\}$.

6. CONCLUSIONS

This paper is part of our ongoing work in the field of semantic multimedia analysis and retrieval. It extended on previous work on scene and shot detection, contour extraction and object tracking, descriptor extraction and matching and semantic document analysis, in the direction of automated extraction of semantic user preferences. Such preferences can then be utilized towards the personalization of the multimedia retrieval process.

The techniques of this paper are based to a great extend on the utilization of fuzzy relational knowledge representation [6], fuzzy algebra [2] and a novel definition of semantic context [9]. Using the latter, a document to document similarity measure may be defined using the height of their common context. Thus, documents may be clustered; each cluster that is not too small to be considered reliable corresponds to an distinct user interest.

The methodology presented in this paper has been developed in the framework of the EU IST-1999-20502 FAETHON project [11] for the analysis of multimedia usage history towards the automated generation of semantic user profiles.

7. REFERENCES

[1] Zhao, R. and W.I. Grosky, Narrowing the Semantic Gap-Improved Text-Based Web Document Retrieval Using Visual Features, IEEE Trans. on Multimedia, Special Issue on Multimedia Database, Vol. 4, No 2, June 2002

[2] G. Klir and B. Yuan, Fuzzy Sets and Fuzzy Logic, Theory and Applications, New Jersey, Prentice Hall, 1995.

[3] Angelides, M.C., Special issue on Multimedia content modeling and personalization, IEEE Multimedia 10(4)

[4] Tsechpenakis G., Akrivas G., Andreou G., Stamou G. and Kollias S., Knowledge - Assisted Video Analysis and Object Detection, EUNITE, Albufeira, Portugal, September 2002

[5] ISO/IEC JTC 1/SC 29 M4242, Text of 15938-5 FDIS Information Technology - Multimedia Content Description Interface - Part 5 Multimedia Description Schemes, 2001

[6] G. Akrivas, G. B. Stamou and S. Kollias Semantic Association of Multimedia Document Descriptions through Fuzzy Relational Algebra and Fuzzy Reasoning IEEE Transactions on Systems, Man, and Cybernetics, part A, accepted for publication.

[7] Wallace, M., Akrivas, G. and Stamou, G. Automatic Thematic Categorization of Documents Using a Fuzzy Taxonomy and Fuzzy Hierarchical Clustering FUZZ-IEEE 2003, St. Liouis, MO, USA, May 2003

[8] Wallace, M., Akrivas, G., Mylonas, P., Avrithis, Y., Kollias, S. Using context and fuzzy relations to interpret multimedia content CBMI, IRISA, Rennes, France, September 2003

[9] Akrivas, G., Wallace, M., Andreou, G., Stamou, G. and Kollias, S. Context – Sensitive Semantic Query Expansion ICAIS, Divnomorskoe, Russia, 2002

[10] Wallace, M., Akrivas, G., Stamou, G. and Kollias, S. Representation of user preferences and adaptation to context in multimedia content -- based retrieval, SOFSEM 2002, Milovy, Czech Republic, November 22-29, 2002

[11] Avrithis Y. and Stamou G., FAETHON: Unified Intelligent Access to Heterogenous Audiovisual Content, Proceedings of VLBV, Athens, Greece, Oct. 2001