A Region Thesaurus Approach for High-Level Concept Detection in the Natural Disaster Domain

Evaggelos Spyrou and Yannis Avrithis

Image, Video and Multimedia Systems Laboratory,
School of Electrical and Computer Engineering
National Technical University of Athens
9 Iroon Polytechniou Str., 157 80 Athens, Greece,
espyrou@image.ece.ntua.gr,
WWW home page: http://www.image.ece.ntua.gr/∼espyrou/

Abstract. This paper presents an approach on high-level feature detection using a region thesaurus. MPEG-7 features are locally extracted from segmented regions and for a large set of images. A hierarchical clustering approach is applied and a relatively small number of region types is selected. This set of region types defines the region thesaurus. Using this thesaurus, low-level features are mapped to high-level concepts as model vectors. This representation is then used to train support vector machine-based feature detectors. As a next step, latent semantic analysis is applied on the model vectors, to further improve the analysis performance. High-level concepts detected derive from the natural disaster domain.

1 Introduction

High-level concept detection in both image and video documents remains still an unsolved problem. However, due to the continuously growing volume of audiovisual content, this problem attracts a lot of interest within the multimedia research community. Its two main and most interesting aspects appear the selection of the low-level features that will be extracted and the approach that will be used for assigning these low-level descriptions to high-level concepts, a problem commonly referred to as the "Semantic Gap".

There exist plenty of works towards the solution of this problem. In [1] a multimedia analysis and retrieval system using multi-modal machine learning techniques in order to model semantic concepts in video is presented. Moreover, in [2], a region-based approach in content retrieval that uses Latent Semantic Indexing (LSI) techniques is presented. In [3] the features are extracted by segmented regions of an image. Also, in [4], a region-based approach is presented, that uses knowledge encoded in the form of an ontology. Moreover, a hybrid thesaurus approach for semantic object recognition and identification within video news archives is presented in [5]. Finally, a lexicon, used in an approach for an interactive video retrieval system is presented in [6].

Fig. 1. An input image and its coarse segmentation.

2 Low-Level Feature Extraction

For the representation of the low-level features, descriptors from the MPEG-7 standard [7] were used and more specifically, the *Color Layout Descriptor*, the *Scalable Color Descriptor*, the *Color Structure Descriptor* and the *Homogeneous Texture Descriptor*. For the extraction of the aforementioned descriptors, the MPEG-7 eXperimentation Model (XM)[8] was used.

Instead of extracting descriptors globally, the color and texture descriptors are extracted from image regions. A multiresolution variation of the RSST color segmentation algorithm [9] is first applied, tuned to produce a coarse segmentation. This way, one can intuitively describe the image with respect to the image segments. To explain this, an input image along with its coarse segmentation is depicted in figure 1. In this example, one could easily describe the input image as a set of regions. Here a user could see "a light blue region" (sky), "two green regions" (vegetation), "an orange region" (fire) etc. Then, the MPEG-7 descriptors are extracted locally, from each image region.

3 Region Thesaurus Construction

Given the entire set of images and their extracted low-level features as described in section 2, it is rather obvious that regions belonging to similar semantic concepts, also have similar low-level descriptions. Also, images that contain the same semantic concepts include similar regions. To exploit these observations, we try to formalize an image description in terms of the regions it is consisted of.

A hierarchical clustering algorithm [10] is applied on the low-level descriptions of all the regions that occur from segmenting all images from the training set. After this clustering, the number of clusters to keep is selected experimentally. This way, by keeping the centroids of those clusters, we select the more often encountered regions. These regions will be referred to as "region types" and form the region thesaurus. Its purpose is to formalize a conceptualization between the low and the high-level features and facilitate their association.

Each region type is represented as a feature vector that contains all the extracted low-level information for it. As it is obvious, a low-level descriptor does not carry any semantic information. On the other hand, a high-level concept

carries only semantic information. A region type lies in-between and contains the necessary information to formally describe the color and texture features, but can also be described with a "lower" description than the high-level concepts. I.e., one can describe a region type as "a green region with a coarse texture".

After this clustering procedure, we can easily observe that each cluster may or may not contain regions from the same high-level feature and regions from the same high-level feature may be encountered in more than one clusters. For example, the high-level concept *vegetation* can have more than one instances differing in i.e. the color of the leaves of trees. Moreover, in a cluster that contains instances from the semantic entity i.e. *sea*, these instances could be mixed up with parts from i.e. *sky*.

4 Image Analysis

Having calculated the distance of each region of the image to all the words of the constructed thesaurus we construct a model vector to semantically describe the visual content of the image, in terms of the region thesaurus. This vector is formed by keeping the smaller distance for each region type among all image regions. The distances are calculated using the MPEG-7 XM [8] and linearly combined. Let: $d_i^1, d_i^2, ..., d_i^j, i = 1, 2, 3, 4$ and $j = N_C$, where N_C is the number of region types and d_i^j the distance of the i-th region of the clustered image to the j-th region type. Then, the model vector D_m is described by equation 1.

$$D_m = [min\{d_i^1\}, min\{d_i^2\}, ..., min\{d_i^{N_C}\}], i = 1, 2, 3, 4 \quad (1)$$

Using these model vectors to describe low-level image features, we train one Support Vector Machine for each high-level concept. We also perform experiments using a Latent Semantic Analysis [11](LSA) approach. This way, we try to exploit the relationships between the high-level concepts and the region types they contain more often. Images correspond to documents and region types to terms. This way, the co-occurrence matrix is formed. Then, matrices $\mathbf{\Sigma}$ and $\mathbf{U}$ are determined using the training set of images. The value of k is selected experimentally. Finally, each input model vector is driven to the concept space. This way, the model vectors are transformed and used to train the concept detectors.

5 Experimental Results

For the evaluation of the presented framework a dataset[1] from various images collected from the world wide web. This set consists of approximately 600 images from the following semantic classes: *fire, rocks, smoke, snow, trees, water*. A separate detector was trained for each concept. Results are shown in table 1, before and after the application of Latent Semantic Analysis. The presented approach was also tested on TRECVID 2007 [12] development data for the same concepts. Considering the complexity of the TRECVID data, the results appear promising.

[1] Special thanks to Javier Molina for sharing his collection.

Table 1. Accuracy for all 6 concepts. Set 1 contains images collected from the web, set 2 from TRECVID 2007 development data.

Concept	Set 1 Without LSA	Set 1 With LSA	Set 2 Without LSA	Set 2 With LSA
Fire	76.0%	84.0%	20.0%	46.0%
Rocks	69.5%	73.9%	21.0%	63.0%
Smoke	60.0%	64.0%	-	-
Snow	80.9%	90.5%	46.0%	72.0%
Vegetation	94.7%	89.4%	31.0%	47.0%
Sea	65.3 %	72.0%	-	-
Sky	72.0 %	75.0%	41.0%	47.0%

6 Acknowledgements

The work presented in this paper was partially supported by the European Commission under contracts FP6-027026 K-Space and FP6-027685 MESH. Evaggelos Spyrou is funded by PENED 2003 Project Ontomedia 03ED475.

References

1. IBM: MARVEL Multimedia Analysis and Retrieval System. (IBM Research White paper)
2. Souvannavong, F., Mérialdo, B., Huet, B.: Region-based video content indexing and retrieval. In: CBMI 2005, Fourth International Workshop on Content-Based Multimedia Indexing, June 21-23, 2005, Riga, Latvia. (2005)
3. Saux, B., G.Amato: Image classifiers for scene analysis. In: International Conference on Computer Vision and Graphics. (2004)
4. Voisine, N., Dasiopoulou, S., Mezaris, V., Spyrou, E., Athanasiadis, T., Kompatsiaris, I., Avrithis, Y., Strintzis, M.G.: Knowledge-assisted video analysis using a genetic algorithm. In: 6th International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS 2005). (April 13-15, 2005)
5. Boujemaa, N., Fleuret, F., Gouet, V., Sahbi, H.: Visual content extraction for automatic semantic annotation of video news. In: IS&T/SPIE Conference on Storage and Retrieval Methods and Applications for Multimedia, part of Electronic Imaging symposium. (2004)
6. Snoek, C.G., Worring, M., Koelma, D.C., Smeulders, A.W.: Learned lexicon-driven interactive video retrieval. (2006)
7. Chang, S.F., Sikora, T., Puri, A.: Overview of the mpeg-7 standard. IEEE trans. on Circuits and Systems for Video Technology **11**(6) (2001) 688–695
8. MPEG-7: Visual experimentation model (XM) version 10.0. ISO/IEC/ JTC1/SC29/WG11, Doc. N4062 (2001)
9. Avrithis, Y., Doulamis, A., Doulamis, N., Kollias, S.: A stochastic framework for optimal key frame extraction from mpeg video databases. (1999)
10. Duda, R.O., Hart, P.E., Stork, D.G.: Pattern Classification. 2 edn. Wiley Interscience (2000)
11. Deerwester, S., Dumais, S., Furnas, G.W., Landauer, T.K., Harshman, R.: Indexing by latent semantic analysis. Journal of the Society for Information Science **41**(6) (1990) 391–407
12. Smeaton, A.F., Over, P., Kraaij, W.: Evaluation campaigns and trecvid. In: MIR '06: Proceedings of the 8th ACM International Workshop on Multimedia Information Retrieval, New York, NY, USA, ACM Press (2006) 321–330