Semantic Image Annotation via Hierarchical Classification

Nicolas Tsapatsoulis
Cyprus University of Technology
31 Archbishop Kyprianos Str.,
CY-3603, Limmasol, Cyprus

Klimis Ntalianis
National Technical University of Athens
9 Heroon Polytechniou Str.,
15780 Zografou, Greece

Abstract: In this paper we address some of the issues commonly encountered in automatic image annotation systems such as simultaneous labeling with keywords corresponding to both abstract terms and object classes, multiple keyword assignment, and low accuracy of labeling due to concurrent categorization to multiple classes. We propose a hierarchical classification scheme which is based on predefined XML-dictionaries of tree form. Every node of such a tree defines a particular classification task while the childs of the node correspond to classification categories. The winning class (subnode) defines the subsequent classification task and the process continues until the leafs of the tree are reached. The final classification task is performed at image segment level; that is every image segment is assigned a particular keyword corresponding to a tree leaf. The path followed from the root of the XML tree to the leafs along with the union of labels assigned to the image segments compose the list of annotation keywords for the input image.

Keywords: Semantic image annotation, image segmentation, region-growing, supervised learning, hierarchical classification, multimedia content description interface

1 Introduction

The last few decades we are facing an enormous increase in the number of digital image collections that are available through the Web or in personal repositories. The problem of searching these repositories created the need for efficient and intelligent schemes for content-based image retrieval (CBIR) [15]. CBIR computes relevance based on the visual similarity of image low-level features such as color, texture and shape [13]. Users can query by example and the system automatically returns images according to relevance [16]. Early CBIR systems were based on the query-by-example paradigm [6], [7], which defines image retrieval as the search for the best database match to a user-provided query image. However, it was quickly realized that the ultimate users were willing to search for images by text-based mode, and, therefore, the design of fully functional retrieval systems would require support for semantic queries [14]. In such systems images in the database are annotated with semantic labels, enabling the user to specify the query through a natural language description of the visual concepts of interest. This realization, combined with the cost of manual image labeling, generated significant interest in the problem of automatically extracting semantic descriptors from images.

Automatic annotation of digital images with semantic labels is usually coped with by utilizing machine learning emphasizing on classification. Semantic labels may refer to an abstract term, such as indoor, outdoor, athletics, etc, or to an object class such as human, car, tree, foam mats, etc. In contrary to object classes, abstract terms cannot be related to specific image regions. Instead annotation using abstract terms consider an image as whole. On the other hand, semantic labels corresponding to object classes are assigned every time a particular object instance is encounter in an image. This object instance almost always correspond to an image region (part of the image). In the literature of automatic semantic image annotation, proposed approaches tend to classify images using only abstract terms or using holistic image features for both abstract terms and object classes. The latter is obviously wrong; semantic labeling using object class labels is actually an object detection task. Therefore, region-based features must be used instead of holistic ones. It is fair to say, however, that there are approaches [19] that use some kind of object detection in order to assign semantic labels to images [3]. Namely, given a set of training images with keywords that describe image semantic contents, region-based low-level features of the training images are extracted. Then, classifiers are constructed with low-level features to give the class decision. The trained classifiers are, then, used to classify new instances and annotate unlabeled images automatically. Since each classifier is trained in the "one-vs-all" (OVA) mode (the concept of interest versus everything else), this semantic labeling framework is referred as supervised OVA. Needless to mention that both image classification using abstract terms and semantic labeling via object detection are extremely difficult to solve; the latter being solvable using strict constraints.

In this paper we propose a unified framework, based on hierarchical classification, for semantic image annotation with both abstract terms and object class labels. To do so we model the structure of semantic labels using an XML-dictionary having a tree form. Root labels correspond to abstract terms while leafs correspond to object classes. Classification is performed in various steps, each step providing different semantic labels to (automatically) choose from. In this way there is no need to handle the classification problem with an extremely large number of labels concurrently available (such a classification task is obviously highly unreliable both in terms of classifiers' training and automatic annotation accuracy). Instead, classification is performed using a few labels at a time, depending on the tree branches at the current level in the hierarchy. In order to allow for semantic labeling at the object class level we use an innovative method for image segmentation using region-growing toward "meaningful" objects. Each image segment is then evaluated against the object class models to be assigned the best matching label. Given that each image is composed -in general- of several regions it is possible one image to be assigned several semantic labels corresponding to object classes. Classification into abstract terms is achieved using holistic instead of region-based features of images. In fact, the MPEG-7 visual descriptors [8] are used for both image and image region classification; the difference is that in the first case the shape descriptor is excluded while in the second case all descriptors are computed per image segment. Concept (semantic label) models were build using the Support Vector Machines (we have used the libSVM [5] library integrated with Weka [17])

The paper is organized as follows: Section 2 introduces the overall system operations performed during automatic image annotation. In Section 3 we discuss the creation of the XML dictionary that is used for semantic annotation. In Section 4 we deal with the MPEG-7 descriptors used in concept modeling. Learning and training stuff preparation is explained in Section 5. In Section 6 we present the image segmentation process used for object detection and labeling. In Section 7 we present the evaluation protocol we have employed along with extended experimental results. Finally conclusions are drawn and further work hints are given in Section 8.

2 System Architecture

The proposed system operates on two different modes: learning and automatic annotation. Learning is involved for the creation of concept models being either abstract terms or object classes and is further explained in Section 5. Automatic image annotation uses the concept models and image processing techniques for the creation of keywords assigned to individual images. The overall annotation process is illustrated as a flowchart in Figure 1. Every time an input image is fed to the system the following procedures take place: (a) The XML-dictionary tree is loaded (details on the form of this dictionary will be given in Section 3), (b) the MPEG-7 visual descriptors of the whole image are computed (see Section 4 for more information), (c) starting from the root of the XML tree the appropriate concept models are loaded (if, for example, at root level there are the concepts 'indoor' and 'outdoor' only the models for these concepts are used), (d) the image is classified to one of the currently loaded concepts using MPEG-7 descriptors computed in step (b) (it must be mentioned here that the confidence of classification is also computed so as to identify situations where there is no match to one of the existing concepts), (e) once the leaf level of the XML tree is reached the appropriate concept class models are loaded (note that these models depend on the classifications made in the previous steps, i.e., the path followed within the XML tree), (f) input image is partitioned into large homogeneous areas using a segmentation technique which is driven by the currently loaded object class models (see Section 6 for details), (g) each image partition is evaluated against the loaded object class models and the winning class is assigned as a label to the corresponding partition (note that classification confidence is always involved to identify no match to a particular object class), and (h) the list of object class labels attached to the individual image partitions along with the abstract terms assigned to the whole image at the various levels of the XML tree compose the list of keyword labels that must be assigned to input image.

3 Hierarchy of Concepts

Dictionaries used for manual image annotation have usually a tree structure (see Figure 2 for an example). Broader concepts ('indoors', 'outdoors') correspond to higher level nodes (short distance to root) while specific items or object classes correspond to leafs or lower level nodes. We have adopted a similar structure for the dictionaries used in automatic annotation. An example of such a dictionary, used for annotating images taken from athletics, in XML format

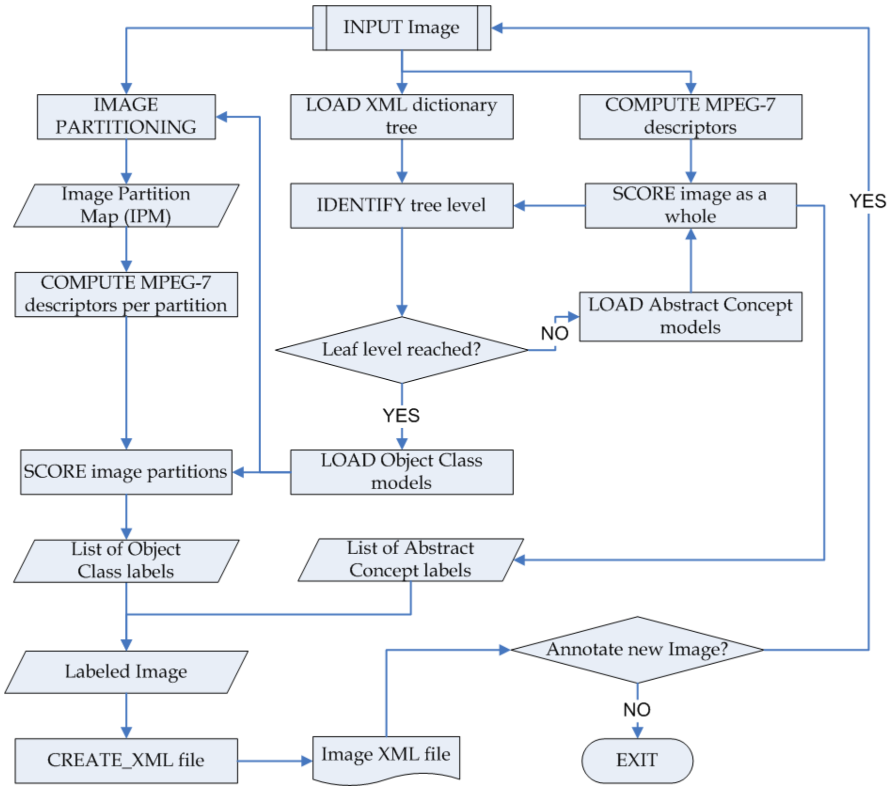

Figure 1: Flowchart of automatic image annotation

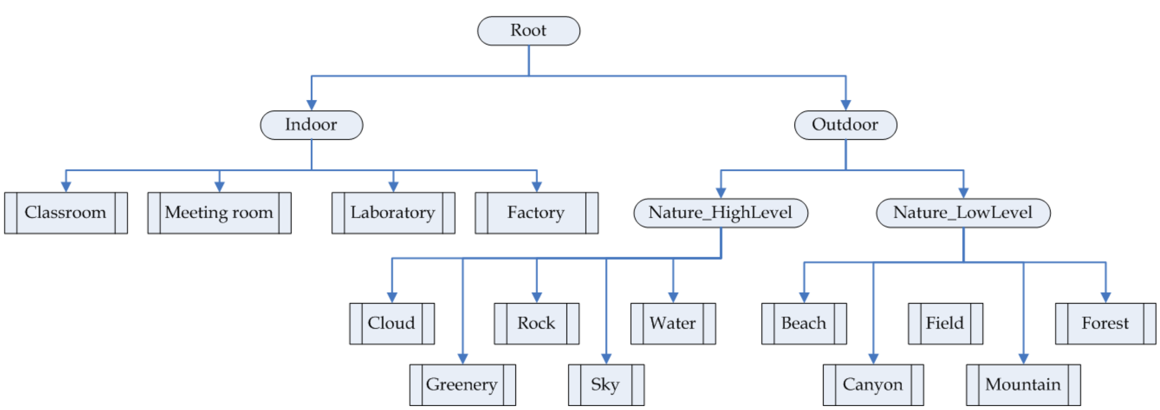

Figure 2: The tree structure of an annotation lexicon

```
- <root>
  - <IndoorEvent>
    - <Jumping>
      + <HighJump>
      - <PoleVault>
          <Athlete />
          <Pillars />
          <HorizontalBar />
          <Pole />
        </PoleVault>
      + <LongJump>
      + <TripleJump>
      </Jumping>
    - <Throwing>
      + <DiscusThrow>
      + <HammerThrow>
      + <JavelinThrow>
      + <ShotThrow>
      </Throwing>
    + <Running>
    </IndoorEvent>
  - <OutdoorEvent>
    + <Marathon>
    + <Walking>
    </OutdoorEvent>
  </root>
```

Figure 3: An example of an XML dictionary related with athletics domain

is shown in Figure 3. We will explain the hierarchical classification approach we follow with the aid of this scheme. For this purpose we assume that (a) for each node there is a trained multiclass classifier able to classify an input image (or image segment for lowest level nodes) to one of the categories defined by node childs, and (b) for each concept there is a confidence estimator mechanism (we have implement these confidence estimators using one-SVM classifiers [18]).

Let I an input image, $\underline{x} \in \Re^N$ a feature vector extracted from I, and $\Phi^y(\underline{x}) = [\phi_1^y(\underline{x})\ \phi_2^y(\underline{x})\ ...\ \phi_M^y(\underline{x})]$ a mapping function $\Re^N \rightarrow \underbrace{[0\ 1] \times [0\ 1] \times ...[0\ 1]}_{M\ times}$. If function $\Phi^y(\underline{x})$ denotes the classifier at node y the winning class is computed by:

$$c^y = argmax_i(\phi_i^y) \tag{1}$$

Let also $\psi_i^y(\underline{x})$ denote a confidence estimation function of the i-th subnode of node y ($\psi_i^y(\underline{x})$ performs a mapping $\Re^N \rightarrow [0\ 1]$). In case the confidence value corresponding to the winning class subnode $\psi_{c^y}^y(\underline{x})$ is higher than a specific threshold T (an indicative value is $T = 0.4$) the label corresponding to subnode c^y will be assigned to input image I and the next classification task will be performed on node c^y with the categorization classes corresponding to its subnodes.

Let us now present an example to clear things out. By referring to Figure 3 the first classification task involves the classes ‘Indoor’ and ‘Outdoor’. Assuming that the winning class is ‘Indoor’ and the confidence, that input image is indeed an indoor image, is higher than T then input image is assigned the label ‘Indoor’. The next classification task involves classes ‘Jumping’, ‘Throwing’, and ‘Running’ (childs of node ‘Indoor’). Assume now that the process continues following the path ‘Jumping’$\rightarrow$ ‘PoleVault’. The last classification task involves classes ‘Athlete’, ‘Pillars’, ‘HorizontalBar’, and ‘Pole’ and is applied to all image segments created by image segmentation (see Section 6). As a result it is possible all the previous labels to be assigned to input image, provided that at least one of image segments classified to each one of these categories. The last thing to mention is that the classification process stops at the tree level where a non-confident classification occurred.

4 Image and Region-based features

Automatic image annotation has gained great attention in the research community because it deals with a real world problem which is laborious to be handled with human intervention exclusively: Searching in image repositories of thousands of images which they have not got explicit metadata assigned to them by humans. In the MPEG-7 framework there is a special foresight for this problem through the definition of the MPEG-7 visual descriptors [8]. These descriptors are low-level image features proposed after an extended evaluation procedure [11]. No doubt that much of the attention paid recently to automatic image annotation and CBIR systems is due to the MPEG-7 visual content description interface, which provide a unified framework for experimentation. Furthermore, the MPEG-7 experimentation model [12] provides parctical ways for the computation of the MPEG-7 descriptors.

MPEG-7 visual descriptors include the color, texture and shape descriptor. A total of 22 different kind of features are included, nine for color, eight for texture and five for shape. The various feature types are shown in Table 1. In the third column of this Table is indicated whether or not the corresponding feature type is used in holistic image and/or image segment description during automatic annotation. The number of features shown in the fourth column in most cases is not fixed and depends on user choice; we indicate there the settings in our implementation. The dominant color features include color value, percentage and

Table 1: MPEG-7 visual descriptors used in the proposed classification scheme

Descriptor	*Type*	*# of features*	*Usage level*	*Comments*
Color	DC coefficient of DCT (Y channel)	1	Both	Part of the Color Layout descriptor
	DC coefficient of DCT (Cb channel)	1	Both	Part of the Color Layout descriptor
	DC coefficient of DCT (Cr channel)	1	Both	Part of the Color Layout descriptor
	AC coefficients of DCT (Y channel)	5	Both	Part of the Color Layout descriptor
	AC coefficients of DCT (Cb channel)	2	Both	Part of the Color Layout descriptor
	AC coefficients of DCT (Cr channel)	2	Both	Part of the Color Layout descriptor
	Dominant colors	Varies	None	Includes color value, percentage and variance
	Scalable color	16	Both	
	Structure	32	Both	They used in both holistic image and image segment description
Texture	Intensity average	1	Both	Part of the Homogeneous Texture descriptor
	Intensity standard deviation	1	Both	Part of the Homogeneous Texture descriptor
	Energy distribution	30	Both	Part of the Homogeneous Texture descriptor
	Deviation of energy's distribution	30	Both	Part of the Homogeneous Texture descriptor
	Regularity	1	None	Part of the Texture Browsing descriptor
	Direction	1 or 2	None	Part of the Texture Browsing descriptor
	Scale	1 or 2	None	Part of the Texture Browsing descriptor
	Edge histogram	80	Both	Includes the spatial distribution of five types of edges
Shape	Region shape	35	Segment	A set of angular radial transform coefficients
	Global curvature	2	Both	Part of the Contour Shape descriptor
	Prototype curvature	2	Both	Part of the Contour Shape descriptor
	Highest peak	1	Both	Part of the Contour Shape descriptor
	Curvature peaks	Varies	Both	Describes curvature peaks in term of amplitude and distance from highest peak

variance and require especially designed metrics for similarity matching. Furthermore, their length is not known a priori since they are image dependent (for example an image may be composed from a single color whereas others vary in color distribution). The previously mentioned difficulties cannot be easily handled in machine learning schemes, therefore we decided to exclude these features for the current implementation of our system. The texture browsing features (regularity, direction, scale) have not been included in the description vectors (for image and image segments) because in the current implementation of the MPEG-7 experimentation model [12] the corresponding descriptor cannot be reliably computed (it is a known bag of the implementation software). The shape descriptor features are computed only on specific image regions (they are not used in the holistic image description). The number of Peaks values of the contour shape descriptor vary depending on the form of an input object. Furthermore, they require a specifically designed metric for similarity matching because they are computed based on the HighestPeak value. For these reason they have been excluded also from the segment description vector at this stage.

5 Data preparation and Concept modeling

It is already stated that dictionary concepts for semantic annotation or retrieval can be divided into two categories: abstract terms and object classes. The latter occupy a fraction of the images that contain them. Hence, most images are a combination of various concepts and, ideally, the assembly of a training set for each semantic class should be preceded by 1) careful semantic segmentation, and 2) identification of the image regions containing the associated visual feature vectors. In practice, the manual segmentation of all database images with respect to all concepts of interest is infeasible. On the other hand, automated segmentation methods are usually not able to produce a decomposition of each image into a plausible set of semantic regions. Manual annotation (for training purposes) of images with abstract terms is much easier since it does not require explicit identification of image regions. In both situations we have used the NOTIS [9] software for manual annotation of training data. The overall work took place within the BOEMIE project [2] and for this reason we chose the athletics application domain. Athletics is a highly hierarchical domain and fits well with the proposed approach. The semantic domain as far as the images is concerned is modeled through an ontology with a structure similar to the XML dictionary shown in Figure 3. Based on this dictionary we manually annotated 1046 images containing a total of 3546 concept instances of 33 different concepts.

Concept classifiers were created using a multiclass SVM methodology [10]. The LIBSVM package of Chen and Lin [13] was employed for this purpose. The input to the SVM were the concept MPEG-7 feature vectors of the relevant images and image segments. Eighteen multiclass classifiers corresponding to the nodes (non leafs) of the XML tree were created and used for automatic classification. Ten of them were involved in object classification and the remaining were used for abstract term classification.

In addition to multiclass classifiers a total of 33 one-SVM classifiers were build for confidence estimation purposes. In particular a Support Vector Machine was trained for each concept. The margin, that is the distance of an unseen concept feature vector to the separating hyperplane, is a measure of confidence for the category membership of the respective image. By varying the acceptance threshold for the margin, precision and recall of the concept categories can be controlled. Confidence estimators were used as concept detectors; that is, irrespectively of the result of the classification task performed on a parent node the winning class is accepted as a label for the input image if the confidence of detection -for the concept corresponding to the winning class- is higher than a predefined threshold.

6 Image Segmentation

Image segmentation is required to form candidate image regions for object detection and subsequent image labeling. Although there are several image segmentation algorithms used in the image analysis research community none of them produces segments that match real objects. The majority of them create homogeneous regions that share common color or intensity properties. In order to create 'meaningful' objects during the image segmentation step we employ the confidence estimators of the concepts corresponding to object classes. Initially an input image I is decomposed into image partitions P_i, such as $I = \bigcup P_i$. Image partitioning is achieved with the

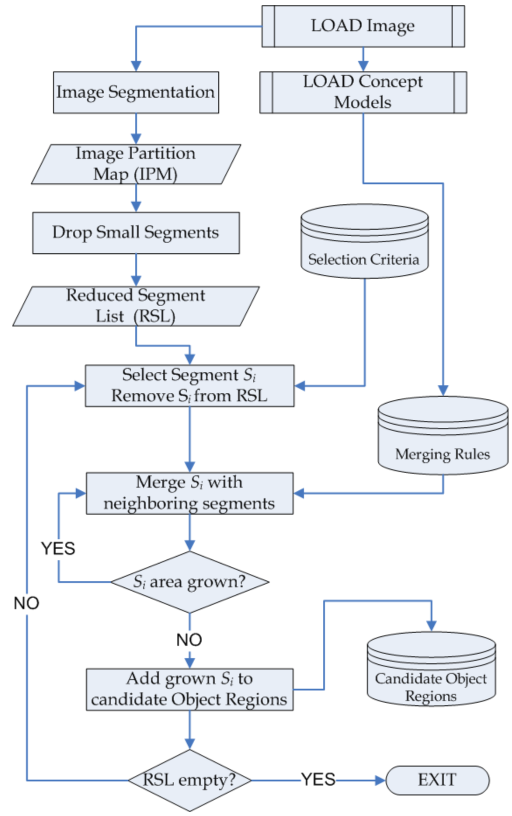

Figure 4: Flowchart of the image segmentation process

aid of the well-known JSEG [4] algorithm. By setting the merging threshold m of this algorithm to low values (e.g., $m = 0.05$) over-segmentation, which allows for more accurate object contour detection, is obtained. The next step involves a region growing procedure. Initial segments P_i are grown toward maximization of the confidence values corresponding to object class models. The growing process stops whenever merging with adjacent segments does not increase the confidence value of the best matching model. All grown regions (which may overlap) are saved as candidate objects. The final decision, however, depends on the results of the classification task performed in the parent node as well on the confidence of the winning class (see also Section 5). Figure 4 illustrates the flowchart of the proposed model driven segmentation approach.

7 Experimental results

The basic aim of our experimental study was to identify the overall performance of the proposed automatic image annotation scheme. In addition we examined

ISSN: 1790-2769

ISBN: 978-960-474-012-3

Table 2: Concept detection results

Concept	*Precision*	*Recall*
Indoor	0.852	0.873
Outdoor	0.786	0.815
Jumping	0.813	0.841
Running	0.825	0.838
Throwing	0.846	0.862
Marathon	0.541	0.557
Walking	0.469	0.448
High Jump	0.608	0.623
Long Jump	0.411	0.453
Pole Vault	0.651	0.672
Triple Jump	0.439	0.477
Sprint	0.705	0.726
Long distance	0.618	0.624
Discus Throw	0.554	0.568
Hammer Throw	0.568	0.581
Javelin Throw	0.497	0.522
Shot Throw	0.601	0.639
Total	0.644	0.656

the difference in identifying abstract term and object class labels. Finally, we compared object class label identification using holistic and region based features. We have used the manually annotated data mentioned in Section 5 as ground truth. The training date were taken from 628 images containing 2098 concept instances. The remaining 418 images and 1448 concept instances were used for testing. Table 2 presents precision and recall per concept (only abstract term concepts are shown). Table 3 compares detection of object class concepts using holistic and image segment features. Though in both cases the results are far from being satisfactory it is clear that image segment features increase significantly the precision/recall figures. By comparing Tables 2 and 3 it is obvious that abstract terms are more easily and accurately identified than object class labels. This is not a surprise since object detection in non-constraint environments is a very hard (even unsolvable) problem.

8 Conclusion

In this work, we have presented a unifying framework for automatic image annotation with semantic labels. The proposed method performs classification tasks along the paths defined by an XML-dictionary tree, and thus, minimizes the number of classes to which categorization is performed. The result of classification at a parent node defines the subsequent classification task, whereas confidence estimation of the detected concepts is used to decide on label acceptance or not. Our evaluation study included 1046 images and 3546 concept instances. The main conclusions drawn from this study are: (a) Classification to abstract terms achieves high rates of both precision and recall, (b) Object detectors (that is labeling corresponding to image segments) perform rather poor but they still be better than detectors which use holistic image features, and (c) involvement of the proposed image segmentation scheme enhances significantly the object detection task. Further work involves: (a) improvement of concept modeling through investigation of alternatively machine learning techniques and training data filtering, (b) examination of the appropriate subset of the MPEG-7 feature set to identify subspaces that may increase classification accuracy, (c) improvement of the proposed image segmentation method in order to achieve plausible objects with higher probability, (d) investigation of ways for including the dominant color features in the image and image segment feature vectors, and (e) experimentation with other domains and benchmark datasets such as the Corel database.

Acknowledgements: This work was undertaken in the framework of the PSIFIORIKSI (Audiovisual Content Digitisation and Multimedia Metadata Extraction, Authoring and Storing based on MPEG-7) project funded by the Cyprus Research Promotion Foundation (CRPF) under the contract PDE/0406/07.

References:

[1] W. H. Adams, C.-Y. Lin, G. Iyengar, B. L. Tseng and J. R. Smith, IBM Multimodal Annotation Tool, IBM Alphaworks, August 2002.

[2] BOEMIE - Bootstrapping Ontology Evolution with Multimedia Information Extraction http://www.boemie.org

[3] G. Carneiro, A. B. Chan, P. J. Moreno, and N. Vasconcelos, "Supervised Learning of Semantic Classes for Image Annotation and Retrieval," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 3, pp. 394-410, March 2007.

[4] Y. Deng and B. S. Manjunath, "Unsupervised segmentation of color-texture regions in images and video," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 8, pp. 800-810, August. 2001.

[5] R.-E. Fan, P.-H. Chen, and C.-J. Lin, "Working set selection using the second order information for

Table 3: Comparison of holistic and region based features for object class label identification

	Holistic features		*Region-based features*	
Concept	*Precision*	*Recall*	*Precision*	*Recall*
Athlete	0.352	0.381	0.650	0.491
Horizontal Bar	0.181	0.227	0.574	0.721
Pillar	0.206	0.215	0.538	0.742
Pole	0.173	0.196	0.566	0.657
Sandpit	0.428	0.574	0.549	0.603
Discus	0.083	0.272	0.337	0.615
Net	0.487	0.725	0.416	0.178
Lane	0.356	0.611	0.551	0.633
Hurdle	0.071	0.219	0.093	0.114
Hammer	0.090	0.263	0.225	0.516
Javelin	0.164	0.205	0.467	0.538
Globe	0.075	0.192	0.241	0.497
People_Countable	0.523	0.610	0.335	0.362
People_Uncountable	0.544	0.597	0.328	0.411
People_Face	0.254	0.462	0.486	0.705
Place of Interest	0.161	0.258	0.133	0.191
Total	0.271	0.351	0.413	0.502

training SVM", *Journal of Machine Learning Research*, no. 6, pp. 1889-1918, 2005

[6] J. Hafner, H. Sawhney, W. Equitz, M. Flickner, and W. Niblack, "Efficient Color Histogram Indexing for Quadratic Form Distance Functions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, no. 7, pp. 729-736, July 1995.

[7] J. Huang, S. Kumar, M. Mitra, W. Zhu, and R. Zabih, "Spatial Color Indexing and Applications," *International Journal of Computer Vision*, vol. 35, no. 3, pp. 245-268, Dec. 1999.

[8] ISO/IEC 15938-3:2001 Information Technology - Multimedia Content Description Interface - Part 3: Visual, Version 1.

[9] A. Kounoudes, N. Tsapatsoulis, Z. Theodosiou, and M. Milis, "A multi-level Video Annotation Tool based on XML-dictionaries," *10th WSEAS International Conference on Mathematical Methods, Computational Techniques and Intelligent Systems*, Corfu, Greece, October 2008

[10] Y. Lee, Y. Lin, and G. Wahba, "Multicategory Support Vector Machines: Theory and Application to the Classification of Microarray Data and Satellite Radiance Data," *Journal of the American Statistical Association*, vol. 99, no. 465, pp. 67-81, March 2004

[11] B. S. Manjunath, J. -R. Ohm, V. V. Vasudevan, and A. Yamada, "Color and Texture Descriptors," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 11, no. 6, pp. 394-410, June 2001.

[12] MPEG-7 Visual Experimentation Model (XM), Version 10.0, ISO/IEC/JTC1/SC29/WG11, Doc. N4063, Mar. 2001.

[13] W. Niblack, R. Barber, W. Equitz, M. Flickner, E. Glasman, D. Pektovic, P. Yanker, C. Faloutsos, and G. Taubin, "The QBIC Project: Querying Images by Content Using Color, Texture, and Shape," *Proceedings of SPIE Storage and Retrieval for Image and Video Databases*, pp. 173-181, 1993.

[14] R. Picard, "Light-Years from Lena: Video and Image Libraries of the Future," *Proceeding of International Conf. on Image Processing*, Oct. 1995.

[15] A. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-Based Image Retrieval: The End of the Early Years," *IEEE Trans. Pattern. Analysis and Machine Intelligence*, vol. 22, no. 12, pp. 1349-1380, Dec. 2000.

[16] J. Smith and S. Chang, "Visually Searching the Web for Content," *Multimedia*, vol. 4, no. 3, pp. 12-20, July-Sept. 1997.

[17] I. H. Witten and E. Frank, "Data Mining: Practical machine learning tools and techniques", 2nd Edition, Morgan Kaufmann, San Francisco, 2005.

[18] Y. Yajima and T. -F. Kuo, "Efficient Formulations for 1-SVM and their Application to Recommendation Tasks," *Journal of Computers*, vol. 1, no. 3, pp. 27-34, June 2006.

[19] S.J. Yang, S.K. Kim, K.S. Seo, Y.M. Ro, J.Y. Kim, Y.S. Seo, "Semantic categorization of digital home photo using photographic region templates," *Information Processing and Management*, vol. 43, no. 2, pp. 503-514, 2007.