Regimvid at ImageCLEF 2015 Scalable Concept Image Annotation Task: Ontology based Hierarchical Image Annotation

Mohamed Zarka, Anis Ben Ammar and Adel M. Alimi

mohamed.ezzarka@ieee.org ben.ammar.regim@gmail.com adel.alimi@ieee.org
REGIM: Reaserch Groups on Intelligent Machines, University of Sfax, ENIS, BP 1173, Sfax, 3038, Tunisia
http://www.regim.org/

Abstract. In this paper, we describe our participation in the *ImageCLEF 2015 Scalable Concept Image Annotation* task. In this participation, we display our approach for an automatic image annotation by the use of an ontology-based semantic hierarchy handled at both learning and annotation steps. While recent works focused on the use of semantic hierarchies to improve concept detector accuracy, we are investigating the use of such hierarchies to reduce detector complexity and then, to handle efficiently large-scale image datasets. Our framework is based on two steps: (1) constructing a fuzzy ontology through analyzing learning dataset, and (2) guiding the annotation process through a reasoning engine. The obtained results confirm that this approach is promising for scalable image annotation.

Keywords: Image Annotation, Classification, Concept Detection, fuzzy Ontology, fuzzy Reasoning

1 Introduction

For the *ImageCLEF2015 Scalable Concept Image Annotation* task, our aim is to construct an automated image annotation framework that focuses on the scalability aspect through reducing semantic concept detection cost and complexity.

Automatic photo annotation is considered as a classification problem that consists in assigning a set of semantic concepts to a semantic content of a given image [33, 39].

Image collections are increasing staggeringly. Thus, retrieving from large-scale image datasets is a challenging task [36, 35, 16, 34]. The access to such enormous contents has forced the image retrieval community to look for advanced approaches and techniques in order to make the availability of automated and efficient semantic annotation for such contents [4, 30, 29].

In the previous *ImageCLEF Scalable Concept Image Annotation* task, [29] focused on the use of a knowledge based approach. Thus, an ontology was generated and used both: (1) in training phase to select images that should be used for optimizing classifiers, and (2) in testing phase for deducing new annotations through concept inter-relationships. This approach was considered as the winner within *ImageCLEF2014 Scalable Concept Image Annotation* task.

Aspiring to reconcile and exploit the semantic assets provided by the use of knowledge based approaches, number of initiatives have investigated the knowledge engineering for image retrieval ([26, 2, 10] to cite a few). Yet, ontologies (as a knowledge database) are powerful tools to design concepts and their interrelationships. In general, ontology-based approaches consists in defining a knowledge conceptualization and a reasoning process in order to handle and enhance a semantic interpretation. An immediate effect of such efforts is the alleviation of the semantic barriers and many promising works raised [18].

Aiming to contribute towards this direction, in [38], we presented a fuzzy ontology based framework for enhancing a multimedia content indexing accuracy. Key dimensions of this inquiry constitute the three main issues addressed by the existing ontologies, namely a generic ontology structure aspect, an automated knowledge extraction process for populating an ontology content, and a machine-driven context detection for a multimedia content. What was accomplished in this study is a novel ontology management method which is intended to a machine-driven knowledge database construction. The experiment that we conducted on the *ImageCLEF2012* dataset displayed semantic improvements over a classical image annotation framework used in large-scale multimedia contents.

Our submitted runs within *ImageCLEF2015 Scalable Concept Image Annotation* rely on a visual analysis of the provided testing dataset. As visual features, we used a *k-means* [32] algorithm to classify training local feature extract by SURF algorithm [3]. For scalability, our runs aim to show that we can go further in such aspect by reducing computing cost. In fact, we propose an ontology based approach that alleviates the computing cost for labeling a given testing image by candidate semantic concepts. By reading papers that have been published within *ImageCLEF Labs* [25, 36, 35, 7, 34], it can be clearly seen that there are a serious focus made on scalability through reducing the candidate concept list to be analyzed within an image content. Mainly, these works rely on dividing candidate concepts into: (1) initial concepts that can be detected directly through analyzing an image content, and (2) extended concepts that can be detected through reasoning with the initial ones. In [38], we presented a fuzzy framework for enhancing a semantic interpretation through reasoning with a given initial concept set. In our submitted runs, we focused on detecting initial concepts. Thus, our contribution consists in developing a fuzzy ontology to guide the annotation process through reducing the number of concepts to be detected.

The present working note is organized as follows: Section 2 describes the proposed framework. Section 3 describes our submitted runs to *ImageCLEF2015 Scalable Concept Image Annotation* task as well as comparison results with other participants runs. Finally, a conclusion and some future directions of our work are described in Section 4.

2 The RegimVid Image Annotation System

2.1 Framework overview

The RegimVid [11, 12, 19] is a semantic video indexing, retrieval and visualization system developed within our team. In this paper, we propose a scalable image annotation framework based on hierarchical annotators. We investigate research works on semantic hierarchies for hierarchical image annotation. Our framework relies on constructing and managing a fuzzy ontology that handle a semantic hierarchy. Such a hierarchy is used then to train more accurate image annotators (see figure 1).

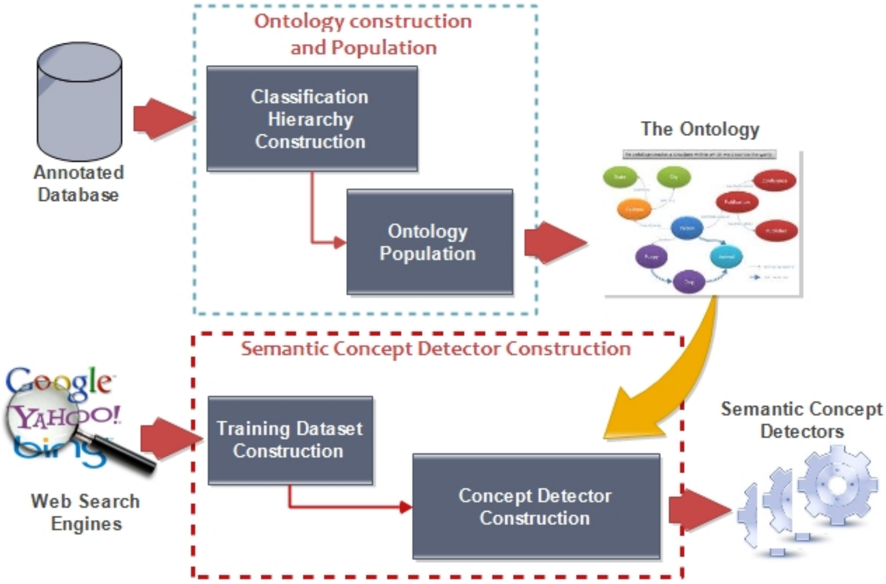

Fig. 1. Ontology based semantic annotator hierarchy for image annotation

Image annotation is considered as a multi-class classification problem. Many approaches were proposed to handle the annotation scalability aspect (large number of concept to annotate with) through combining semantic hierarchical structures with classification techniques (like SVM : Support Vector Machine) [8, 22, 1, 24].

Mainly, two different approaches were proposed for constructing the semantic hierarchy. The first one is qualified as top-down method: the semantic hierarchy is build through recursive class set clustering [8]. The second one is qualified as bottom-up method: the hierarchy is defined by agglomerative partitioning of the classes [22]. Furthermore, two different approaches were proposed also for hierarchical image classification: the first is the *Binary Hierarchical Decision Trees*

(BHDTS)[8], and the second is the *Decision Directed Acyclic Graphs* (DDAGS) [15].

Let $C = \{c_1, c_2, \ldots, c_N\}$ be a set of N semantic concept. The DDAGS approach trains $N * (N-1)/2$ binary classifiers and uses a DAG to decide if an image *image* belongs or not to a semantic concept class $c_i \in C$. At each given node at a distance d from the tree root, d semantic concept classes are eliminated, and $N - d$ decision nodes remain to be evaluated. The BHDTS approach handle the semantic hierarchy as a binary tree: concept classes are clustered hierarchically into two subsets. This clustering step is iterated until a single concept class set is reached. For every clustering step, an SVM classifier is trained in order to decide if an image *image* could be annotated by the first or the opposite semantic concept class. A total of $log_2(N)$ SVM classifiers are trained and used for analyzing a test image. Despite the fact that these two approaches enable accurate classifiers, they handle semantic hierarchy as binary structures which requests a considerable structure to handle with large amount of concept classes.

In our proposed framework, we aim to define a new method for constructing a hierarchical classifiers for scalable image annotation. At first, an annotated image dataset is analyzed to construct the hierarchy tree for concept classes. Then, and for every level of the defined tree structure, an SVM is trained for predicting if a test image *image* belongs to the first concept class set or the second one. By starting by the first level (root node), the hierarchy is walked until reaching leafs nodes through computing classifier votes (see figure 2).

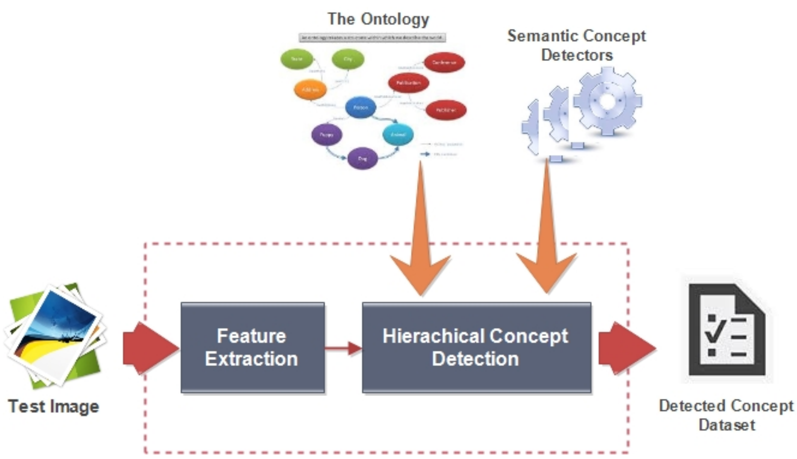

Fig. 2. Ontology based Hierarchical image classification

Our method is inspired from fuzzy decision tree based method [37, 6] to extract uncertain knowledge in a classification problem. Fuzzy set theory is used to model the tree structure. Thus, our proposed approach is based on a fuzzy ontology that handles such a decision tree. In what follows, we discuss the structure of our fuzzy ontology, we show how we populate its content, and how to infer available knowledge in order to use the hierarchical classifiers to annotate a test image accurately.

Ontology Structure The ontology structure is based on three conceptual classes : the semantic concept Concept, the hierarchical node Node, and the test image Image.

We define also a set of relationships between these conceptual classes (see table 1).

Table 1. Semantic Relationships between conceptual classes

Relationships	Definition	Meaning
isIndexedBy	$(\langle \mathsf{Image}, \mathsf{Concept} \rangle : \mathsf{isIndexedBy}) \geq p_1$	The image Image is annotated by the concept Concept by a fuzzy weight p_1
votesFor	$(\langle \mathsf{Node}, \mathsf{Image} \rangle : \mathsf{votesFor}) \geq p_2$	The SVM for the node Node votes for the image image by a fuzzy weight p_2
existsIn	$(\langle \mathsf{Concept}, \mathsf{Node} \rangle : \mathsf{existsIn}) \geq p_3$	The image image exists in the node Node by a fuzzy weight p_3
isChildOf	$(\langle \mathsf{Node}, \mathsf{Node} \rangle : \mathsf{isChildOf})$	The first node Node has a semantic concept subset of the second node Node

The relationship isChildOf depicts that a node $node_1 \in \mathsf{Node}$ is a child of another node $node_2 \in \mathsf{Node}$. This relationship is used then for modeling the semantic hierarchy for concept classes.

The relationship existsIn enumerates for each node $node \in \mathsf{Node}$ the contained set of concept classes. A concept $concept \in \mathsf{Concept}$ can exists in many nodes, but for separate levels.

The relationship votesFor is used when an image *image* is being annotated and the hierarchy is walked from the root node to the leafs. A node $node \in \mathsf{Node}$ votes for an image $image \in \mathsf{Image}$ by a fuzzy weight p_2 when a SVM classification on that image predicts that the image *image* could annotated by the set of semantic concepts that exists in the node *node*.

Finally, the relationship isIndexedBy depicts that an image $image \in \mathsf{Image}$ is annotated by the concept $concept \in \mathsf{Concept}$ by a fuzzy weight equal to p_1.

The proposed ontology structure is used to enable handling the hierarchical classifiers, to trace the hierarchy walk for classifying a given test image, and then to model the set of semantic concepts that annotate that image (see figure 3). In what follows, we expose the population process for our ontology, then, we discuss the reasoning process used to guide and assist the hierarchical annotation.

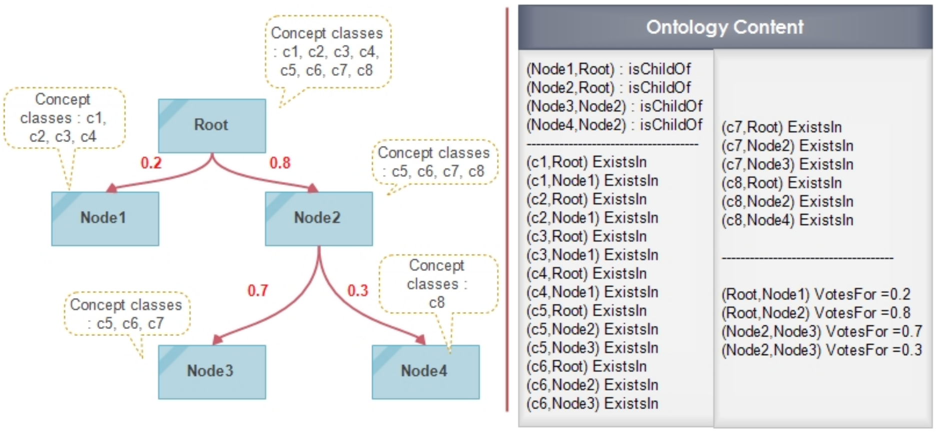

Fig. 3. Ontology based hierarchical image classification

Ontology population Given a defined set of semantic concepts, we start by clustering it through analyzing annotated image dataset provided by the *ImageCLEF 2015 Scalable Concept Image Annotation* task.

At first, we apply a binary clustering for the whole concept set, and we define two new nodes in the ontology $node_1$ and $node_2$. We use a *k-means* clustering algorithm with $k = 2$. Then, each concept is instantiated within the ontology, and for every concept *concept* that belongs to the node *node*, a new relationship existsIn is instantiated between *concept* and *node*. This process is recursively called on $node_1$ and $node_2$ until a sub-node contains only one semantic concept class, or the clustering process seems unable to cluster a given semantic concept classes. At each iteration, the new defined nodes are populated within the ontology through instantiating the isChildOf relationships.

Hierarchical classificators construction Once the hierarchical structure is defined through the above mentioned recursive binary clustering, an SVM based classifier is trained for all the nodes that belong to the same level. As training images, we select some development images for every concept that belongs to a node. In section 2.2, we detail the development image dataset used for the training task.

At a given level, two possible nodes are figuring (see figure 4). By exploring existsIn relationships, we construct a training image dataset. For the first node (see node *Root* in figure 4), and for each concept that belongs to that node, a subset of images that are annotated by this concept are selected to be training images for corresponding node.

For a leaf node (see node *Node1* in figure 4), we proceed as follow: let $C_m = \{c_1, c_2, \ldots, c_k\}$ be a set of k concepts that belongs to the node $node_m$. We construct then k classifiers. Each classifier is related to a given concept and trained against the other concepts. Then, and for a classifier f of a concept

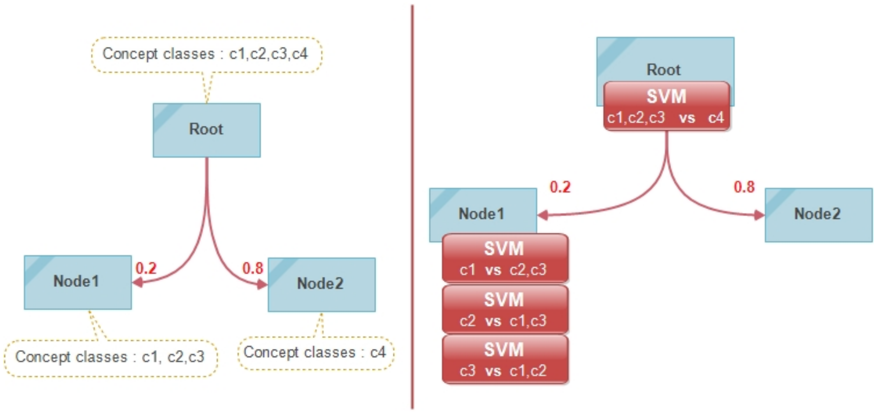

Fig. 4. Hierarchical SVM classificator construction

$c_f \in C_m$, we train an SVM classifier based on two image sets: the first set is based on images that are annotated by the concept c_f, and the other set is based on images that are annotated by the other concepts $(C_m \setminus c_f)$.

For a leaf node that contain only one concept class (see node *Node2* in figure 4), no SVM classifier will be constructed. And an image annotation for this concept will be computed through the leaf node classification vote.

Reasoning We start reasoning from the root node (top node) of the constructed fuzzy tree (see figure 3). For a given node, we compute the values of the membership functions (μ) for the child nodes through firing the corresponding SVM classifiers. The classification results (the vote) are populated into the fuzzy ontology through instantiating the votesFor relationship.

In order to improve reasoning accuracy and to minimize the decision tree walk (which will also minimize the number of SVM classifiers to be fired), we define a *Fuzziness control threshold* $\theta_r = 0.1$: given two sub-nodes $node_1$ and $node_2$, firing the SVM classifier at this level provides two membership function values μ_1 for $node_1$ and μ_2 for $node_2$. Then, we compute $\theta_r = |\mu_1 - \mu_2|$.

if $\theta_r \leq 0.1$, then we could not be sure if the SVM classifier is discriminative to judge if the content of a test image belongs to the first or to the second node. We proceed so to walk both sub-nodes ($node_1$ and $node_2$). For the opposite case ($\theta_r > 0.1$), the reasoner walks only the node that has the greater membership function value (μ).

Given the example in figure 3, the SVM classifier of the node *root* computed $\mu_1 = 0.2$ for the node $Node1$, and $\mu_2 = 0.8$ for the node $Node2$. Then, the reasoning algorithm stops walking the node $Node1$ and proceeds to walk the $Node2$ since $\theta_r = 0.8 - 0.2 = 0.6$ and $0.1 \leq 0.6$.

A leaf node can contain a set of concept classes, or only one concept class. In the first case, and for every contained concept class, an SVM classifier is fired for that concept against the other contained concept classes. The classification result is populated in the ontology through the instantiation of the relationship isIndexedBy between the concept class and the test image. The fuzzy weight for the new relationship is computed as an average of μ values computed from the root node to the leaf one. In case of a single concept class, a new isIndexedBy relationship is instantiated within the ontology between that concept and the test image. The fuzzy weight as in the first case.

Our proposed fuzzy decision tree reasoner assists the annotation of a given test image through firing recursive trained SVM classifiers in order to optimize the number of concept to be detected. Such an optimization should reduce also the computing cost of a given test image annotation process.

In the next section, we expose how we construct an SVM classifier for each node in the constructed fuzzy hierarchical semantic structure of concept classes.

2.2 Svm Classifier Construction

In our participation within *ImageCLEF 2015 Scalable Concept Image Annotation* task, we aimed basically to evaluate the scalability aspect of our preliminary automatic annotation framework. For semantic concept detector/annotator, we have not really defined an original approach, but we implemented state-of-the-art bags of quantized local features and linear classifiers learned by support vector machines. In fact, and as pointed in [28], bag-of-features and codebook approach has gained a great attention by image classification and annotation community as it showed notable semantic accuracy [26, 17, 9]. In what follow, we expose how we construct SVM classifiers for semantic concept detection and annotation.

Construct a learning dataset Image annotation has always been heavily dependent to good development datasets. First, datasets were mainly hand-collected. However, and recently, several researches attempt to automate such a laborious task. Re-ranking images gathered from popular Image search engines (GOOGLE, YAHOO!, BING, ...) can construct automatically an image learning dataset [14, 13, 31].

As a development dataset, we have not used one provided by the *ImageCLEF 2015 Scalable Concept Image Annotation task*. In fact, not all the concepts were annotated. We relied then on FLICKR image search engine to obtain image set and construct a learning dataset. We used so the information provided with concept list to query the search engine and we gathered first 100 result images for each given concept.

At the outset, it seems to be curious to use an external data source as a development dataset. Our aim is to explore available on-line data-sources (like search engines) to train non annotated semantic concepts.

Local Feature Extraction Our framework extracts features from an input image through a robust local feature extractor. We followed a basic and state-of-the-art framework for such purpose (as described in [28]). Leading extractors for such a purpose includes *Scale Invariant Feature Transform* (SIFT) and

Speeded Up Robust Features (SURF). Local feature descriptors handle a pixel within an image by analyzing its neighborhood pixels. Many different descriptors and interest-point detectors were proposed and discussed in the literature. While the SIFT descriptor [23] is considered as the most widely used descriptor, SURF [3] is known as robust local feature extraction to various image perturbations.

Our framework extracts local features and descriptors using SURF. Such a choice is argued by SURF concise descriptor length (64 floating point values). The SURF implementation that we used is provided by OPENCV [5].

For query image analysis, local features are extracted and mapped into nearest computed cluster centroids. The query image is then handled by a vector that represents defined visual bag-of-words.

Classification of local Features and Contructing the bag-of-words model After extracting local features, a bag-of-words model is used to represent these descriptors. The latter are extracted from training images and are grouped into N clusters of visual words using *k-means*. Each defined descriptor is classified into its cluster centroid by computing the *Euclidean distance* metric. For our runs, we choose a value of $N = 100$. This value is argued by a balance between high bias (under-fitting) and high variance (over-fitting).

In order to alleviate the computing cost of *k-means* clustering, we used *Mini Batch k-means* [32] as an alternative to the *k-means* algorithm for clustering massive datasets. *Mini Batch k-means* reduces the computational cost by handling fixed size subsample instead of all the data in the database. This strategy reduces the amount of distance to be computed at each clustering iteration.

Learning Algorithm The learning algorithm consists in training one-vs.-one linear SVM to operate in the bag of SURF feature space. Training images are classified through a histogram vector conctructed in the *k-means* based clustering. We used a linear kernel for our SVM based learning algorithm in view of its simplicity and computational efficiency in training and classification: $K(x, y) = x^T y + c$.

Basically, SVM are binary classifier. For a given detector, an image is annotated by one of two distinct groups. A one-vs.-one scheme is used in which each SVM trained for each combination of individual classes. The SVM implementation used in our runs is given by SCIKIT-LEARN library [27].

Decision As an SVM decision function, a class membership probability estimation fits the decision values.

SCIKIT-LEARN library uses a PLATT SCALING in order to calibrate the SVM classifier to produce, in addition to class predictions, probabilities. When the SVM is trained, an optimization process is called to optimize parameter vectors A and B such that : $P(y|X) = 1/(1 + exp(A * f(X) + B))$ where $f(X)$ is the signed distance of a sample from the hyperplane.

2.3 Object Localization

Our developed framework does not handle yet concept localization. As future work, we are motivated to use state-of-the-art based techniques (such as [21, 20]). In our submitted runs, we considered the whole image content as a localization for all annotated concepts.

3 Experiments and Results

3.1 Submitted Runs

We submitted two runs, where the only difference is the value of the threshold weight for all computed annotation:

- Run 1 (*regimvid_at_imageclef2015_task1*): In this run, we considered all the annotation weights performed by our annotation framework.
- Run 2 (*regimvid_at_imageclef2015_task1_0.7*): In this run, we considered only annotation weights that are greater or equal to 0.7.

3.2 Results

We would like to notice that we annotated only 300 000 images of the 500 000 images provided in the test dataset. And, since many test images were not accessible on-line, we extracted features from the provided image thumbnails[1] (with low resolutions). Due to these facts, our runs haven't reached an advanced position compared to the other runs (see tables 2 and 3). Furthermore, our system used a state-of-the-art based Svm classifier. We think that a complete image annotation on real (full size) images with more tweaked Svm classifiers should give better results. Furthermore, fuzzy ontology based semantic enhancement (described in [38]) should also enhance our framework annotation accuracy.

Table 2. MAP_0_Overlap Runs evaluation

	MAP
Best run	0, 795403 (/SMIVA/21.run)
Worst run	0, 0305398 (/REGIM/regimvid_at_imageclef2015_task1_0.7.txt)
Average	0, 31046
Our best run	0, 0366072 (position 85/89)

3.3 Runtime

Training process: Training the Svm classifiers task elapsed about 4 days (we used 100 learning images per a concept). This task was executed on a moderm machine (Intel $i5$ processor with 16 GB RAM memory).

[1] compressed in the `webupv15_data_visual_images.zip` file.

Table 3. MAP_0.5Overlap Runs evaluation

	MAP
Best run	$0,659507$ (/SMIVA/21.run)
Worst run	$0,000231898$ (/MLVISP6/run_blur1.txt)
Average	$0,18673$
Our best run	$0,0161687$ (position 75/89)

Annotation process: The annotation task was done on 10 Vps machines (each one has one core CPU and 1 GB of RAM). The annotation of 300 000 images elapsed about 1 633 hours (without taking into consideration the Vps parallel computing).

Our framework annotates a test image with an average of 19.615 seconds (the maximum record was 597.250 seconds and the minimum one was 0.066 second).

And for a given test image, an average of 52 Svm classifiers were fired (the maximum was 175 and the minimum was 6). Our framework has reduced the number of Svm classifiers to be fired in order to annotate a given test image.

4 Conclusion

In this working note, we described our annotation framework for the *ImageCLEF 2015 Scalable Concept Image Annotation* task. We discussed our ontology based framework for reducing the number of concepts to be detected for a given image. We developed a state-of-the art bag-of-words based concept detector (that uses Surf feature extractor and *k-means* classification). Then, concept detectors are selected through reasoning with a fuzzy ontology content. Thus, not all the concept detectors are used for a given image.

In our experiment, we showed how the use of such a method could reduce the number of concept detectors to be used in order to efficiently annotate a large-scale image dataset. While the obtained results were not really impressive, we still believe that our framework can reveal better results through tweaking local feature extraction and training, and exploiting semantic enhancement through fuzzy reasoning. Thus, we are considering potential future directions to further improve our proposed framework.

Acknowledgment

The authors would like to acknowledge the financial support of this work by grants from *General Direction of Scientific Research* (DGRST), Tunisia, under the **ARUB** program. The authors would like to acknowledge also the Image-CLEF2015 Organising Committee.

References

1. Bannour, H., Hudelot, C.: Hierarchical image annotation using semantic hierarchies. In: Proceedings of the 21st ACM International Conference on Information and Knowledge Management. pp. 2431–2434. CIKM '12, ACM, New York, NY, USA (2012)
2. Bannour, H., Hudelot, C.: Building and using fuzzy multimedia ontologies for semantic image annotation. Multimedia Tools and Applications pp. 1–35 (2013)
3. Bay, H., Ess, A., Tuytelaars, T., Gool, L.V.: Speeded-up robust features (surf). Computer Vision and Image Understanding 110(3), 346 – 359 (2008), similarity Matching in Computer Vision and Multimedia
4. Benavent, X., Castellanos, A., de Ves, E., Hernández-Aranda, D., Granados, R., García-Serrano, A.: A multimedia ir-based system for the photo annotation task at imageclef2013. In: Working Notes for CLEF 2013 Conference , Valencia, Spain, September 23-26, 2013. (2013)
5. Bradski, D.G.R., Kaehler, A.: Learning Opencv, 1st Edition. O'Reilly Media, Inc., first edn. (2008)
6. Bujnowski, P., Szmidt, E., Kacprzyk, J.: Intuitionistic fuzzy decision tree: A new classifier. In: Angelov, P., Atanassov, K., Doukovska, L., Hadjiski, M., Jotsov, V., Kacprzyk, J., Kasabov, N., Sotirov, S., Szmidt, E., Zadrony, S. (eds.) Intelligent Systems'2014, Advances in Intelligent Systems and Computing, vol. 322, pp. 779–790. Springer International Publishing (2015)
7. Cappellato, L., Ferro, N., Jones, G., San Juan, E. (eds.): CLEF 2015 Labs and Workshops, Notebook Papers. No. 994 in CEUR Workshop Proceedings (CEUR-WS.org) (2015), http://ceur-ws.org/Vol-1391
8. Cevikalp, H.: New clustering algorithms for the support vector machine based hierarchical classification. Pattern Recognition Letters 31(11), 1285 – 1291 (2010)
9. Elleuch, N., Ammar, A.B., Alimi, A.M.: A generic framework for semantic video indexing based on visual concepts/contexts detection. Multimedia Tools Appl. 74(4), 1397–1421 (2015)
10. Elleuch, N., Zarka, M., Ben Ammar, A., Alimi, M.A.: A fuzzy ontology: based framework for reasoning in visual video content analysis and indexing. In: Proceedings of the Eleventh International Workshop on Multimedia Data Mining. pp. 1–1. MDMKDD '11, ACM, New York, NY, USA (2011)
11. Elleuch, N., Zarka, M., Feki, I., Ammar, A.B., Alimi, A.M.: REGIMVID at TRECVID2010: semantic indexing. In: TRECVID 2010 workshop participants notebook papers, Gaithersburg, MD, USA, November 2010 (2010)
12. Feki, G., Ksibi, A., Ammar, A.B., Amar, C.B.: Regimvid at imageclef2012: Improving diversity in personal photo ranking using fuzzy logic. In: CLEF 2012 Evaluation Labs and Workshop, Online Working Notes, Rome, Italy, September 17-20, 2012 (2012)
13. Fergus, R., Fei-Fei, L., Perona, P., Zisserman, A.: Learning object categories from google's image search. In: Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on. vol. 2, pp. 1816–1823 Vol. 2 (Oct 2005)
14. Fergus, R., Perona, P., Zisserman, A.: A visual category filter for google images. In: Pajdla, T., Matas, J. (eds.) Computer Vision - ECCV 2004, Lecture Notes in Computer Science, vol. 3021, pp. 242–256. Springer Berlin Heidelberg (2004)
15. Gao, T., Koller, D.: Discriminative learning of relaxed hierarchy for large-scale visual recognition. In: Proceedings of the 2011 International Conference on Computer Vision. pp. 2072–2079. ICCV '11, IEEE Computer Society, Washington, DC, USA (2011)

16. Gilbert, A., Piras, L., Wang, J., Yan, F., Dellandrea, E., Gaizauskas, R., Villegas, M., Mikolajczyk, K.: Overview of the ImageCLEF 2015 Scalable Image Annotation, Localization and Sentence Generation task. In: CLEF2015 Working Notes. CEUR Workshop Proceedings, CEUR-WS.org, Toulouse, France (September 8-11 2015)
17. Kanehira, A., Hidaka, M., Makuta, Y., Tsuchiya, Y., Mano, T., Harada, T.: MIL at imageclef 2014: Scalable system for image annotation. In: Working Notes for CLEF 2014 Conference, Sheffield, UK, September 15-18, 2014. pp. 372–379 (2014)
18. Kannan, P., Bala, P., Aghila, G.: A comparative study of multimedia retrieval using ontology for semantic web. In: Advances in Engineering, Science and Management (ICAESM), 2012 International Conference on. pp. 400–405 (March 2012)
19. Ksibi, A., Ammar, B., Ammar, A.B., Amar, C.B., Alimi, A.M.: Regimrobvid: Objects and scenes detection for robot vision 2013. In: Working Notes for CLEF 2013 Conference , Valencia, Spain, September 23-26, 2013. (2013)
20. Lampert, C.H., Blaschko, M., Hofmann, T.: Beyond sliding windows: Object localization by efficient subwindow search. In: Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on. pp. 1–8 (June 2008)
21. Lazebnik, S., Schmid, C., Ponce, J.: Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In: Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on. vol. 2, pp. 2169–2178 (2006)
22. Li, L.J., Wang, C., Lim, Y., Blei, D., Fei-Fei, L.: Building and using a semantivisual image hierarchy. In: Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on. pp. 3336–3343 (June 2010)
23. Lowe, D.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision 60(2), 91–110 (2004)
24. McNamara, D.S., Crossley, S.A., Roscoe, R.D., Allen, L.K., Dai, J.: A hierarchical classification approach to automated essay scoring. Assessing Writing 23(0), 35 – 59 (2015)
25. Mller, H., Clough, P., Deselaers, T., Caputo, B.: ImageCLEF: Experimental Evaluation in Visual Information Retrieval. Springer Publishing Company, Incorporated, 1st edn. (2010)
26. Mylonas, P., Spyrou, E., Avrithis, Y., Kollias, S.: Using visual context and region semantics for high-level concept detection. Multimedia, IEEE Transactions on 11(2), 229–243 (2009)
27. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E.: Scikit-learn: Machine learning in Python. Journal of Machine Learning Research 12, 2825–2830 (2011)
28. Piras, L., Giacinto, G.: Open issues on codebook generation in image classification tasks. In: Machine Learning and Data Mining in Pattern Recognition - 10th International Conference, MLDM 2014, St. Petersburg, Russia, July 21-24, 2014. Proceedings. pp. 328–342 (2014)
29. Reshma, I.A., Ullah, M.Z., Aono, M.: KDEVIR at imageclef 2014 scalable concept image annotation task: Ontology based automatic image annotation. In: Working Notes for CLEF 2014 Conference, Sheffield, UK, September 15-18, 2014. pp. 386–397 (2014)
30. Sahbi, H.: CNRS - TELECOM paristech at imageclef 2013 scalable concept image annotation task: Winning annotations with context dependent svms. In: Working Notes for CLEF 2013 Conference , Valencia, Spain, September 23-26, 2013. (2013)

31. Schroff, F., Criminisi, A., Zisserman, A.: Harvesting image databases from the web. IEEE Transactions on Pattern Analysis and Machine Intelligence 33(4), 754–766 (2011)
32. Sculley, D.: Web-scale k-means clustering. In: Proceedings of the 19th International Conference on World Wide Web. pp. 1177–1178. WWW '10, ACM, New York, NY, USA (2010)
33. Snoek, C.G.M., Worring, M.: Concept-based video retrieval. Foundations and Trends in Information Retrieval 2(4), 215–322 (2009)
34. Villegas, M., Müller, H., Gilbert, A., Piras, L., Wang, J., Mikolajczyk, K., de Herrera, A.G.S., Bromuri, S., Amin, M.A., Mohammed, M.K., Acar, B., Uskudarli, S., Marvasti, N.B., Aldana, J.F., del Mar Roldán García, M.: General Overview of ImageCLEF at the CLEF 2015 Labs. Lecture Notes in Computer Science, Springer International Publishing (2015)
35. Villegas, M., Paredes, R.: Overview of the imageclef 2014 scalable concept image annotation task. In: CLEF 2014 Evaluation Labs and Workshop, Online Working Notes (2014)
36. Villegas, M., Paredes, R., Thomee, B.: Overview of the imageclef 2013 scalable concept image annotation subtask. In: CLEF 2013 Evaluation Labs and Workshop, Online Working Notes, Valencia, Spain (2013)
37. Wang, X., Liu, X., Pedrycz, W., Zhang, L.: Fuzzy rule based decision trees. Pattern Recognition 48(1), 50 – 59 (2015)
38. Zarka, M., Ben Ammar, A., Alimi, A.: Fuzzy reasoning framework to improve semantic video interpretation. Multimedia Tools and Applications pp. 1–32 (2015)
39. Zhang, D., Islam, M.M., Lu, G.: A review on automatic image annotation techniques. Pattern Recogn. 45(1), 346–362 (Jan 2012)