# Detection of Visual Concepts and Annotation of Images using Predictive Clustering Trees

Ivica Dimitrovski[1,2], Dragi Kocev[1], Suzana Loskovska[2], and Sašo Džeroski[1]

[1] Department of Knowledge Technologies, Jožef Stefan Institute
Jamova cesta 39, 1000 Ljubljana, Slovenia
[2] Department of Computer Science, Faculty of Electrical Engineering and Information Technology
Karpoš bb, 1000 Skopje, Macedonia
ivicad@feit.ukim.edu.mk, Dragi.Kocev@ijs.si, suze@feit.ukim.edu.mk, Saso.Dzeroski@ijs.si

**Abstract.** In this paper, we present a multiple targets classification system for visual concepts detection and image annotation. Multiple targets classification (MTC) is a variant of classification where an instance may belong to multiple classes at the same time. The system is composed of two parts: feature extraction and classification/annotation. The feature extraction part provides global and local descriptions of the images. These descriptions are then used to learn a classifier and to annotate an image with the corresponding concepts. To this end, we use predictive clustering trees (PCTs), which are capable to classify an instance to multiple classes at once, thus exploit the interactions that may occur among the different visual concepts (classes). Moreover, we constructed ensembles (random forests) of PCTs, to improve the predictive performance. We tested our system on the image database from the visual concept detection and annotation task part of ImageCLEF 2010. The extensive experiments conducted on the benchmark database show that our system has very high predictive performance and can be easily scaled to large number of images and visual concepts.

## 1 Introduction

An ever increasing amount of visual information is becoming available in digital form in various digital archives. The value of the information obtained from an image depends on how easily it can be found, retrieved, accessed, filtered and managed. Therefore, tools for efficient archiving, browsing, searching and annotation of images are a necessity.

A straightforward approach, used in some existing information retrieval tools for visual materials, is to manually annotate the images by keywords and then to apply text-based query for retrieval. However, manual image annotation is an expensive and time-consuming task, especially given the large and constantly growing size of image databases.

The image search provided by major search engines, such as Google, Bing, Yahoo! and AltaVista, relies on textual or metadata descriptions of images found

on the web pages containing the images and the file names of the images. The results from these search engines are very disappointing when the visual content of the images is not mentioned, or properly reflected, in the associated text.

A more sophisticated approach to image retrieval is automatic image annotation: a computer system assigns metadata in the form of captions or keywords to a digital image [6]. These annotations reflect the visual concepts that are present in the image. This approach begins with the extraction of feature vectors (descriptions) from the images. A machine learning algorithm is then used to learn a classifier, which will then classify/annotate new and unseen images.

Most of the systems for detection of visual concepts learn a separate model for each visual concept [8]. However, the number of visual concepts can be large and there can be mutual connections between the concepts that can be exploited. An image may have different meanings or contain different concepts, multiple targets classification (MTC) can be used for obtaining annotations (i.e., labels for the multiple visual concepts present in the image) [8]. The goal of MTC is to assign to each image multiple labels, which are a subset of a previously defined set of labels.

In this paper, we present a system for detection of visual concepts and annotation of images. For the annotation of the images, we propose to exploit the interactions between the target visual concepts (inter-class relationships among the image labels) by using predictive clustering trees (PCTs) for MTC. PCTs are able to handle multiple target concepts, i.e., perform MTC. To improve the predictive performance, we use ensembles (random forests) of PCTs for MTC. For the extraction of features, we use several techniques that are recommended as most suitable for the type of images at hand [8].

We tested the proposed approaches on the image database from the visual concept detection and annotation task part of ImageCLEF 2010 [10]. The visual concept detection and annotation task is a multiple labels (targets) classification challenge. It aims at the automatic annotation of a large number of consumer photos with multiple annotations. The concepts used in this annotation task are for example abstract categories like Family/Friends or Partylife, the time of day (day, night, sunny, ...), Persons (no, single, small or big group), Quality (blurred, underexposed, ...) and etc.

The remainder of this paper is organized as follows. Section 2 presents the proposed large scale visual concept detection system. Section 3 explains the experimental design. Section 4 reports the obtained results. Conclusions and a summary are given in Section 5.

## 2 System for Detection of Visual Concepts

### 2.1 Overall architecture

Fig. 1 presents the architecture of the proposed system for visual concepts detection and image annotation. The system is composed of a feature extraction part and a classification/annotation part. We use two different sets of features

to describe the images: global and local features extracted from the image pixel values. We employ different sampling strategies and different spatial pyramids to extract the visual features (both global and local) [5].

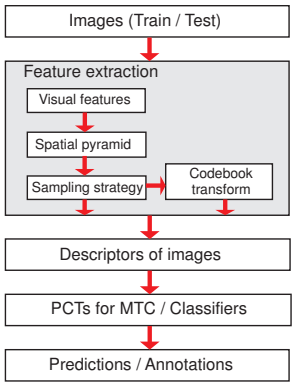


**Fig. 1.** Architecture of the proposed system for detection of visual concepts and annotation of images.

As an output of the feature extraction part, we obtain several sets of descriptors of the image content that can be used to learn a classifier to annotate the images with the visual concepts. Tommassi et al. [14] show that usage of various visual features that bring different information about the visual content of the images clearly outperform single feature approaches. Following these findings, in our research we use 'high level' feature fusion scheme.

The high level fusion scheme (depicted in Fig. 2) is performed as follows. First, we learn a classifier for each set of descriptors separately. The classifier outputs the probabilities with which an image is annotated with the given visual concepts. To obtain a final prediction, we combine the probabilities output from the classifiers for the different descriptors by averaging them. Depending on the domain, different weights can be used for the predictions of the different descriptors.

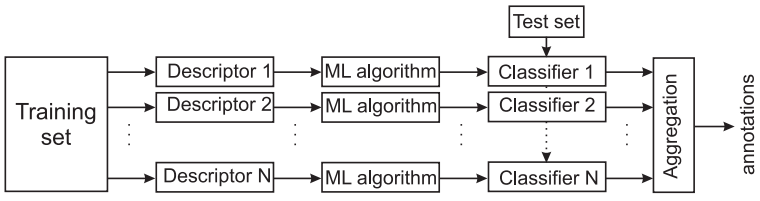


**Fig. 2.** High level fusion of the various descriptors.

### 2.2 Multiple Targets Classification

Following the reccomendations from [3], we formally describe the machine learning task that we consider here - multiple targets classification.

We define the task of multiple targets prediction as follows:
**Given:**

- A *description space* $X$ that consists of tuples of primitives (boolean, discrete or continuous variables), i.e. $\forall X_i \in X, X_i = (x_{i_1}, x_{i_2}, ..., x_{i_D})$, where $D$ is the size of a tuple (or number of descriptive variables),
- a *target space* $Y$, where each tuple consists of several variables that can be either continuous or discerete, i.e., $\forall Y_i \in Y, Y_i = (y_{i_1}, y_{i_2}, ..., y_{i_T})$, where $T$ is the size of a tuple (or number of target variables),
- a *set of examples/instances* $E$, where $E = \{(X_i, Y_i) | X_i \in X, Y_i \in Y, 1 \leq i \leq N\}$ and $N$ is the number of examples of $E$ ($N = |E|$), and
- a *quality criterion* $q$ (which rewards models with high predictive accuracy and low complexity).

**Find:** a function $f : X \rightarrow Y$ such that $f$ maximizes $q$. Here, the function $f$ is presented with decision trees, i.e., predictive clustering trees.

If the tuples from $Y$ (the target space) consist of continuous/numeric variables then the task at hand is multiple targets regression. Likewise, if the tuples from $Y$ consist of discrete/nominal variables then the task is called multiple targets classification.

### 2.3 Ensembles of PCTs for MTC

In the PCT framework [1], a tree is viewed as a hierarchy of clusters: the top-node corresponds to one cluster containing all data, which is recursively partitioned into smaller clusters while moving down the tree.

PCTs are constructed with a standard "top-down induction of decision trees" (TDIDT) algorithm. The heuristic for selecting the tests is the reduction in variance caused by partitioning the instances, where the variance $Var(S)$ is defined by equation (1) below. Maximizing the variance reduction maximizes cluster homogeneity and improves predictive performance.

A leaf of a PCT is labeled with/predicts the prototype of the set of examples belonging to it. With appropriate variance and prototype functions, PCTs can handle different types of data, e.g., multiple targets [4], hierarchical multi-label classification [15] or time series [12]. A detailed description of the PCT framework can be found in [1]. The PCT framework is implemented in the CLUS system, which is available for download at `http://www.cs.kuleuven.be/~dtai/clus`.

The prototype function returns a vector containing the probabilities that an example belongs to a given class for each target attribute. This afterwards can be used to calculate the majority class for each target attribute. The variance function is computed as the sum of the entropies of class variables:

$$Var(E) = \sum_{i=1}^{T} GiniCoefficient(E, y_i) \quad (1)$$

For a detailed description of PCTs for MTC the reader is referred to [1, 4]. Next, we explain how PCTs are used in the context of an ensemble classifier, namely ensembles further improve the performance of PCTs.

**Random Forests of PCTs** To improve the predictive performance of PCTs, we use ensemble methods. An ensemble classifier is a set of classifiers. Each new example is classified by combining the predictions of each classifier from the ensemble. These predictions can be combined by taking the average (for regression tasks) or the majority vote (for classification tasks) [2]. In our case, the predictions in a leaf are the proportions of examples of different classes that belong to it. We use averaging to combine the predictions of the different trees. As for the base classifiers, a threshold should be specified to make a prediction.

We use random forests as an ensemble learning technique. A random forest [2] is an ensemble of trees, obtained both by bootstrap sampling, and by randomly changing the feature set during learning. More precisely, at each node in the decision tree, a random subset of the input attributes is taken, and the best feature is selected from this subset (instead of the set of all attributes). The number of attributes that are retained is given by a function f of the total number of input attributes $x$ (e.g., $f(x) = x, f(x) = \sqrt{x}, f(x) = \lfloor \log_2 x \rfloor + 1$).

### 2.4 Feature Extraction

We use different commonly used types of techniques for feature extraction from images. We employ two types of global image descriptors: gist features [11] and a RGB color histogram, with 8 bins in each color channel for the RGB color space.

Local features include scale-invariant feature transforms (SIFT) extracted densely on a multi-scale grid [7]. The dense sampling gives an equal weight to all key-points, independent of their spatial location in the image. To overcome this limitation, one can use spatial pyramids of 1x1, 2x2 and 1x3 regions [13].

We computed four different sets of SIFT descriptors over the following color spaces: RGB, opponent, normalized opponent and gray. For each set of SIFT descriptors, we use the codebook approach to avoid using all visual features of an image [13].

The generation of the codebook begins by randomly sampling 50 key-points from each image and extracting SIFT descriptors in each key-point (i.e., each key-point is described by a vector of numerical values). Then, to create the codewords, we employ k-means clustering on the set of all key-points. We set the number of clusters to 4000, thus we define a codebook with 4000 codewords (a codeword corresponds to a single cluster and a codebook to the set of all clusters). Afterwards, we assign the key-points to the discrete codewords predefined in the codebook and obtain a histogram of the occurring visual features. This histogram will contain 4000 bins, one for each codeword. To be independent of the total number of key-points in an image, the histogram bins are normalized to sum up to 1.

The number of key-points and codewords (clusters) are user defined parameters for the system. The values used above (50 key-points and 4000 codewords) are recommended for general images [13].

## 3 Experimental Design

### 3.1 Definition and Parameter Settings

We evaluated our system on the image database from the visual concept detection and annotation task part of ImageCLEF 2010. The image database consists of training (8000) and test (10000) images. The images are labeled with 93 visual concepts [10]. A list of the visual concepts is presented in Table 2. The goal of the task is to predict which of the visual concepts are present in each of the testing images.

We generated six sets of visual descriptors for the images: four sets of SIFT descriptors (one detector, dense sampling, over four different color spaces) with 32000 bins for each set (8 sub-images, from the spatial pyramids: 1x1, 2x2 and 1x3, 4000 bins each). We also generated two sets of global descriptors (gist features with 960 bins and RGB color histogram with 512 bins).

The parameter values for the random forests were as follows: we used 100 base classifiers and the size of the feature subset was set to 10% of the number of descriptive attributes.

### 3.2 Performance measures

The evaluation of the results is done using three measures of performance suggested by the organizers of the challenge: mean average precision (MAP), F-measure and average ontology score (AOS) [10]. The first score evaluates the performance for each visual concept (concept-based evaluation), while the second and the third evaluate the performance for each testing image (example-based evaluation).

The *mean average precision* is widely used evaluation measure. For a given target concept, the average precision can be calculated as the area under the precision-recall curve for that target. Hence, it combines both precision and recall into a single performance value. The average precision is calculated for each visual concept separately and the obtained values are then averaged to obtain the mean average precision.

The *F-measure* is also widely used measure and it is well known. F-measure is calculated as the weighted harmonic mean of precision and recall:

$$F - measure = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \tag{2}$$

The *AOS measure* calculates the misclassification cost for each missing or wrongly annotated concept per image. The AOS score is based on structure information (distance between concepts in the provided ontology of concepts),

relationships from the ontology and the agreement between annotators for a concept extend with misclassification cost that incorporates the Flickr context similarity costs map [9].

Additionally, we report Equal Error Rate (EER) and Area Under the ROC curve (AUC). However, we do not discuss these evaluation measures because they were not used for the evaluation of the submitted runs by the organizers of the competition [10]. This was the case because precision/recall analysis gives more intuitive and sensitive evaluation than the ROC analysis.

### 3.3 Submitted runs

We have submitted four different runs (see Table 1). We do not use the EXIF metadata and the Flickr user tags provided for the photos. This means that all our runs consider automatic annotation using only visual information.

The runs can be divided using the following criteria: used descriptors and rescaling of the outputs. We used two different sets of descriptors: only SIFT (local descriptors) and SIFT combined with global descriptors (RGBHist and Gist). Since the AOS measure uses threshold 0.5 to determine wheter an image is annotated with a concept, we lineraly scale the probabilities to cope with the skewed distribution of the visual concepts. The linear scaling can be done on low level or high level. With the low level approach, we linearly scale the outputs from each classifier (obtained from the separate descriptors) and then average these values. For the high level approach, we linearly scale the averaged output of the classifiers.

The runs can be summarized as follows:

- SIFT + RGBHist + Gist (HLScale): local descriptors (SIFT) and global descriptors (RGBHist and Gist) with high level linear scaling.
- SIFT (HLScale): local descriptors (SIFT) with high level linear scaling.
- SIFT + RGBHist + Gist (LLScale): local descriptors (SIFT) and global descriptors (RGBHist and Gist) with low level linear scaling.
- SIFT (LLScale): local descriptors (SIFT) with low level linear scaling.

## 4 Results and Discussion

Table 1 presents the results on the testing set of images from the 4 runs. The results show that by using both local and global descriptors we get better predictive performance. The incorporation of global descriptors in the learning phase lifts the predictive performance of AP by 6% on average for the following visual concepts: Flowers, Baby, Fancy, Summer, Sunset-Sunrise, Park-Garden, Plants, Still-Life and Food. We can note decrease in the predictive performance (by 7% for AP) only for the visual concept Travel. Considering the linear scaling, we obtain better results by applying high level scaling.

In the following we focus on the prediction scores for the individual visual concepts as evaluated by average precision score. We can note the higher AP

values for the following visual concepts: Neutral-Illumination, No-Visual-Season, No-Blur, No-Persons, Outdoor, Sky, Day, Landscape-Nature, No-Visual-Time, Clouds, Natural, Plants. We obtain lower AP values for the concepts that are less represented in the training set of images (e.g., rain, horse, skateboard, graffiti...) and the 'difficult' concepts (e.g., abstract, technical, boring). The agreement of human annotators on the 'difficult' concepts is $\sim 75\%$ [9].

**Table 1.** Results of the experiments evaluated using Mean Average Precision, Average F-measure and Ontology Score incorporating the Flickr Context Similarity costs.

| Run name | MAP | F-measure | OS with FCS |
|---|---|---|---|
| SIFT + RGBHist + Gist (HLScale) | 0.334 | 0.596 | 0.595 |
| SIFT (HLScale) | 0.318 | 0.574 | 0.570 |
| SIFT + RGBHist + Gist (LLScale) | 0.334 | 0.556 | 0.541 |
| SIFT (LLScale) | 0.317 | 0.545 | 0.543 |

Further improvements can be expected if different weighting schemes are used (to combine the predictions of the various descriptors). The weight of the descriptors can be adapted for each visual concept separately. For instance, the SIFT descriptors are invariant to color changes, and they do not predict well concepts where illumination is important. Thus, the weight of the SIFT descriptors in the combined predictions for those concepts should be decreased. Also we should find better descriptors for these concepts, such as estimating the color temperature and overall light intensity.

Another approach is to tackle the problem with the skewed distribution of concepts over the images. On approach can be generation of virtual images containing the under-represented visual concepts. These virtual images can be obtained with re-scaling, translation, rotation, changing the brightness of the images from the under-represented concepts.

## 5 Conclusion

Multiple targets classification (MTC) problems are encountered increasingly often in image annotation. However, flat classification machine learning approaches are predominantly applied in this area. In this paper, we propose to exploit the dependencies between the different target attributes by using ensembles of trees for MTC. Our approach to MTC builds a single classifier that simultaneously predicts all of the visual concepts present in the images at once. This means adding new visual concepts will just slighly decrease the computational efficiency. While, for the other approaches that create a classifier for each visual concept separately this means learning an additional classifier.

Applied on the image database from the visual concept detection and annotation task part of ImageCLEF 2010 our approach was ranked fourth for the example-based performance measures (Ontology Score with FCS and Average

**Table 2.** Results per concept for our best run in the Large-Scale Visual Concept Detection Task using the Average Precision, Equal Error Rate and Area Under the Curve. The concepts are presented in descending order by Average Precision.

| Concept | AP | EER | AUC |
|---|---|---|---|
| Neutral-Illumination | 0.979 | 0.275 | 0.791 |
| No-Visual-Season | 0.954 | 0.297 | 0.768 |
| No-Blur | 0.873 | 0.289 | 0.792 |
| No-Persons | 0.873 | 0.281 | 0.799 |
| Outdoor | 0.860 | 0.232 | 0.855 |
| Sky | 0.826 | 0.163 | 0.915 |
| Day | 0.826 | 0.268 | 0.816 |
| Landscape-Nature | 0.753 | 0.156 | 0.924 |
| No-Visual-Time | 0.742 | 0.266 | 0.819 |
| Clouds | 0.728 | 0.136 | 0.933 |
| Natural | 0.713 | 0.413 | 0.629 |
| Plants | 0.704 | 0.243 | 0.842 |
| Male | 0.674 | 0.269 | 0.809 |
| Sunset-Sunrise | 0.659 | 0.102 | 0.965 |
| Partly-Blurred | 0.644 | 0.269 | 0.813 |
| Cute | 0.620 | 0.420 | 0.610 |
| Portrait | 0.572 | 0.225 | 0.853 |
| No-Visual-Place | 0.565 | 0.294 | 0.785 |
| Water | 0.550 | 0.202 | 0.877 |
| Female | 0.524 | 0.272 | 0.803 |
| Trees | 0.521 | 0.231 | 0.851 |
| Adult | 0.493 | 0.292 | 0.773 |
| Single-Person | 0.493 | 0.305 | 0.762 |
| Citylife | 0.487 | 0.266 | 0.815 |
| Indoor | 0.484 | 0.310 | 0.773 |
| Night | 0.467 | 0.180 | 0.902 |
| Building-Sights | 0.467 | 0.237 | 0.849 |
| Family-Friends | 0.465 | 0.273 | 0.798 |
| Macro | 0.460 | 0.257 | 0.817 |
| Sea | 0.445 | 0.120 | 0.943 |
| Food | 0.436 | 0.207 | 0.891 |
| Park-Garden | 0.419 | 0.178 | 0.896 |
| Sunny | 0.408 | 0.329 | 0.740 |
| Mountains | 0.403 | 0.159 | 0.919 |
| Visual-Arts | 0.385 | 0.478 | 0.528 |
| Flowers | 0.357 | 0.223 | 0.846 |
| Still-Life | 0.356 | 0.267 | 0.814 |
| Vehicle | 0.343 | 0.298 | 0.785 |
| Animals | 0.334 | 0.285 | 0.781 |
| Aesthetic-Impression | 0.328 | 0.389 | 0.656 |
| Street | 0.305 | 0.245 | 0.829 |
| Beach-Holidays | 0.298 | 0.169 | 0.899 |
| Architecture | 0.297 | 0.315 | 0.740 |
| Summer | 0.297 | 0.287 | 0.782 |
| Autumn | 0.286 | 0.236 | 0.848 |
| Car | 0.281 | 0.253 | 0.826 |
| Small-Group | 0.272 | 0.304 | 0.765 |

| Concept | AP | EER | AUC |
|---|---|---|---|
| Partylife | 0.265 | 0.249 | 0.831 |
| Big-Group | 0.257 | 0.244 | 0.829 |
| Teenager | 0.256 | 0.279 | 0.789 |
| Fancy | 0.242 | 0.432 | 0.594 |
| Overall-Quality | 0.240 | 0.394 | 0.648 |
| Bodypart | 0.236 | 0.308 | 0.754 |
| Underexposed | 0.228 | 0.207 | 0.852 |
| Lake | 0.212 | 0.172 | 0.907 |
| Overexposed | 0.206 | 0.275 | 0.801 |
| Dog | 0.192 | 0.286 | 0.799 |
| Insect | 0.191 | 0.190 | 0.892 |
| Motion-Blur | 0.177 | 0.353 | 0.712 |
| Out-of-focus | 0.168 | 0.316 | 0.756 |
| Child | 0.162 | 0.335 | 0.734 |
| Toy | 0.161 | 0.276 | 0.804 |
| Bird | 0.161 | 0.261 | 0.816 |
| Baby | 0.152 | 0.277 | 0.817 |
| Painting | 0.148 | 0.317 | 0.761 |
| River | 0.144 | 0.179 | 0.899 |
| Artificial | 0.138 | 0.410 | 0.618 |
| Shadow | 0.135 | 0.323 | 0.725 |
| Snow | 0.117 | 0.254 | 0.814 |
| Winter | 0.113 | 0.292 | 0.787 |
| Desert | 0.111 | 0.152 | 0.930 |
| Fish | 0.105 | 0.361 | 0.692 |
| Boring | 0.105 | 0.439 | 0.592 |
| Travel | 0.105 | 0.319 | 0.738 |
| Sports | 0.099 | 0.337 | 0.715 |
| Birthday | 0.098 | 0.320 | 0.718 |
| Train | 0.098 | 0.236 | 0.819 |
| Ship | 0.085 | 0.237 | 0.838 |
| Bicycle | 0.077 | 0.302 | 0.777 |
| Bridge | 0.077 | 0.210 | 0.864 |
| Spring | 0.076 | 0.210 | 0.884 |
| Musical Instrument | 0.075 | 0.292 | 0.787 |
| Technical | 0.070 | 0.390 | 0.653 |
| Church | 0.067 | 0.290 | 0.802 |
| Airplane | 0.060 | 0.255 | 0.823 |
| Old-person | 0.055 | 0.310 | 0.745 |
| Work | 0.053 | 0.390 | 0.666 |
| Cat | 0.032 | 0.377 | 0.686 |
| Graffiti | 0.026 | 0.394 | 0.669 |
| Abstract | 0.022 | 0.399 | 0.646 |
| Horse | 0.013 | 0.324 | 0.760 |
| Rain | 0.007 | 0.353 | 0.708 |
| Skateboard | 0.001 | 0.452 | 0.622 |
| Average | 0.334 | 0.281 | 0.788 |

F-measure) and fifth for the concept-based evaluation (Mean Average Precision), out of 17 competing groups.

The system we presented is general. It can be easily extended with new feature extraction methods, and it can thus be easily applied to other domains, types of images and other classification schemes. In addition, it can handle arbitrarily sized hierarchies organized as trees or directed acyclic graphs.

## References

1. Blockeel, H., De Raedt, L., Ramon, J.: Top-down induction of clustering trees. In Proc. of the 15th ICML, 55–63 (1998)
2. Breiman, L.: Random Forests. Machine Learning, 45, 5–32 (2001)
3. Džeroski, S.: Towards a General Framework for Data Mining. In Proc. of the 5th KDID, LNCS vol. 4747, 259–300 (2007)
4. Kocev, D., Vens, C., Struyf, J., Džeroski, S.: Ensembles of Multi-Objective Decision Trees. Proc. ECML 2007, LNAI vol.4701, 624–631 (2007)
5. Lazebnik, S., Schmid, C., Ponce, J.: Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. CVPR, 2169–2178 (2006)
6. Li, J., Wang, J. Z.: Real-Time Computerized Annotation of Pictures. IEEE Trans. on Pattern Analysis and Machine Intelligence, 30(6), 985–1002 (2008)
7. Lowe, D. G.: Distinctive Image Features from Scale-Invariant Keypoints. International Journal of Computer Vision, 60(2), 91–110 (2004)
8. Nowak, S., Dunker, P.: Overview of the CLEF 2009 Large-Scale Visual Concept Detection and Annotation Task. Multilingual Information Access Evaluation Vol. II Multimedia Experiments: 10th Workshop of the CLEF 2009, to appear in LNCS, Corfu, Greece (2010)
9. Nowak, S., Lukashevich, H.: Multilabel classification evaluation using ontology information, Workshop on IRMLeS, Heraklion, Greece (2009)
10. Visual Concept Detection and Annotation Task at ImageCLEF 2010: http://www.imageclef.org/2010/PhotoAnnotation
11. Oliva, A., Torralba, A.: Modeling the Shape of the Scene: A Holistic Representation of the Spatial Envelope. International Journal of Computer Vision, 42(3), 145–175 (2001)
12. Slavkov, I., Gjorgjioski, V., Struyf, J., Džeroski, S.: Finding explained groups of time-course gene expression profiles with predictive clustering trees. Molecular BioSystems, vol.6, no.4, 729–740 (2010)
13. Van de Sande, K., Gevers, T., Snoek., C.: A comparison of color features for visual concept classification. CIVR, 141–150 (2008)
14. Tommasi, T., Orabona, F., and Caputo, B.: Discriminative cue integration for medical image annotation. Pattern Recognition Letters, 29(15), 1996–2002 (2008)
15. Vens, C., Struyf, J., Schietgat, L., Dzeroski, S., Blockeel, H.: Decision trees for hierarchical multi-label classification. Machine Learning 73(2), 185–214 (2008)