Multi-level Image Annotation Using Bayes Classifier and Fuzzy Knowledge Representation Scheme

MARINA IVASIC-KOS[1], IVO IPSIC[1], SLOBODAN RIBARIC[2]
[1]Department of Informatics
University of Rijeka
R. Matejčić 2, 51000 Rijeka
CROATIA
{marina, ivoi}@uniri.hr, http://www.inf.uniri.hr/

[2]Department of Electronics, Microelectronics, Computer and Intelligent Systems
Faculty of Electrical Engineering and Computing, University of Zagreb
Unska 3, 10000 Zagreb
CROATIA
slobodan.ribaric@zemris.fer.hr, http://www.fer.unizg.hr/

Abstract: - Automatic image annotation (AIA) is the process by which metadata, in form of keywords or text descriptions are automatically assigned to an unlabeled image. Generally, two problems can be distinguished: the problem of semantic extraction, due to the gap between the image features and object labels, and the problem of semantic interpretation, due to the gap between the object labels and the human interpretation of images.

In this paper, a model for multi-level image annotation that is performed in two phases is proposed. In the first phase, a Naïve Bayes classifier is used to classify low-level image features into elementary classes. In the second phase, a knowledge representation scheme based on Fuzzy Petri Net is used to expand the level of vocabulary and to include multi-level semantic concepts related to images into image annotations. In the paper, a knowledge representation scheme for outdoor image annotation is given. Procedures for determining concepts related to an image using fuzzy recognition and inheritance algorithms on a knowledge representation scheme are presented, as well as experimental results of image annotation.

Key-Words: - Multi-level Image Annotation, Fuzzy Petri Net, Knowledge Representation, Naïve Bayes

1. Introduction

Digital images are becoming more present in everyday life, both in private and in business applications, so the number of digital images is rapidly increasing. Due to the large number of images, efficient search and retrieval are becoming demanding, as well as organization and storage of the images.

The problem of image search and retrieval has been intensively investigated using a variety of methods that can be roughly classified into two paradigms: text based image retrieval (keyword-based queries) and content based image retrieval (image-based queries) [1]. It turned out that users prefer keyword-based queries for image retrieval mainly because of the simplicity of placing the query, successful use of existing technologies and intuitive results, close to the user's expectations.

To make it possible to retrieve images by keywords, images must be annotated and complemented by metadata. Metadata may contain various types of information about the image such as date, location, resolution, size, keywords, free-text description, etc.

Manually providing image annotation is a tedious and expensive task, especially when dealing with a large number of images, thus automatic image annotation has emerged as an alternative solution.

Automatic image annotation (AIA) is the process by which metadata in form of keywords or text descriptions are automatically assigned to an unlabeled image. If an image is represented with low-level features such as color, shape or texture extracted from image content, AIA techniques try to find reliable mapping between these features and the concepts (keywords) that people would use to interpret that image.

As a representation that one can get from the raw image data cannot be simply transformed into an interpretation of the image inherent to humans, the so called semantic gap occurs [2].

Bridging the semantic gap is the main challenge in AIA. Generally two problems can be distinguished: the problem of semantic extraction due to the gap between the imaging features and object labels, and the problem of semantic interpretation due to the gap between object labels and human image interpretation.

AIA approaches that tackle the first problem usually use classification or probabilistic methods, whereas for introduction of multi-level semantics in the second problem, semantic modeling and knowledge representation schemes specific to the application domain are needed.

In this paper, automatic multi-level image annotation is performed in two phases. In the first phase a Bayesian classifier is used for mapping low-level features to object labels - classes that correspond to keywords from a controlled vocabulary. Obtained classes are inputs to the second phase in which a knowledge representation scheme based on Fuzzy Petri Net (KRFPN) [3] is used. The KRFPN scheme is used to enable inference of semantic concepts at different semantic levels related to an image and to expand the level of vocabulary by linking concepts with inheritance and compositional relationships and synonyms.

The reminder of the paper is organized as follows. In Section 2 a brief review of recent related work is given. Then, in Section 3 the proposed model is presented in detail. An example of knowledge representation scheme adopted for outdoor image annotation and determination of concepts related to images using fuzzy inference algorithms is given in Section 4 and 5, respectively. The experimental set-up, results and conclusion are given in Section 6 and 7, respectively.

2. Related work

AIA has been an active research topic in recent years due to its potential impact on both image interpretation and image retrieval or search. AIA approaches that have been proposed so far can be divided into two types according to the semantic level of concepts that are used for annotation and interpretation of images [4].

Classical AIA approaches look for a mapping between image features and concepts in a flat controlled vocabulary. For that purpose, classification, probabilistic modeling and lately graph based methods have been extensively used. Methods based on classification like [5] treat each semantic keyword or concept as an independent class and assign each keyword to one classifier. The probabilistic methods aim to learn a relevance model to represent correlation between images and keywords. Methods based on translation model [6] and models which use latent semantic analysis [7] fall into this category. A recent survey of research made in the field can be found in [1, 8].

In general, the amount of knowledge that is needed for the classification of images increases with the semantic level of concepts used to interpret the images. Thus, AIA approaches that tackle the problem of multi-level image annotation and semantic interpretation usually use logical reasoning and a knowledge base to introduce rich semantics. In recent years, several different approaches have been proposed and hereafter some of them will be mentioned.

A hierarchical model for generating words that correspond to class labels has been proposed in [9]. The model is inspired by the Hofmann's hierarchical clustering model and a model of soft clustering.

In [10] a SVM classifier is used for learning the elementary classes of natural scenes which are then linked using a probabilistic modeling method into concepts of a higher semantic level to achieve multi-level image annotation.

In [11, 12] ontology was used for the semantic description of the image content and descriptive logic was used in [12] for verification of the classification results.

To explore the ontology of words that is used for image interpretation and multi-level annotation in [13] a WordNet has been proposed. This idea is further extended in [14]. The authors intend to illustrate each of the concepts from the WordNet ontology with 500-1000 images in order to create public image ontology, the ImageNet.

Within the project aceMedia, in [15] ontology with fuzzy logic is combined to generate concepts from beach domain with appropriate reliability. Later on, in [16], the same group of authors have used a combination of different classifiers for learning concepts and fuzzy spatial relationships. Authors have reported that the environment used by the ontology is shown to be incompatible with that of fuzzy reasoning engines. In [17] a framework based on fuzzy Petri Nets for semantic content image analysis is proposed.

3. Proposed model for automatic multi-level image annotation

The proposed automatic multi-level image annotation of an unlabeled image can be roughly decomposed into following procedures: image

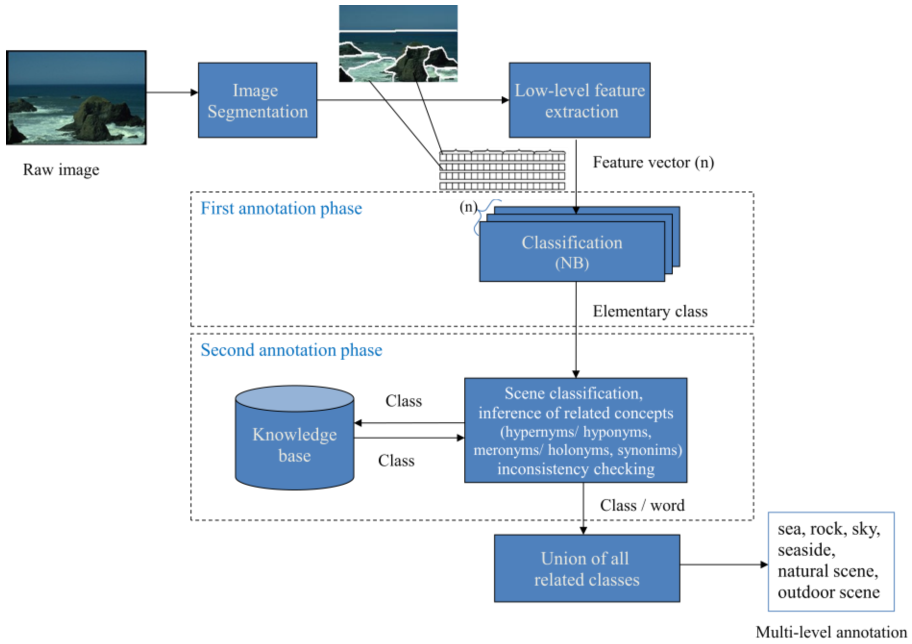

Fig. 1 Two phases of proposed automatic multi-level image annotation

segmentation, extraction of low-level features, definition of feature vectors, construction of classifiers and determination of more abstract concepts related to that image using inference algorithms defined on the knowledge representation scheme. The overview of the proposed model is given in Fig. 1.

An unlabeled image is first segmented using an algorithm for automatic segmentation. The segmentation algorithm divides the image into different regions based on feature homogeneity. From these regions, low-level region-based features are extracted such as average color, position, size, shape, etc.

Then, for each region a feature vector is defined and used for classification. Hence, image annotation problem is considered as a multi-class classification problem. In the first annotation phase we have used the Naive Bayes (NB) classifier to estimate the parameters necessary for classification. NB is one of the simplest probabilistic classifiers that can be trained very efficiently using a small amount of training data. The training data consists of low-level feature vectors obtained from the image segments and keywords from a controlled vocabulary.

The basic assumption is that components of a feature vector $\boldsymbol{x} = (x_1, x_2, \dots, x_n)$ are mutually independent, so that (1) applies:

$$P(x_1, x_2, \dots, x_n \,|C_j) \approx \prod_{i=1}^{n} P(x_i \,|C_j) \tag{1}$$

Based on the Bayes' theorem for each new segmented unlabeled image represented by $\boldsymbol{x}^{new}$ a classification result is determined by (2):

$$c_{MAP} = \underset{C_i \in C}{argmax}\, P(\boldsymbol{x}^{new}|C_i)\, P(C_i) \tag{2}$$

Note that evidence $P(\boldsymbol{x})$ is a scaling factor with a constant value when values of the feature variables are known. The values of $P(\boldsymbol{x}|C_i)$ and $P(C_i)$, $\forall C_i \in C$ are estimated on the basis of data in a training set.

The obtained classification results are elementary classes that correspond to the objects that can be directly recognized on the image like "*train*", "*airplane*" or "*sky*", as referred in [18].

The classes obtained after the first phase of annotation belong to the flat controlled vocabulary and have names written in lowercase. These classes are used as an entry to the knowledge representation scheme, based on a Fuzzy Petri Net (KRFPN) [3] in the second annotation phase.

The KRFPN scheme is used to enable inference of concepts at different semantic levels related to an image and to expand the level of initial vocabulary by inclusion of concepts in inheritance and compositional relationships, synonyms and other derived concept related to a given image. Classes obtained after second phase of annotation belong to the hierarchical structured vocabulary and have names written with an initial capital letter.

Therefore, union of all obtained concepts (classes) at different semantic levels makes the final results of multi-level image annotation.

To determine the parent classes in inheritance relationships (hypernyms/hyponyms also known as parent/child) the fuzzy inheritance algorithm is used for a given class. For example, for the elementary class "*airplane*", the parent classes defined in knowledge base are "*Vehicle*" and "*Man-made object*". These classes are according to [18] referred as generalized.

To determine the classes of the whole in the compositional relationships (holonyms/meronyms also known as whole-part) such as scene classes "*Mountain view*", "*Seaside*" etc., the fuzzy recognition algorithm with elementary classes as parts is used.

Both procedures of determining concepts on higher semantic levels that are related to an image will be discussed in detail in section 5.

The KRFPN scheme and fuzzy intersection algorithm [3] can also be used to improve classification results from the first step but this is beyond the scope of this paper.

4. Fuzzy Knowledge Representation Scheme

The goal of using the knowledge representation scheme is to enrich image annotation with words that are as similar as possible to the terms people use when they interpret these images. Therefore, concepts in a flat controlled vocabulary used for classification in the first annotation phase are associated with new concepts, using inheritance (is_a) relationship and compositional (is_part_of) relationship according to the expert knowledge.

A useful property of the utilized fuzzy knowledge representation scheme is that the degree of uncertainty to a particular concept or relationship can be expressed. That property is particularly important when handling with information that is neither entirely reliable nor valid, such as classification of automatically segmented regions.

4.1 Concepts in the KRFPN Scheme

The concepts of the KRFPN scheme include classes from a given domain at different semantic levels and their attributes.

In the knowledge base adopted for image annotation, a set of appropriate generalized classes (e.g. *GC = {OutdoorScenes, NaturalScenes, Man-madeObjects, Landscape, Vehicles, Wildlife, WildCatScene, AnimalScene, ... }*), a set of corresponding scene classes (e.g. *SC = {Seaside, Inland, Sea, Underwater, Space, SceneAirplane, SceneTrain, ...}*) and a set of derived or abstract classes that are "common" to human interpretation (e.g. *AC = {Summer ...}*) are defined according to the expert knowledge.

A set of attributes in the proposed model consists of the elementary classes that correspond to object that can be directly recognized in the image like *C = {airplane, train, shuttle, building, road, grass, ground, cloud, sky, coral, dolphin, bird, lion, ...}.*

4.2 Relations in the KRFPN Scheme

Three types of relationships are defined in the knowledge base: inheritance, compositional (attributed) and spatial relationships. Inheritance relationships are defined according to the expert knowledge while compositional and spatial relations are determined according to data in the training set.

The inheritance relationship $\Sigma_1 = \{is\_a\}$ is defined in order to take advantage of the hierarchies between parent classes from the set *GC* and child classes from the set *SC* or *GC*. These relationships are also used to link classes from the sets *SC* or *GC* to appropriate derived or abstract class from set *AC*.

The compositional relationship $\Sigma_2 = \{is\_part\_of\}$ is defined between a class that plays the role of the whole and its components (parts). It is treated as a special case of an attributed relation defined between a class (whole) and values of its attributes (parts). Here, a scene class is considered as a whole consisting of the elementary classes that represent its semantic parts.

To determine which elementary classes are the characteristic parts of a particular scene class, a modified Bayes rule is applied. It is assumed that components of each scene are independent and that a scene may contain several characteristic elementary classes. Therefore, instead of choosing an attribute with a maximum posterior probability, all those elementary classes with a posterior probability $P(SC_i|C_k), \forall_k C_k$ exceeding the marginal value ε for a given scene SC_i $\forall_i$ are selected (3):

$$M_s(SC_i) = \{C_j : \underset{i}{arg}\, P(SC_i|C_k) \approx \underset{k}{arg} \frac{P(C_k|SC_i)}{P(C_k)} \geq \varepsilon\}. \quad (3)$$

The spatial and pseudo-spatial relations are defined according to the spatial location of objects in real scenes and according to the co-occurrence of objects, respectively, $\Sigma_3 = \{is\ below,\ is\ above,\ occurs\ with,\ occurs\ not\ with\ ...\}$. Spatial and pseudo-spatial relations are defined between classes from set *C* and can be used to validate and adjust the results of classification obtained in the first annotation phase.

4.3 Degree of Truth of Relations

Given that a fuzzy knowledge representation scheme is used, the degree of truth from the interval [0,1] can be assigned to any relationship where 0 means "not true" and 1 "always true" [19].

For the inheritance relationship, the degree of truth is set to 1 because any exception, if it exists, can be modeled using a set of contradictions.

For the compositional relation between a scene and an elementary class, the degree of truth was determined using the posterior probability $P(SC_i|C_j), \forall_j\, C_j \in M(SC_i)$ estimated from the training data set, separately for each elementary class C_j that is chosen as a component of a given scene SC_i, (4).

$$f(t_k) = P(SC_i|C_j) = \frac{P(C_j|SC_i)P(SC_i)}{\sum_{i=1}^{s} P(C_j|SC_i)P(SC_i)}, \quad C_j \in M(SC_i),\ s = |SC| \tag{4}$$

4.4 Graphical Representation of the KRFPN

The KRFPN can be represented by a directed graph containing two types of nodes: places and transitions. Graphically, a place $p_i \in P$ is represented by a circle and transition $t_j \in T$ by a bar. The relationships based on input $I{:}T \rightarrow P^{\infty}$ and output $O{:}T \rightarrow P^{\infty}$ functions are represented by directed arcs. In a semantic sense, each place from a set *P* corresponds to a concept and any transition from a set *T* to a relation.

In Fig. 2 a part of the knowledge base is presented, showing compositional relationships between a particular scene class and appropriate

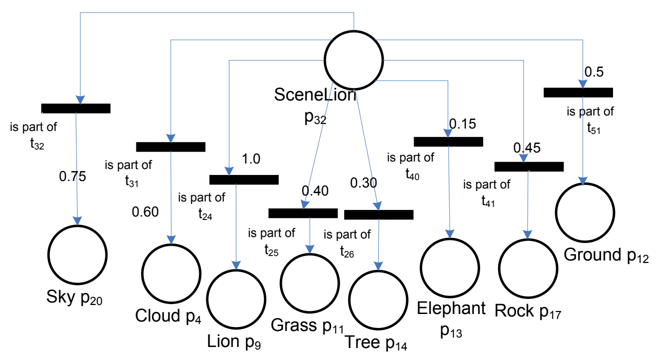

Fig. 2 Part of KRFPN scheme depicting the class "*SceneLion*" and compositional relations to appropriate elementary classes

elementary classes defined by (3) and (4). For example, the degree of truth of relation between the class "*SceneLion*" and its component elementary class "*lion*" is set to 1.0. As each concept in a knowledge base is associated with a place, the class "*SceneLion*" is assigned to place p_{32} and class "*lion*" to place p_9. Also, each relationship between concepts is assigned to a transition, so for example relationship "*is_part_of*" between classes "*SceneLion*" and "*lion*" is assigned to transition t_{75}.

5. Inference of Semantic Concepts Related to the Image

Fuzzy inference algorithms defined on knowledge representation scheme are used for determining concepts related to an image based on relationships defined in the knowledge base. More precisely, fuzzy recognition algorithm is used for scene classification and fuzzy inheritance algorithm for parent class assessment.

For both algorithms, assumption is that results of first image annotation phase are obtained. This means that unknown image is segmented, and then low-level region-based features are extracted and classified according to the maximum posterior probability (2).

The obtained classification results belong to the set *C* of elementary classes and are input to the KRFPN scheme.

5.1 Determining Scene Classes Using Fuzzy Recognition Algorithm

For the task of scene classification for a new, unknown image, a fuzzy recognition algorithm on the inverse KRFPN scheme is used. The inverse KRFPN scheme is obtained by replacing the input and output functions of KRFPN scheme and is denoted as –KRFPN [3].

For the part of a KRFPN scheme presented in Fig. 2 appropriate part of the inverse scheme, –KRFPN, is presented in Fig. 3.

The fuzzy recognition algorithm finds the class whose properties best match the given set of attributes and relations. Therefore, the elementary classes C_i obtained after first annotation phase are treated as parts (or attributes) of an unknown scene class X. It is assumed that these elementary classes exist in a knowledge base, so the places associated to them are marked with tokens $m_i \in M$.

More precisely, if elementary class "*grass*" is obtained after the first annotation phase, then the place p_{11} associated to that class will be marked with a token (Fig. 3). A token is represented as a black dot in a place.

To each token, a token value $c(m_k)$ is assigned. A token value of each input place corresponds to the

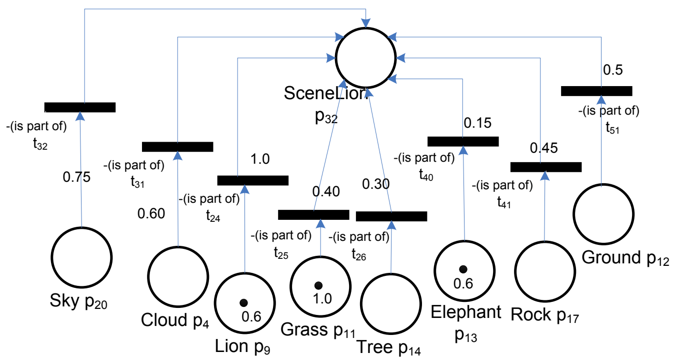

Fig. 3 Part of -KRFPN scheme depicting scene '*SceneLion*' and marked elementary classes

probability of the appropriate elementary class mapped to that place and is computed according to:

$$\mathrm{c}(\mathrm{m_k}) = \mathrm{c} + \mathrm{P}(\mathrm{C_i}),\ \ \mathrm{c} \geq 0 \qquad (5)$$

where c is an experimentaly determinated constant value and $P(C_i)$ is the a priori probability of the class C_i which was estimated from the relative frequency of occurrence of the class C_i in the training set.

According to the initial token distribution the root node π_0 of the recognition tree will be formed. Thus, for a token in place p_9, with token value 0.6, a root node $\pi_0(\mathrm{p9}, 0.6)$ will be created. The procedure is repeated for each marked place, so that the final number of the root nodes and consequently recognition trees depends on the number of marked places. For instance, for the knowledge scheme in Fig. 3 three recognition trees $\pi^\mathrm{i}, \mathrm{i} = 1{,}2{,}3$ with root nodes ${\pi_0}^i, i = 1{,}2{,}3$ will be formed:

$$\pi_0^1(p9, 0.6), \pi_0^2(p13, 0.6), \pi_0^3(p11, 1.0)$$

The tokens give dynamic properties to the net and define its execution by firing an enabled transition. The transition is enabled when every input place of the transition is marked, i.e. if each input place of the transition has at least one token. By firing, a token moves from all its input places to the corresponding output places.

In Fig. 4a, transition t_{24} is enabled and token moves from input place p_9 to the output place p_{32}. In the output place a new token value is computed as:

$$c(m_2) = c(m_1) * f\left(t_j\right) \qquad (6)$$

The new token value in the output place equals to the product of token value in input places $c(m_1)$ defined according to (4) and the degree of truth $f\left(t_j\right)$ assigned to transition according to (2) for compositional reletionships, or is set to 1 for inheritance relationships, (Fig. 4b).

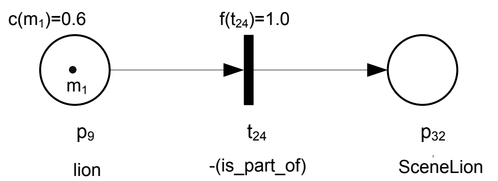

Fig. 4a Transition t_{24} is enabled.

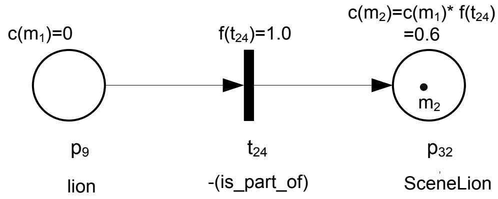

Fig. 4b Transition t_{24} has fired.

Firing of the enabled transition creates new nodes at the next higher level of the recognition tree with the new token values computed according to (6). Appropriate recognition tree with two new nodes formed from the root node $\pi_0^1(\mathrm{p9}, 0.6)$ by firing the enabled transitions t_{24} and t_{37}, is given in Fig. 5. Arcs of the recognition tree are marked with the transition value $f(t_j)$ and label of the transition t_j.

In the same manner an appropriate recognition tree will be formed for each root node (Fig. 6). The depth of search for recognition trees in Fig. 6 is set to 2.

Among these newly formed nodes only those whose components best suit the initial set of elementary classes should be selected.

To simplify the execution of the fuzzy

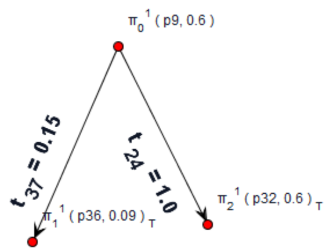

Fig. 5 Recognition tree formed by firing the enabled transition t_{24} and t_{37}

recognition algorithm each node is represented as a vector. For example node $\pi_1^1(\mathrm{p32}, 0.6)$ is represented by the vector $\boldsymbol{\pi}_1^1(0,..0{,}0.6{,}0, \dots ,0)$ that has at the 32$^{\mathrm{th}}$ position value 0.6, i.e. $\boldsymbol{\pi}_{0\,32}^1$=0.6.

Then, the total sum of all nodes in all b recognition trees corresponds to a total sum of

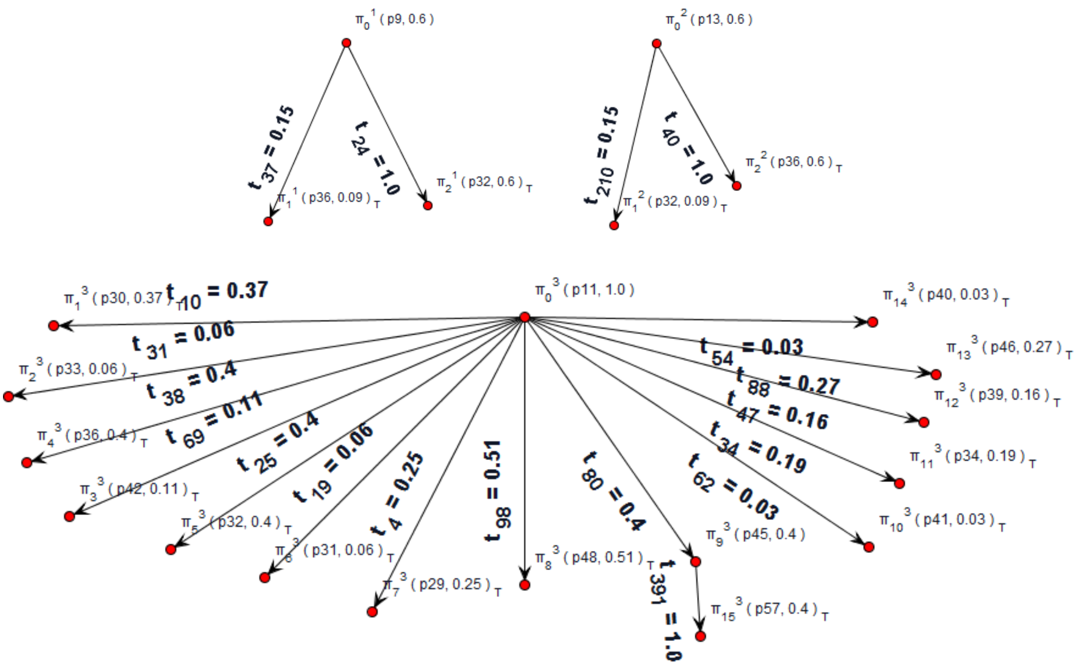

Fig. 6 Recognition trees formed after firing the enabled transitions for each of the root nodes that match initially obtained elementary classes.

vectors $\boldsymbol{\pi}_i{}^k, i = 1,2,..p \le n \;\; k = 1,2,..b \le |M|$ and is computed as:

$$\mathbf{Z} = \sum_{k=1}^{b} \sum_{i=1}^{p^k} \boldsymbol{\pi}_i^k. \qquad (7)$$

In the example in Fig. 6 number of recognition trees is b=3 and the number of nodes in each recognition tree is 2, 2 and 15, respectively. Hence, the total sum computed according to (7) is:

$\mathbf{Z} = \sum_{i=1}^{2} \boldsymbol{\pi}_i^1 + \sum_{i=1}^{2} \boldsymbol{\pi}_i^2 + \sum_{i=1}^{15} \boldsymbol{\pi}_i^3 =$ (0…0, 0.25$|_{29}$, 0.37, 0.06, 0, **1.09**$|_{32}$, 0.06, 0.19, 0, **1.09**$|_{36}$, 0, 0, 0.16, 0.03, 0.03, 0.11, 0, 0, 0.4, 0.27, 0, 0.51, 0 …0, 0, 0.4$|_{57}$, 0 …0).

The dimension of the vector **Z** corresponds to the number of nodes in the knowledge representation scheme. In order to clarify to which component of the vector **Z** a value applies, some components are marked with $|_{ind}$ where ind corresponds to indices of a vector component.

In the computed vector $\mathbf{Z} = (\mathbf{Z_1}, \mathbf{Z_2}, \dots, \mathbf{Z_n})$, the indices of elements with the highest sum are selected:

$$\mathrm{I}^* = \{\mathrm{i}^*: \arg \max_{i\,=\,1,..,n} \mathbf{Z}_\mathrm{i}\} \qquad (8)$$

Consequently, the scene class assigned to the place with indices $i^* \in I^*$ is chosen as the best match for a given set of elementary classes used as inputs to the scheme.

In this example, the maximum value of the vector **Z** of 1.09 is on the 32th and 36th components, $I^* = \{32,36\}$ so the results of the fuzzy recognition algorithm are concepts assigned to places p_{32} and p_{36}, that are *SceneLion* and *SceneElephant*, respectively.

Obtained classes can be used as root nodes for the next recognition process that will infer classes from the higher semantic levels.

5.2 Determining Parent Classes Using Fuzzy Inheritance Algorithm

To enrich the image annotation mainly with parent classes, synonyms and other concepts that can be assumed to be related to the image, the fuzzy inheritance algorithm on the KRFPN scheme is used.

The fuzzy inheritance algorithm determines attributes of classes, first locally and then at higher hierarchical levels. For a given $k \in \mathbb{N}$ the final tree of inheritance at the most k+1 level is constructed during the process of inheritance.

As the class of interest can be at different level of abstraction, whether at the level of the elementary class or the scene class or its parent classes, the key feature of the inheritance algorithm is that allows the representation of knowledge at different levels of abstraction.

For a given class that exists in the knowledge base, a root node is formed $\pi_0(\,p_k, 1.0)$ so that p_k corresponds to a place in a knowledge base that is associated to a given class. The inheritance tree is formed by firing the enabled transitions until the condition for stopping the algorithm is satisfied or the desired depth of inheritance tree reached.

Fig. 7 shows a 4-level inheritance tree of the KRFPN scheme for one of the scene classes, a "*WildCatScene*" and corresponding root node $\pi_0(p_{49}, 1.0)$. After semantic interpretation of nodes and arches displayed in Fig. 7 following conclusion is obtained: *WildCatScene (p49) is WildLife (p50) and AnimalScene (p51) and NaturalScene (p53) and OutdoorScene (p55).*

If a class apart from inheritance relations has defined composition relationships, the inheritance tree will show the parent class as well as its components (attributes).

In Fig. 8 a 2-level inheritance tree for the class *SceneCheetah* is presented. The highlighted nodes correspond to the parent classes (e.g. *WildCatScene*

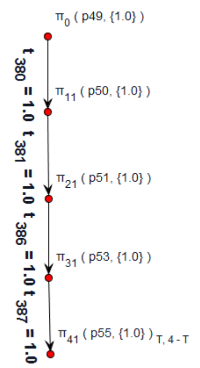

Fig. 7 The inheritance tree for a concept "*WildCatScene*"

(p_{49}) is_a *WildLife* (p_{50})), and the other nodes represent its components, the elementary classes.

6. Experimental Results

To evaluate the proposed automatic model for multi-level image classification, we have used a part of Corel image dataset [20].

Images were automatically segmented based on

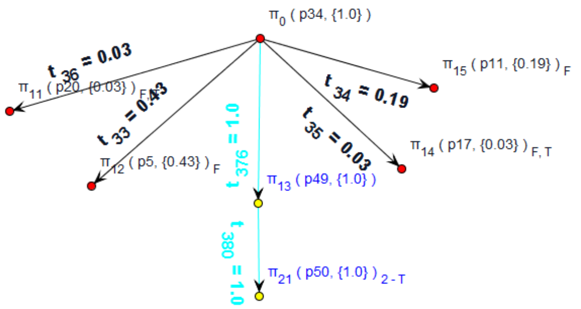

Fig. 8 The inheritance tree for a concept "*SceneCheetah*"

visual similarity of pixels using the Normalized Cut algorithm [21], so the segments do not fully correspond to objects. Every segmented region of each image is more precisely characterized by a set of 16 features based on color (average CIE L*a*b* color, standard deviation and skew of L*a*b* components), position (horizontal, vertical), size and shape of the region (width, height, boundary/ area ratio, convexity) [6].

Also, each image segment of interest was manually annotated with the first keyword from a set of corresponding keywords provided by [20] and used for supervised learning of the model. The vocabulary used to denote the segments have 28 words related to natural and artificial objects such as *'airplane'*, *'bird'*, etc. and the background like *'ground'*, *'sky'*, etc.

The data set used for the experiment consists of 3960 segments divided into training and test subsets by 10-fold cross validation with 20% of observations for holdout cross-validation.

We have used the Naïve Bayes classification algorithm to classify image segments into elementary classes according to (2). The results of automatic classification of image segments are compared with ground truth, so the precision and recall measures are calculated according to (9):

$$\text{Precision} = \frac{N_{cor}}{N_{Auto}}, \text{Recall} = \frac{N_{cor}}{N_{Man}} \qquad (9)$$

where N_{cor} is the number of segments that are correctly labeled with a given keyword (true positives), N_{Auto} is the number of segments that are automatically labeled with a given keyword (sum of true positives and false positives), N_{Man} is the number of segments that contain given word in the ground truth (sum of true positives and false negatives).

The results of classification of an image segment into elementary class are shown in Fig. 9 for each class. The obtained results differ greatly for various classes. The highest recall was obtained for classes *cheetah* (id=5), *coral* (id=7), *polar-bear* (id=15), and *tracks* (id=23). The highest precision was achieved for classes that was most common in images as *trees* (id=25), *sky* (id = 20), *grass* (id=11), and *water* (id =26). It should be mentioned that in determining the annotation results, the semantic similarity of words such as sky and cloud were not taken into account.

The obtained average precision and recall of the first annotation phase for all 28 classes are 34% and 26%, respectively.

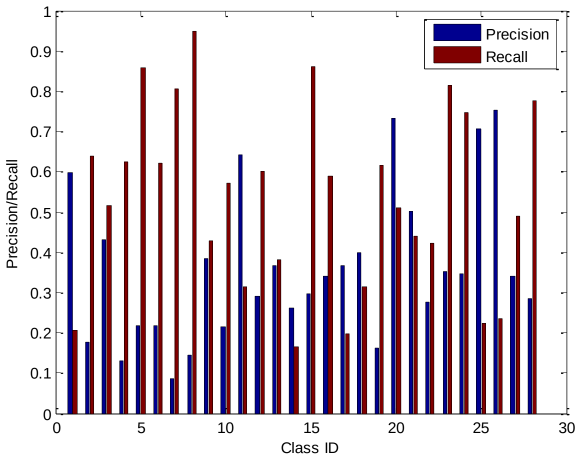

Fig. 9 A precision/recall graph for classification of image segments into elementary classes (displayed by ID)

Afterwards, based on the classification results of the first annotation phase and the knowledge base developed for a particular domain, an automatic image annotation on a higher semantic level was performed following the fuzzy recognition algorithm as explained in subsection 5.1 and the fuzzy inheritance algorithm as explained in subsection 5.2. In Table 1 few examples of the proposed multi-level image annotation are presented. The results of the first phase of image annotation are presented in the 1st row below each image and results of the second phase in the 2nd row below each image. The results in the 2nd row include scene classes and their parent classes.

Table 1: Examples of multi level image annotation

			
Phase 1. (NB)	train, tracks, sky	grass, tiger	water, sand, sky, road
Phase 2. (KRFPN)	Vehicle, Man-Made Object, Outdoor	Wildcat, Wildlife, Natural Scenes, Outdoor Scene	Coast, Landscape, Natural Scenes, Outdoor Scene

7. Conclusion

The aim of this paper is to propose a model for multi-level image annotation using the Naïve Bayes classifier and knowledge representation formalism based on fuzzy Petri Net (KRFPN). The Naïve Bayes classifier is used to classify low-level image features into elementary classes. The obtained elementary classes are used as inputs to the KRFPN scheme that is used to expand the level of vocabulary including multi-level semantics related to an image that will be used for image annotation.

To be more precise, a fuzzy inheritance algorithm on the KRFPN scheme is used for determining the parent concepts of a given class (hypernyms/hyponyms). A fuzzy recognition algorithm is used to determine the most appropriate scene class according to its compositional relationships with elementary classes (holonyms/meronyms). The complexity of the algorithms is polynomial, O(nm) where n is the number of places and m number of transitions in the KRFPN scheme.

The KRFPN has a hierarchical structure, and can be easily used as an upgrade and expansion of any classifier in order to enlarge a vocabulary used for annotation.

To improve the classification results, we plan several changes in the first annotation phase. We will consider using some saliency detection method as one proposed in [22] as classification results heavily depend on segmentation accuracy. Also, we plan to expand the feature set using invariant feature descriptors for interesting points on the object or regions like SIFT (Scale Invariant Feature Transform) [23], GLOH (Gradient Location-Orientation Histogram) [24] or some image descriptors for local interest regions more robust to illumination changes [24]. Then, the semantic similarity of words such as sky and cloud can be taken into account according to [21]. In the second annotation phase a fuzzy intersection algorithm on the KRFPN scheme can be explored for purifying the classification results obtained in the first phase. Also, it will be explored how the use of spatial and pseudo-spatial relations defined between elementary classes can improve classification results obtained in the first annotation phase.

This research is limited to the domain of outdoor scenes and the knowledge base includes knowledge that is relevant to that domain. However, the methodology of acquiring knowledge and reasoning in the KRFPN scheme is expandable and adaptable to the acquisition of new knowledge of a particular domain.

References

[1] Datta, R., Joshi, D., Li, J. 2008. “Image Retrieval: Ideas, Influences, and Trends of the New Age”, *ACM Transactions on Computing Surveys*, vol. 20, pp. 1-60, April 2008.

[2] Smeulders AWM, Worring M, Santini S, Gupta A, Jain R. 2000. Content-based image retrieval at the end of the early years. *IEEE Transaction on Pattern Analysis and Machine Intelligence*; 22 (12): 1349–1380.

[3] Ribarić, S., Pavešić, N., 2009. "Inference Procedures for Fuzzy Knowledge Representation Scheme", *Applied Artificial Intelligence*, vol. 23, January 2009, pp. 16-43.

[4] Tousch, A. M., Herbin, S., and Audibert, J. Y. (2012). Semantic hierarchies for image annotation: A survey. *Pattern Recognition*, 45(1), 333-345.

[5] J. Li and J. Z. Wang, ``Real-Time Computerized Annotation of Pictures," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, 2008, pp. 985-1002.

[6] Duygulu, P., Barnard, K., Freitas, J.F.G. de, Forsyth, D. A., 2002. "Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary", *ECCV* 2002, UK, May 2002, pp. 97–112.

[7] Monay F. and Gatica-Perez D., "On image auto-annotation with Latent Space Models", Proc. *ACM Multimedia*, Berkeley, CA, 2003, pp. 275–278.

[8] Zhang, D., Islam, M. M., and Lu, G. (2012). A review on automatic image annotation techniques. *Pattern Recognition*, 45(1), 346-362.

[9] Barnard, K., Duygulu, P., Forsyth, D., Freitas, N. de, Blei, D. M., Jordan, M. I., 2003. "Matching words and pictures," *Journal of Machine Learning Research* vol. 3, pp. 1107–1135, 2003.

[10] Fan, J., Gao, Y., Luo, H., Jain, R., 2008. "Mining Multilevel Image Semantics via Hierarchical Classification", *IEEE Transactions on Multimedia*, vol. 10, 2008, pp. 167-187.

[11] Benitez, A.B., Smith, J. R., Chang, S.F., 2000. "MediaNet: A Multimedia Information Network for Knowledge Representation", *Proc. IS&T/SPIE*, v. 4210, MA, November 2000.

[12] Schreiber, A. T. G., Dubbeldam, B., Wielemaker, J., Wielinga, B., 2001. "Ontology-based photo annotation," *IEEE Intelligent Systems*, vol. 16, no. 3, pp. 66–74, 2001.

[13] Srikanth, M., Varner, J., Bowden, M., Moldovan, D., 2005. "Exploiting ontologies for automatic image annotation", *Proc. SIGIR'05*, 2005, p. 552–558.

[14] Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L., 2009. "ImageNet: A Large-Scale Hierarchical Image Database".

[15] Papadopoulos, G.T., Mylonas, P., Mezaris, V., Avrithis, Y., Kompatsiaris, I., 2006. "Knowledge-Assisted Image Analysis Based on Context and Spatial Optimization", *Int. Journal on Semantic Web and Information Systems*, vol. 2, no. 3, pp. 17-36. July-Sept. 2006.

[16] Athanasiadis, T. et al. 2009. "Integrating Image Segmentation and Classification for Fuzzy Knowledge-based Multimedia", *Proc. MMM2009*, France, 2009.

[17] Ivašić-Kos, M. Ribarić, S.; Ipšić, I. "Image Annotation using Fuzzy Knowledge Representation Scheme", IEEE Proceedings of the 2010 International Conference of Soft Computing and Pattern Recognition. Pariz, France, 2010. 218-223

[18] Ivašić-Kos, M.; Ribarić, S.; Ipšić, I. Low- and High-level Image Annotation Using Fuzzy Petri Net Knowledge Representation Scheme. // *International Journal of Computer Information Systems and Industrial Management* (IJCISIM). 4 (2012) ;

[19] Chen, S.M., Ke, J.S., Chang, J.F., 1990. "Knowledge Representation Using Fuzzy Petri Nets", *IEEE Transactions on Knowledge and Data Engineering*, vol. 2, 1990, pp. 311-319.

[20] Carbonetto, P., Freitas, N. de, Barnard, K., 2004. "A Statistical Model for General Contextual Object Recognition", *Proc. ECCV* 2004, Czech Republic, May 2004, pp. 350-362.

[21] Shi, J., Malik, J, 2000. "Normalized cuts and image segmentation", *IEEE Trans. PAMI*, vol. 22, no. 8, pp. 888–905, 2000.

[22] Pesaranghader A., Muthaiyah S., 2013. Semantic Similarity Using First and Second Order Co-occurrence Matrices and Information Content Vectors, *WSEAS Transactions on Computers*, Vol. 12 (3), March 2013

[23] Luo S., Chen S. Saliency Detection Based on Path Price and Fuzzy Reasoning, *WSEAS Transactions on Computers*, Vol. 12 (5), May 2013

[24] Lowe D.G., Distinctive image features from scale-invariant keypoints, *International Journal of Computer Vision*, Vol.2, No.60, 2004, pp. 91-110

[25] Mikolajczyk K., Schmid C., A performance evaluation of local descriptors, *IEEE Trans Pattern Analyses and Machine Intelligence*, Vol.27, No.10, 2005, pp. 1615-1630.

[26] Shi W., Chen S., Fang L., Local Pixel Class Pattern Based on Fuzzy Reasoning for Feature Description, *WSEAS Transactions on Signal Processing*, Vol. 9 (2), April 2013