International Journal of Computer Information Systems and Industrial Management Applications.
ISSN 2150-7988 Volume 4 (2012) pp. 428-–436

Low- and High-level Image Annotation Using Fuzzy Petri Net Knowledge Representation Scheme

Marina Ivasic-Kos[1], Slobodan Ribaric[2], Ivo Ipsic[1]

[1]Department of Informatics, University of Rijeka,
Omladinska 14, 51000 Rijeka, Croatia
{marinai, ivoi}@uniri.hr

[2] Faculty of Electrical Engineering and Computing, University of Zagreb,
Unska3, 10000 Zagreb, Croatia
slobodan.ribaric@zemris.fer.hr

Abstract: **In order to exploit the massive image information and to handle overload, techniques for analyzing image content to facilitate indexing and retrieval of images have emerged. In this paper, a low-level and high-level image semantic annotation based on Fuzzy Petri Net is presented. Knowledge scheme is used to define more general and complex semantic concepts and their relations in the context of the examined outdoor domain. A formal description of hierarchical and spatial relationships among concepts from the outdoor image domain is described. The automatic image annotation procedure based on fuzzy recognition and inheritance algorithm, that maps high-level semantics to image, is presented together with experimental results.**

Keywords**:** image annotation, image interpretation, knowledge representation, Fuzzy Petri Net**.**

I. Introduction

Nowadays, digital images are ubiquitous and their number in databases is growing with an incredible speed.

Describing images by their semantic contents can facilitate users to index, retrieve, organize and interact with huge data using existing text searching techniques.

As the majority of the images are barely documented, current research on semantic image retrieval is closely related to automatic image annotation that works toward finding a solution to the problem of automatically linking keywords to an unlabelled image [1].

The basic premise of image auto annotation approaches is that a model describing how low-level image features like color, texture and shape are related to keywords can be learnt from a training set of images. Obtained model is then applied to un-annotated images in order to automatically generate keywords that describe their content. Usually, the keywords with the highest probability are chosen to annotate the image.

For solving the problem of automatic image annotation, many different approaches have been used. A recent survey of methods of image retrieval is given in [2]. Hereafter we will mention some referent methods to point out different approaches used for automatic image annotation. Methods based on translation model [3] and several extensions have assumed automatic image annotation to be analogous to translation problem between languages. Models which use Latent Semantic Analysis transform the features to a vocabulary of visual terms, which represents a purely visual language [4]. Renewal methods based on classifications are used for classifying images into a large number of categories [5]. Paper [6] gives a comparative analysis of (some existing) multi-label classification methods applied to different problem domains and evaluates the performance of such methods for different tasks among which is automatic image annotation.

Paper [7] introduces an algorithm based on the correlation mining of colors in order to utilize image interpretation and to improve the accuracy of image retrieval.

Under the assumption that the basic goal of annotation is the facilitation and improvement of image retrieval, the annotation should contain keywords which user might use during the retrieval.

According to [8], users' text-based queries consist of two words (on average) although their request is much more subtle, often representing an information or entertainment needs, that would normally require a deeper query of a higher semantic level than keyword or object token itself. For example, during retrieval of images from personal dataset, it is more intuitive to use the keyword "beach" instead of list of concepts like "sand, sea, sky, person" or other objects which can possibly be recognized on images belonging to the mentioned domain. Besides, query "wild-cats" means that someone is looking for "tiger", "lion", "leopard" and other wild cats.

For analyzing high-level semantics and searching images more intelligently, the ontology and descriptive logic are often pointed out. Some early work on semantic description of images using ontology was described in [9]. In [10], the ontology with hierarchical classification of image concepts is used to represent the semantics of the whole image. Then, due to the ambiguity and unreliability of facts, authors [11] have been trying to incorporate elements of fuzzy logic into ontology. Later, the same group of authors [12] has reported that environment used by the ontology is shown to be incompatible with that of fuzzy reasoning engines.

The paper reveals an approach to knowledge-based image annotation. The knowledge base is built using representation scheme based on Fuzzy Petri Net that is briefly presented in section two.

An example of knowledge base with formal description of hierarchical and spatial relationships among concepts from the outdoor image domain is described in the third section. The low-level and high-level image automatic annotation

procedure based on fuzzy recognition algorithm, and generation of abstract concepts at a higher semantic level using the fuzzy inheritance algorithm, are presented in section four and five, respectively. Experimental results of automatic image annotation and discussion are given in section six.

II. Knowledge Representation Scheme Formalism

In our approach, a knowledge representation scheme based on Fuzzy Petri Net theory, named KRFPN, is used for low- and high-level image automatic annotation.

The knowledge representation KRFPN, [13] is defined as 13-tuple:

$$KRFPN = (P, T, I, O, M, \Omega, \mu, f, c, \alpha, \beta, \lambda, Con), \quad (1)$$

where:

$P = \{p_1, p_2 \ldots p_n\}, n \in \mathbb{N}$ is a set of places;
$T = \{t_1, t_2 \ldots t_m\}, m \in \mathbb{N}$ is a set of transitions;
$I: T \rightarrow P^{\infty}$, is an input function;
$O: T \rightarrow P^{\infty}$, is an output function;
$M = \{m_1, m_2 \ldots m_r\}, 1 \leq r < \infty$, is a set of tokens;
$\Omega: P \rightarrow P(M)$, is a tokens' distribution within places;
$\mu: P \rightarrow \mathbb{N}$, is a marking of places;
$f: T \rightarrow [0,1]$, is the degree of truth associated with the transitions;
$c: M \rightarrow [0,1]$, is the degree of truth associated with the token;
$\alpha: P \rightarrow D$, maps each place $p_i \in$ P into single concept $d_j \in$ D, so $d_j = \alpha(p_i)$, i = 1,..n, j = 1,..|D|;
$\beta: T \rightarrow \Sigma$, maps each transition $t_j \in$ T into single relation $r_k \in \Sigma$, thus $r_k = \beta(t_j)$, j = 1,..m, k = 1,..|Σ|;
$\lambda \in [0,1]$, is a threshold related to transitions firing;
$Con \subseteq (D \times D) \cup (\Sigma \times \Sigma)$, contains pairs of a mutually contradictory relations and/or concepts.

The KRFPN can be represented by a bipartite direct graph containing two types of nodes: places and transitions. Graphically, places $p_i \in$ P are represented by circles, transitions $t_j \in$ T by bars. The relationships, based on input and output functions are represented by directed arcs. In the semantic sense, each place p_i corresponds to a concept from the set D, and any transition t_j to a relation from the set Σ (Figure 1).

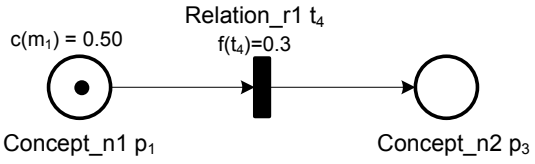

Figure 1. Fuzzy Petri net formalism (place, transition, token) with the associated semantic meaning

A dot in a place represents token $m_l \in$ M, and a place that contains one or more tokens is called the marked place. Complete information about the token is given by a pair $(p_j, c(m_i))$, where the first component specifies the place where the token is, and the second one, its truth value. Tokens give dynamic features to the net and define its execution by firing an enabled transition t_j. The transition is enabled when every input place of transition is marked, i.e. if each of the input places of the transition has at least one token and additionally the value $c(m_i)$, of each token exceeds the threshold value λ. An enabled transition t_j can be fired. By firing, a token is moving from all its input places $I(t_j)$ to the corresponding output places $O(t_j)$. In enabled transitions, token with the maximum value $c(m_i)$ takes the role in firing. After firing, new token value is obtained as $c(m_i) * f(t_j)$ in the output place, as shown in Figure 2.

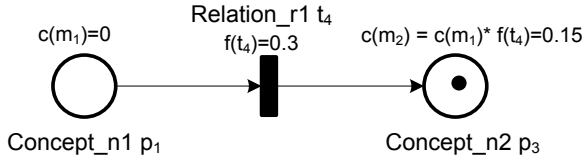

Figure 2. New token value is obtained in the output place after firing

Values $c(m_i)$ and $f(t_j)$ are degrees of truth assigned to a token at the input place $p_i \in I(t_j)$ and a transition t_j, respectively. Semantically, the value $c(m_i)$ expresses the degree of uncertain assignment of concept from the set D to the place p_i while the value $f(t_j)$ represents the degree of uncertain assignment of relationship from the set Σ to the transition t_j. The value of $c(m_i)$, $f(t_j) \in [0, 1]$, can be expressed by truth scales where 0 means «not true» and 1 «always true» [14].

Also, because of the uncertain and ambiguous semantic interpretation and exceptions in inheritance, a set Con of contradiction relations or concepts can explicitly be defined [13].

The inference procedures (inheritance and recognition) defined in KRFPN scheme use the dynamical properties of the net. More details about the KRFPN scheme and inference procedures can be found in [13].

III. Knowledge Base for Domain Images

To demonstrate a model of low and high-level image automatic annotation based on the KRFPN scheme, a part of the image dataset [15], of Corel Photo Library that includes natural and artificial objects and landscape, is used.

Each image from the dataset was annotated with the controlled vocabulary according to [3]. Figure 3 displays image samples and associated annotation.

water trees sky grass tiger

Figure 3. Example of images and annotations

Additionally, images are segmented using the Normalized Cut algorithm [16]. Segmentation is based on grouping of visual similarities of pixels without any concern about the object semantics (Figure 4), so segments do not fully correspond to the objects. We have considered only segments with the area bigger than 2% of the total image area.

Figure 4. Example of segmented images

In our experiment, each image segment of interest was manually annotated only with the first keyword from a set of corresponding keywords provided by [15] and used as ground truth for the training model.

A. Image features

Every segmented region of each image is more precisely characterized by a set of 16 feature descriptors; $A_k, \mathrm{k} = 1, 2 \ldots 16$. Features are based on color, position, size and shape of the region [15].

In order to simplify the model and to emphasize important information, image features are quantized. We have used the k-means algorithm with the squared Euclidean distance to decide how many quantization levels have to be created for every feature. The algorithm converges quickly, but the obtained solution may not be optimal because it depends on the initial set of cluster centroids and the selected values of k that define the number of clusters. The algorithm has to be run several times, each time with a new, randomly selected set of initial cluster centroids. In order to choose the optimal number of clusters (quantization levels), the value of k was increased until the mean of the sum of distances of points within the cluster to cluster centroids terminated to decrease significantly or when it was close to a zero.

B. Model definition

Here, we propose a simple model which maps image features to domain classes represented by keywords.

Analyzing the segments which belong to a certain class, by simple grouping the segments labeled by the same keywords together, the representative descriptor values for each class are computed. Values V_k of certain descriptive variables A_k typical for a certain class C_i have been chosen based on the probability of the intersection of descriptive value occurrence and class occurrence.

Each of the specific value of descriptor V_k is associated with the degree of probability, based on the conditional probability formula of multiple independent subsets of V_k:

$$P(V_k = v_{kj} | C = C_i) = \frac{P(V_k = v_{kj}, C = C_i)}{P(C = C_i)} = \frac{k(v_{k_j} \cap C_i)}{k(C_i)}, \quad (2)$$
$$\forall V_k \in V, k = 1, 2, \ldots, 16;\ V_k = Domain\ (A_k)\ ;$$

where:

$C = \{C_1, C_2 \ \ldots \ C_n\}$ is a set of classes;
$V = \{V_1, V_2 \ \ldots \ V_n\}$ is a set of descriptor values;
$V_k = \{v_{k1}, v_{k2} \ \ldots \ v_{k|\mathrm{Vk}|}\}$ is a set of values of descriptor A_k, k=1, 2, ... , m, where $|\mathrm{V_k}|$ is a number of quantization levels for descriptor A_k.

Descriptor values which have conditional probability lower than the threshold are equally associated to the nearest values of descriptors that are higher than the threshold. In this experiment the threshold was set to 0.05.

Because of intra-class variety, each class has usually more than one associated value of a certain descriptive variable. Thus, the occurrences which correspond to one class can be associated with different values of a descriptor.

In this model we have used two kinds of weighting: first, weighting the descriptors' impact to the classification performance and, second, weighting the descriptor values. We applied Quadratic Discriminate Analyses (QDA) filter [17] separately on each descriptor, and according misclassification error of the QDA algorithm, we assigned more weights to more discriminated descriptors (value) for the classification performance, $w(V_i), i = 1,2, \ldots,16$.

Furthermore, we applied some kind of Inverse Document Frequency (IDF) principle [18] that is mainly used while searching the text to reduce the effect of those terms that appear frequently in many documents and often are neither discriminate nor important for distinguishing documents. In this case, using a modified IDF principle to the attribute values, more weight is given to the values that occur rarely for particular descriptor as follows:

$$w_{IDF\,k,i} = w_{IDF}(v_{k,i}) = \log \frac{N}{1+n_{k,i}},\ \forall v_{k,i} \in V_k, \quad (3)$$

where $\mathrm{n_{k,i}}$ is the number of segments that have the value of $v_{k,i}$ for the attribute $\mathrm{A_k}$, and N is a total number of segments.

To compute the probability of the spatial relationships among classes, we analyzed mutual occurrence of classes in the image annotation. An n x n matrix, where n is a cardinality of the set C, is created. Each element of the matrix can be formally defined as:

$$P(C = C_j | C = C_i) = \frac{P(C=C_j,\, C=C_i)}{P(C=C_i)}, \mathrm{i} \neq \mathrm{j} \quad (4)$$

In order to model different occurrences of two or more equal classes in an image, the probability $\mathrm{P}(C_i|C_i)$ can be experimentally estimated.

C. Forming knowledge base for domain images

A semantic analysis and knowledge representation of domain images are focused on four semantic categories – elementary classes, generalized classes, derived classes and scene classes. Elementary classes correspond to the object labels which were directly identified in the images like "lion", "airplane", "grass", and "sky". Generalized classes include classes created by generalizing objects recognized in the image (e.g. "wildcat" is generalization of elementary classes "tiger" and "lion"). High-level generalization

includes generalization of already generalized classes like: “wildcat” (generalization of elementary class) – “wildlife” (generalization of generalized class) – “animal”, “natural object” (high-level generalization). Same abstract classes that are “common” to human association based on the identified image objects like “winter” for “snow” can be described by derived classes. Scene classes are used to represent the semantics of the whole image like “mountain view” and “outdoor”.

Classes from all semantic categories, according to the model based on KRFPN, are elements of a set D, where $D = C \cup Inst \cup V$.

A subset C includes elementary classes, generalized classes and related more abstract classes as scene and derived classes. Elements of a set C are initially generated according to the segments’ keyword for images from a training model as follows:

C = {Airplane, Bear, Polar-bear, Bird, Fox, Wolf, Lion, Elephant, Tiger, Cloud, Sky, Water, Trees, Grass, Rock, Send, Mountain, Snow, Plane, Train, Tracks, Roads}.

A Subset Inst includes instances of each class at the entrance to the knowledge base that have to be classified, but may also include some specific instances of the class of interest, which can be already stored in a knowledge base. In this experiment, it was not necessary to store instances in the knowledge base, so elements of the set Inst are instances of classes at the entrance to the knowledge base:

$Inst = \{X_i, i=1... n\}$.

A subset V represents class attributes and consists of descriptor values as determined by quantization of image region features as follows: A_1 – size of the region, A_2 - horizontal position (x), A_3 - vertical position (y), A_4 –width, A_5 – height, A_6 - boundary/area ratio, A_7 – convexity, A_8 - luminance (L), A_9 - green-red intensity (a), A_{10} - blue-yellow intensity (b) and A_{11} – std L, A_{12} - std a, A_{13} - std b, A_{14} - L skew coefficients, A_{15} - a skew coefficients, A_{16} - b skew coefficients. Descriptors A_1– A_7 refer to geometrical properties of the region while descriptors A_8 – A_{16} are related to CIE L*a*b color model components and corresponding standard deviation and skew coefficients of that components. A set of descriptor values is:

$V = \{V_1 = (size1, size2, ..., size10), V_2 = (xpos1, ..., xpos10), V_3 = (ypos1, ..., ypos11), V_4 = (hight1, …, hight7), V_5 = (width1, …, width6), V_6 = (bound1, ..., bound7), V_7 = (conv1, conv2, conv3), V_8 = (L1, ..., L5), V_9 = (a1, ..., a7), V_{10} = (b1, ..., b6), V_{11} = (stdL1, …, stdL6), V_{12} = (stda1, …, stda8), V_{13} = (stdb1, … , stdb5), V_{14} = (skewL1, ..., skewL12), V_{15} = (skewa1, skewa2, skewa3), V_{16} = (skewb1, skewb2, skewb3)\}$.

Relations from the set Σ are defined according to the expert knowledge on relations between concepts in the domain. The set Σ of relations is a union of hierarchical relations (Σ_1), relations between class C_i and values of its attributes from set V_k (Σ_2) and spatial and co-occurrence relationships among classes (Σ_3).

A set of relations $\Sigma_1 \cup \Sigma_2 \cup \Sigma_3$ is defined in the following manner:

Σ_1 = {is_a, is_part_of};

Σ_2 = {has_size, has_xpos, has_ypos, has_width, has_hight, has_boundary_area, has_convexity, has_Lum, has_green_red, has_blue_yellow, has_std_Lum, has_std_a, has_std_b, has_skew_Lum, has_skew_a, has_skew_b};

Σ_3 = {is_on, on_top, on_bottom, is_above, is_ below}.

In Figure 5 a part of knowledge base is presented, showing relations among particular classes from the set C and appropriate values of descriptors from the set V defined by the former procedure. To every transition from the set Σ_2 that models the relation among attribute values and a class, a probability (degree of truth) is assigned according to (2). For example, the degree of truth of relation between the particular class “Sky” and the descriptor value “ypos2” is 0.38 (Fig.5).

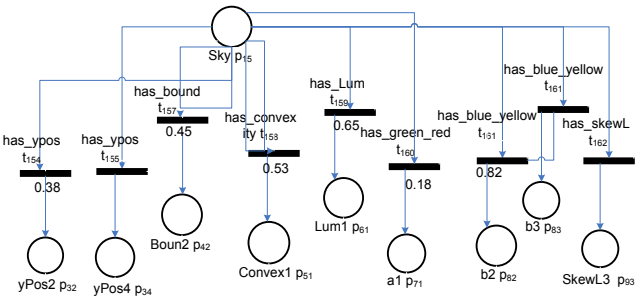

Figure 5. Part of relations among class ‘Sky’ and its attributes

From the set of spatial and co-occurrence relationships, we have mostly used relation “is_on”. The degree of truth is defined according to the frequency of mutual occurrence of classes in the image domain, according to (3). Figure 6 displays relations among class “Airplane” and classes that are usually found on the image when class “Airplane” is detected.

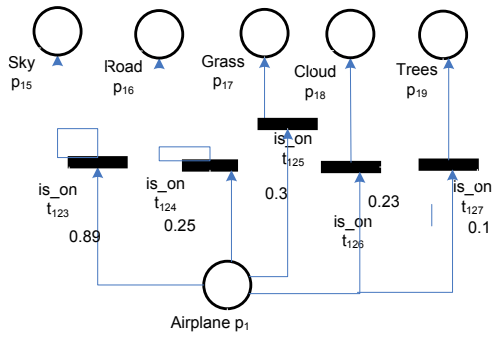

Figure 6. Example of co-occurrence relationships

More accurate spatial relationships (e.g., is_ below, is above, is near) among domain concepts, as presented in the Figure 7, are defined using general knowledge and analyzing the location of objects in the image. These relationships are useful for validating and improving the results of automatic image annotation.

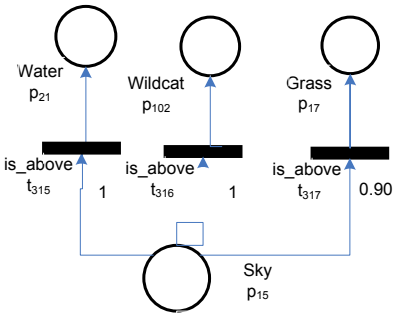

Figure 7. Example of spatial relationships

The set D of semantic concepts is, using expert knowledge, expanded with generalizations of concepts as well as with synonyms that can be useful in image retrieval. When defining the concepts generalization or synonyms, the lexical databases like WordNet [19] can be used.

Generalizations of concepts in the knowledge base are obtained by hierarchical relation as shown in Figure 8. The degree of truth is set according to the expert knowledge.

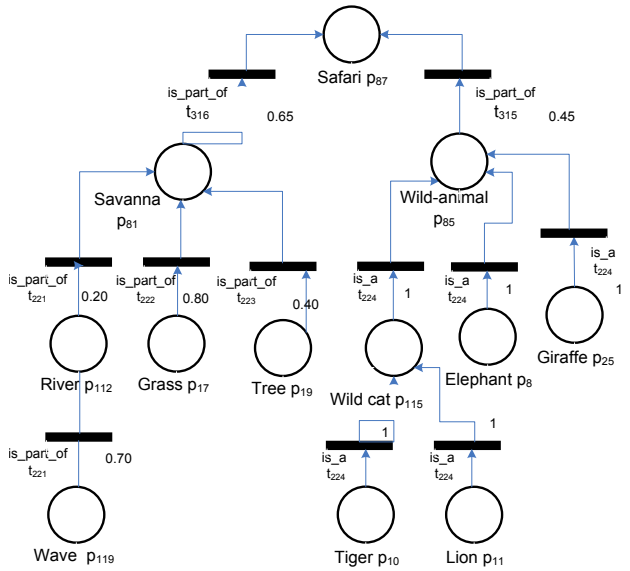

Figure 8. Including concepts of a higher semantic level into the knowledge database using hierarchy relations

Also, knowledge presented in Fig. 8 attempts to connect some keywords that can be used in text query like "safari" or "savannas", with appropriate concepts that can be detected in the corresponding image such as "tree", "grass" or similar. In this way, by including concepts of a higher semantic level into the knowledge base, concept organization in the natural language is transferred into the knowledge base to facilitate retrieval and manipulation of images.

IV. Image Annotation Procedure Based on Fuzzy Recognition Algorithm

For a task of automatic annotation of a new, unknown image, fuzzy recognition algorithm on inverse KRFPN scheme is used [13].

Assumption is that an unknown image is segmented and 16 feature descriptors are obtained from each segment. Thus, if there is a set of attribute values assigned to a segment that exist in the knowledge base, they are mapped to places $\{p_1, p_3, p_7 \ldots, p_k\}$. Corresponding token value $c(m_m)$ assigned to each place corresponds to IDF weight given to attribute value.

For initial token distribution, 16 recognition trees are formed $\boldsymbol{\pi}^1, \boldsymbol{\pi}^2, \ldots \boldsymbol{\pi}^{16}$, with root nodes that correspond to initially marked places.

For instance, if an input image is segmented and feature vector is extracted for every segment, after quantization, the quantization values obtained for each segment will be used as attributes of that segment.

Image in Figure 9 has four segments and corresponding attributes for the first segment are presented.

A1	A_2	A_3	….	A_{15}	A_{16}
size6	xpos10	ypos4		skewa2	skewb3

Figure 9. Image segments and corresponding attributes of a particular segment

Then, if attribute values exist in knowledge base and are assigned to places: $\alpha^{-1}(\text{size6}) = p_{34}$, $\alpha^{-1}(\text{xpos10}) = p48$, $\alpha{-}1\text{ypos4}{=}p53, \ldots$ $\alpha{-}1\text{skewb3}{=}p191$ with corresponding degree of truth: 0.87, 0.61, 0.53… 1, respectively, than 16 root nodes π_0 of recognition trees, will be formed: $\pi_0^1(p_{34}, 0.87), \pi_0^2(p_{48}, 0.61), \ldots \pi_0^{16}(p_{191}, 1)$

By firing of enabled transitions on inverse KRFPN scheme, new nodes on the following higher level of recognition tree are created. Appropriate values of tokens in a new node are obtained as follows:

$$c(m_{k+1}) = c(m_k) * f(t_k) = w_{IDF_{i,j}} * w(V_i)\, P\left(V_i = v_{i,j} \middle| C = C_l\right) P(C = C_l), \quad (5)$$

where $f(t_k)$ is truth value of arc t_k that links attribute value $v_{i,j}$ and place p_j assigned to the class C_l which is set according to (2) with appropriate descriptor weight $w(V_i)$ and $c(m_k)$ is a token value set according to (3).

Figure 10 shows the first and the last corresponding recognition trees in inverse KRFPN scheme with enabled transition starting from the root node for the example mentioned above. Nodes of the tree have a form $\left(p_j, c(m_l)\right)$ where $c(m_l)$ is a value of token m_l in place p_j. Arcs of tree are marked with a value $f(t_j)$ and a label of a transition $t_j \in T$ whose firing creates new nodes linked to elementary classes.

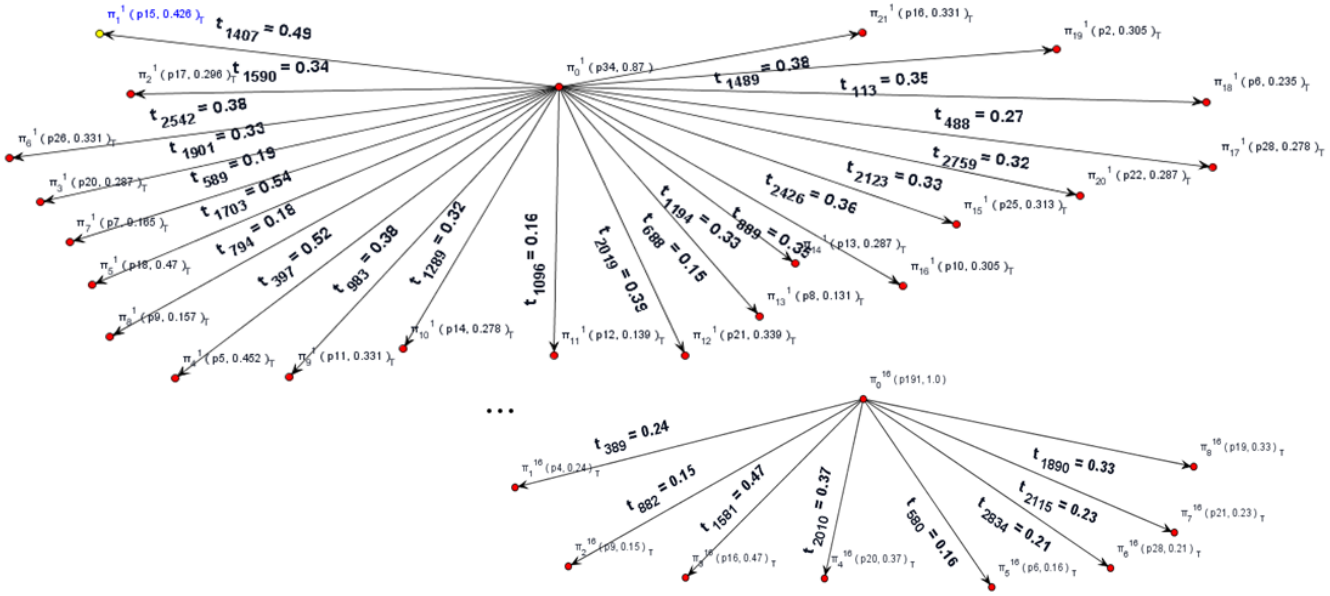

Figure 10. Partial view of recognition trees $\boldsymbol{\pi}^1$ ... $\boldsymbol{\pi}^{16}$ which root nodes correspond to attribute value of segment

For all levels of each recognition tree represented by vector $\boldsymbol{\pi}^k$ (k∈1, 2..., b; b= 16 for this example), the sum of nodes $\mathbf{z}^k$ is computed:

$$\mathbf{z}^k = \sum_{i=1}^{p} \boldsymbol{\pi}_i^k, \quad k = 1, 2, \dots, 16, \qquad (6)$$

where p is the number of nodes in the k-th recognition tree excluding the root node (p=21 for the first recognition tree).

Accordingly, total sum of all nodes for all recognition trees is given by:

$$\mathbf{Z} = \sum_{k=1}^{b} \mathbf{z}^k, \qquad (7)$$

where b is the number of all recognition trees (b=16 for this example).

If there are some initial properties that include the relations from the set Σ_3, then the recognition sub-trees with selective firing are constructed and all nodes without enabled transitions (terminal nodes) are computed by augmenting the total sum in (8). For the example above, the total amount of nodes in all recognition trees calculated by the formula (8) is as follows:

(p1{1.876}, p2{1.907}, p3{1.825}, p4{2.329}, p5{2.341}, p6{5.646}, p7{1.362}, p8{2.439}, p9{2.518}, p10{1.942}, p11{2.524}, p12{3.173}, p13{2.225}, p14{3.852}, p15{3.484}, p16{4.143}, p17{1.525}, p18{3.319}, p19{1.816}, **p20{6.331}**, p21{5.123}, p22{2.358}, p23{1.652}, p24{0.978}, p25{2.073}, p26{3.643}, p27{1.976}, p28{2.059})

Then, semantic concept assigned to the place that corresponds to a class with max argument of $\boldsymbol{Z} = (\boldsymbol{Z_1}, \boldsymbol{Z_2}, \dots, \boldsymbol{Z_n})$ is selected as the best match for given set of properties:

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

For this example $i^* = 20$, thus place p20 (indicated with bold font) is selected and after applying formula for semantic interpretation, $\alpha(p34) = 'Sky'$, class 'Sky' is selected as the best match for a given set of attribute values.

After recognition, image segments are classified into classes with the best matching as shown in Figure 11.

Figure 11. Results of recognition algorithm

Furthermore, obtained classes (that refer to elementary classes) can be used as root nodes for the next recognition process, on higher hierarchical level, that will infer concepts from higher semantic levels.

For instance, if a set of elementary classes with corresponding degree of truth as 'cloud', 'rock', 'sand', 'water' is obtained as image annotation, it will be treated as a set of attributes of an unknown scene class X. Then, four recognition trees $\boldsymbol{\pi}^i, i = 1,2,..4$ will be created applying formula (6) on the elementary classes and scene classes. After summing all nodes in all four recognition trees, a 'Seaside' is an interpretation of the place that is selected as the best match for a given set of elementary classes.

The important property of the fuzzy recognition algorithm is that the recognition trees are finite and that the execution of the recognition procedure is efficient and not computationally nor time demanding. More details and particular cases of inference procedures defined on KRFPN scheme can be found in [13].

V. Class Generalization Using Fuzzy Inheritance Algorithm

If properties of each class in the knowledge base and its relations with parent classes should be displayed or analyzed, fuzzy inheritance algorithm is most suitable for this purpose [13].

For a given class that exists in the knowledge base, the appropriate place is determined by the formula $\alpha^{-1}(d_j) = p_l$, $d_j \in D$. According to the initially marked place and appropriate token value, the initial token distribution is created representing the root node of inheritance tree.

Token value $c(m_k)$ can be set to a value obtained by the recognition algorithm. 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 is reached.

Semantic interpretation of arches that connect nodes in parent - child relationship forms statements and paths of inheritance. The resulting inheritance paths describe: class attributes (elements from set V for elementary classes and elementary classes for scene classes), spatial and pseudo-spatial relationships between elementary classes (e.g. “Airplane occurs with Sky” or “Airplane is above Water”) and parent classes whether of the elementary classes or of the generalized classes (e.g. “Airplane is Vehicle AND is Man-made Object AND is Outdoor Scene”).

For each of the inheritance paths the measure of truth is determined by the token value in the leaf node.

VI. Experimental Results

The data set used for the experiments is taken from Corel Stock photo library (prepared according to [15]). It consists of 475 segmented images (with a total of 4835 segments) that were divided into the training and the testing subsets by 10-fold cross validation with 20% of observations for holdout cross-validation. Each segment in the training set was initially annotated with one of the 22 semantic concepts.

Using a simple model which maps image features to domain classes and expert knowledge, a knowledge base was developed to represent the domain concepts of interest and their hierarchical and spatial relations.

After building the domain knowledge, an automatic semantic annotation of images in the test set can be performed following the fuzzy recognition algorithm on the proposed scheme.

Since the ground truth annotations of the images in the test set, concerning the image classes and suggested by humans, are known, it is possible to determine which obtained image annotations are relevant to a particular image and thus calculate precision and recall.

Precision is the ratio of correctly predicted keywords, and the total number of keywords that were obtained by automatic annotation, while recall is the ratio of correctly predicted keywords and all keywords for the image (ground-truth annotations).

Figure 12 presents the average per-word precision and recall for the automatic annotation experiments. The keywords (classes) are on precision-recall graph marked with class id.

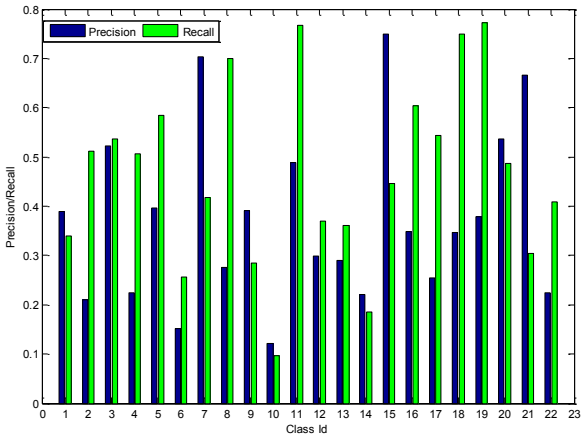

Figure 12. The average per word precision-recall graph

The results in Figure 12 show that better results are achieved for recall than for precision. The average precision rate is 37%, while average recall rate is 48%.

Good results, over 50% of precision are obtained for elementary classes with ID 3, 7, 11, 15, 20 and 21 that have labels 'bird', 'grass', 'polar-bear', 'sky', 'trees', 'water', respectively. The highest recall, over 70% was achieved for classes with ID 8, 11, 18 and 19 that correspond to labels 'ground', 'polar-bear', 'tracks' and 'train'. Only one class has precision and recall result less than 20%, a ‘mountain’ class, while the majority of other classes have results of precision and recall between 20 and 50%.

Some explanations of obtained results are as follows. Instead of class ‘mountain’, in automatic annotation class 'trees' is often used, as the mountains in the image are usually covered by trees, so it is difficult to decide, even when annotating manually, which label would better describe the image.

Furthermore, the class ‘cloud’ is often in automatic annotation replaced with ‘sky’ or vice versa since the boundary between these classes was not precisely defined in manual annotation, which was used for training. Similar case is with the classes 'ground' and 'grass'.

Some classes appear in only few images, as classes ‘fox’ with ID 6 and ‘sand' with ID 14, so the model for them might not be correctly set. Available solution to address this problem is using more samples and fine tuning of truth degree of the particular transitions. A similar solution can be used for classes that are not properly included in the model because of many variations in their appearance, caused either by bad segmentation, difference in illumination, occlusions, etc.

Obtained results can be considered as relatively good, especially when taking into account that a relatively small set was used for learning as well as that images were segmented automatically. Automatic image segmentation rarely achieves accurate segmentation of objects and obtained segments often include several different objects often combined with parts of the background as 'sky', 'grass' and 'trees'.

Moreover, higher semantic concepts that could not be directly identified in an image are not included in results depicted in Figure 12. Namely, elementary classes are obtained as results of low-level image feature classification,

so this is a reason that higher semantic concepts could not even be directly identified or included in evaluation of precision and recall.

We have used fuzzy recognition and inheritance algorithms and elementary classes to obtain the higher semantic level concepts. These concepts are used for image classification at more abstract level, such as scene classification.

In Table 1, examples of image annotation using both recognition and inheritance inference procedures are represented. The first row below each image in Table 1 shows results of low-level image classification, while the second row represents image annotations on more abstract level that involves classification of scenes and results of high-level class generalization.

Table 1. Examples of low-level and high-level image classification

		
'train', 'tracks', 'sky'	'dolphin', 'water'	'water', 'sand', 'sky', 'road'
'Train Scene', 'Vehicle', 'Man-Made Object', 'Outdoor'	'Dolphin scene', 'Sea', 'Natural Scenes', 'Outdoor Scene'	'Coast', 'Landscape', 'Natural Scenes', 'Outdoor Scene'

VII. Conclusion

Automatic image annotation has emerged as an alternative which can enhance image management and retrieval. The aim is to annotate image with concepts of a higher semantic level that will correspond to keywords which users intuitively use during image retrieval.

It is difficult to infer high-level semantics from the image features, because it is necessary to explore all image objects and their relations, and include knowledge necessary for semantic interpretation of overall image.

In this paper, the KRFPN formalism based on Fuzzy Petri Net was used for knowledge representation. This representation uses a simple graphical notation with just a few types of elements and has a well-defined semantics so the model is easily understood. The well-defined inference algorithms can be used for image annotations at various semantic levels of abstraction. The complexity of the algorithm is O (nm) where n is the number of places and m the number of transitions in KRFPN scheme.

Furthermore, an important property of the KRFPN formalism is the ability to show the uncertain knowledge using truth value of the concept and relation.

In the paper, a model which maps feature descriptors to domain classes is shortly specified. In this model two kinds of weighting that concern weighting the descriptors impact to the classification performance and weighting the descriptor values are used.

Also, a part of knowledge base that includes relationships among concepts, particularly generalization, spatial relationships and relationships among class and its attributes, is presented.

Moreover, higher semantic concepts that could not be directly identified in the image, like scene classes and generalization of classes, are inferred using fuzzy recognition and inheritance algorithms and elementary classes.

The preliminary research is limited to a particular domain of outdoor images, but we believe that our approach will be suitable to image databases from different domains because the methodology of acquiring knowledge and inference in the KRFPN scheme is expandable and adaptable.

References

[1] J. S. Hare, P. H. Lewis, P. G. B. Enser and C. J. Sandom, "Semantic facets: an in-depth analysis of a semantic image retrieval system", ACM international conference on Image and video retrieval, Amsterdam, Netherlands, vol. 6, July 2007, pp. 250-257.

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

[3] P. Duygulu, K. Barnard , J. F. G. de Freitas , D. A. Forsyth, "Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary", Proceedings of the 7th European Conference on Computer Vision, London, UK, May 2002, pp. 97–112.

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

[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] A. M. Santos, A. M. P. Canuto and A. F. Neto, "A Comparative Analysis of Classification Methods to Multi-label Tasks in Different Application Domains", International Journal of Computer Information Systems and Industrial Management Applications, (IJCISIM), vol. 3, 2011, pp. 218-227

[7] C. Y. Yue, Z. H. Ping and X. H. Song,"A Study on the Algorithm Based on Image Color Correlation Mining", International Journal of Computer Information Systems and Industrial Management Applications (IJCISIM), vol.1, 2009, pp.279-286

[8] http://comminfo.rutgers.edu/conferences/mmchallenge/2010/02/10/yahoo-challenge-image/ [06.10.2011]

[9] J. P. Schober, T. Hermes and O. Herzog, "Content-based image retrieval by ontology-based object recognition", Proc. of the KI-2004 Workshop on Applications of Description Logics - ADL2004, Ulm, Germany, 2004, pp. 61–67.

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

[11] T. Athanasiadis, P. Mylonas, Y. Avrithis and S. Kollias, “Semantic Image Segmentation and Object Labeling”, IEEE Transactions on Circuits and Systems for Video Technology, vol. 17, March 2007, pp. 298-312.
[12] T. Athanasiadis et al. , “Integrating Image Segmentation and Classification for Fuzzy Knowledge-based Multimedia”, Proc. MMM2009, France, 2009.
[13] S. Ribarić and N. Pavešić, “Inference Procedures for Fuzzy Knowledge Representation Scheme”, Applied Artificial Intelligence, vol. 23, January 2009, pp. 16-43.
[14] S.M. Chen, J. S. Ke and J.F. Chang, “Knowledge Representation Using Fuzzy Petri Nets”, IEEE Transactions on Knowledge and Data Engineering, vol. 2, 1990, pp. 311-319.
[15] P. Carbonetto, N. de Freitas and K. Barnard, “A Statistical Model for General Contextual Object Recognition”, Proceedings of the 8th European Conference on Computer Vision ECCV 2004, Prague, Czech Republic, vol. 1, May 2004, pp. 350-362.
[16] J. Shi and J. Malik, “Normalized cuts and image segmentation”, IEEE Trans. PAMI, vol. 22, no. 8, pp. 888–905, 2000.
[17] G.J McLachlan, Discriminant Analysis and Statistical Pattern Recognition, John Wiley & Sons, Hoboken, New Jersey, 1992, 2004.
[18] S. Robertson, "Understanding inverse document frequency: on theoretical arguments for IDF", Journal of Documentation, vol. 60, 2004, pp.503 – 520
[19] Fellbaum C. WordNet: An Electronic Lexical Database, MIT Press, 1998.

AuthorBiographies

Marina Ivasic-Kos was borne in 1973 in Rijeka, Crotia. She received her B.Sc. degree in mathematics and computer science from the Faculty of Philosophy, University of Rijeka and M.Sc. degree in information science from the Faculty of Philosophy, University of Zagreb in 1997 and 2201, respectively. Then, she has been enrolled in PhD postgraduate study of the computer science at the Faculty of Electrical Engineering and Computing in Zagreb. Since 1998 she is working at the Department of informatics, University of Rijeka, as teaching assistant for computer science courses. Her current research interests belong to the field of pattern recognition, computer vision and knowledge representation. Her work in these fields was presented at the international conferences.

Dr. Slobodan Ribaric is a Full Professor at the Department of Electronics, Microelectronics, Computer and Intelligent Systems, Faculty of Electrical Engineering and Computing, University of Zagreb. His research interests include Pattern Recognition, Computer Architecture, Knowledge Representation and Biometrics. He has published more than one hundred and forty papers on these topics (http://bib.irb.hr), and he is author of six books and co-author of one (An Introduction to Pattern Recognition). Professor Ribaric was involved in COST Action 275 “Biometrics on the Internet” and Network of Excellence FP6 “Biosecure”. He was a leader of Working group 2 of COST Action 2101 “Biometrics for Identity Documents and Smart Cards”. Ribaric is a member of IEEE and MIPRO.

Ivo Ipsic was born in 1963. He received his B.Sc., M.Sc. and Ph.D. degrees in electrical engineering from the Faculty of Electrical Engineering, University of Ljubljana, Slovenia in 1988, 1991 and 1996, respectively. From 1988–1998 he was a staff member of the Laboratory for Artificial Perception at the Faculty of Electrical Engineering, University of Ljubljana. Since 1998 Ivo Ipsic has been a professor of computer science at the University of Rijeka, teaching computer science courses. His current research interests lie within the field of pattern recognition. Ivo Ipsic is author of more than 50 papers presented at international conferences or published in international journals.