Image Annotation Using Fuzzy Knowledge Representation Scheme

Marina Ivasic-Kos, Ivo Ipsic
Department of Informatics
University of Rijeka
Rijeka, Croatia
{marinai, ivoi}@uniri.hr

Slobodan Ribaric
Department of Electronics, Microelectronics, Computer and Intelligent Systems
University of Zagreb
Zagreb, Croatia
slobodan.ribaric@zemris.fer.hr

Abstract— **In order to exploit the massive image information and to handle overload, techniques for analysing image content to facilitate indexing and retrieval of images have emerged. In this paper, a semantic image content analysis framework 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 algorithm, that maps high-level semantics to image, is presented.**

Keywords-component; image annotation; classification; knowledge representation; Fuzzy Petri Net

I. INTRODUCTION

Nowadays, the number of digital images 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 automatic image annotation (auto-annotation) that works toward finding a solution to the problem of automatically linking keywords to an unlabelled image [1].

The basic premise of auto-annotation approaches is that a model describing how low-level image features like colour, 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 auto-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 represent a purely visual language [4]. Renewal methods based on classifications are used for classifying images into a large number of categories [5].

Under the assumption that the basic goal of annotation is to facilitate and improve image retrieval, the annotation should contain words which user might use during the retrieval. According to [6], users' text-based queries consist of two words (on average) although their request is much more subtle, often representing an information or entertainment need, 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 word "beach" instead of list of concepts like "sand, sea, sky, person" or other objects which can be possibly recognised 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 analysing high-level semantics and searching images more intelligently, the ontology is often pointed out. Some early work on semantic description of images using ontologies was described in [7]. In [8], the hierarchical image concept ontology is used to represent the semantics of the whole image.

The paper reveals an approach to knowledge-based image annotation. The knowledge base is built using representation scheme based on Fuzzy Petri Net. A formal description of hierarchical and spatial relationships among concepts from the outdoor image domain is described. The image annotation procedure based on fuzzy recognition algorithm and experimental results of auto-annotation are presented.

II. KNOWLEDGE FORMALIZATION

In the recent years methods for formal knowledge representation by which an image can be described or interpreted, like ontology [8] and descriptive logic (DL) [7], are applied. Lately, due to the ambiguity and unreliability of facts, authors [9] have been trying to incorporate elements of fuzzy logic into ontology.

In our approach a knowledge representation scheme based on Fuzzy Petri Net theory, named KRFPN, is used for image auto-annotation. In this section we shortly describe the KRFPN formalism [10].

The knowledge representation KRFPN is defined as 13-tuple:

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

where:

$P = \{p_1, p_2 ... p_n\}$, $n \in \mathbb{N}$ is a set of places,

$T = \{t_1, t_2 ... t_m\}$, $m \in \mathbb{N}$ 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,
$\Omega: P \rightarrow \mathcal{P}(M)$, is a tokens' distribution within places,
$\mu: P \rightarrow N$, marking of places,
$f: T \rightarrow [0, 1]$, the degree of truth of the transitions,
$c: M \rightarrow [0, 1]$, the degree of truth of the token,
$\alpha: P \rightarrow D$, maps place from set P to concept from set D,
$\beta: T \rightarrow \Sigma$, maps transition from set T to relation in set Σ,
$\lambda \in [0, 1]$, threshold value related to firing of the transitions,
$Con \subseteq (D \times D) \cup (\Sigma \times \Sigma)$, is a set of pairs of 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 $p_j \in P$ 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 a semantic sense, each places p_i correspond to a concept from set D, and any transitions t_j to relation from set Σ (Fig. 1).

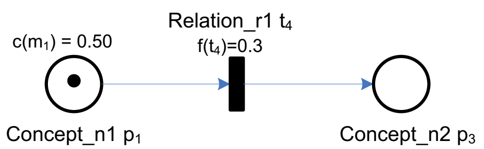

Figure 1. Fuzzy Petri net formalism (place, transition, token) with associated semantic meaning

A dot in a place represents token $m_i \in M$, and the place that contains one or more tokens is called a marked place. The 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 maximum value $c(m_i)$ takes the role in firing. After firing, new token value is obtained as $c(m_i) * f(t_i)$ in the output place, as shown in Fig. 2.

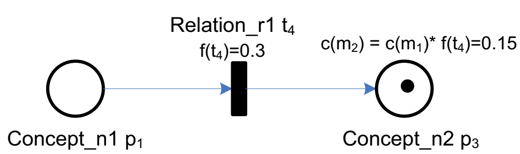

Figure 2. New token value is obtained in output place after firing

Values $c(m_i)$ and $f(t_i)$ are degrees of truth assigned to token at the input place $p_j \in I(t_i)$ and transition $t_i \in T$, respectively. Semantically, value $c(m_i)$ express the degree of uncertain assignment of concept from set D to place p_j and value $f(t_i)$ the degree of uncertain assignment of relationship from set Σ to transition t_i. Value of $c(m_i), f(t_i) \in [0,1]$, can be expressed by truth scales where 0 means «no true» and 1 «always true» [11].

Also, because of uncertain and ambiguous semantic interpretation and exceptions in inheritance, a set Con of contradiction relations or concepts can explicitly be defined [10].

The inference procedures (inheritance and recognition) defined in KRFPN scheme use the dynamical properties of the net. A more details about the KRFPN scheme and inference procedures can be found in [10].

III. Knowledge Base for Domain Images

To demonstrate a model of image annotation based on the KRFPN we have used a part of image dataset of Corel Stock Photo Library which include natural objects, parts of landscape and artificial objects [12].

Each image form the dataset was annotated with controlled vocabulary according to [3]. Fig. 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 (n-cut) algorithm. Segmentation is based on grouping of visual similarity of pixels without any concern about the object semantics (Fig. 4), so segments do not fully correspond to objects [3]. We have considered only segments with area bigger then 2% of the total image area.

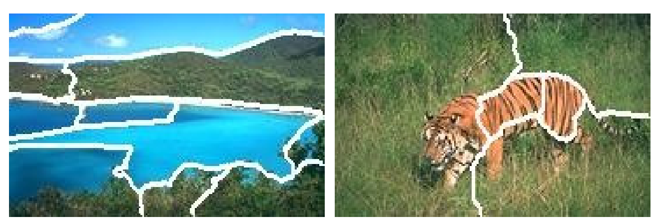

Figure 4. Example of images and annotations

In our experiment, each image segment of interest was manually annotated only with first keyword from a set of corresponding keywords provided by [12] 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, k=1, 2, …,16. Features are based on colour, position, size and shape of the region [12].

In order to simplify the model and to enhance the more important information, image features are quantizated. We have used the Expectation Maximization (EM) algorithm with cross validation to decide how many quantization levels have to be created for every feature.

B. Model definition

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

Analysing the segments which belong to a certain class, by simple grouping the segments labelled 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 is associated with a degree of probability, based on the conditional probability formula of multiple independent subsets of V_k:

$$\forall_i\ C_i \in C,\ i=1,2,\dots,n;\ \forall_k\ A_k \in A,\ k=1,2,\dots,16;$$
$$Dom(A_k) = V_k;$$
$$P(\bigcup_j V_{kj} \mid C_i) = \sum_j P(V_{kj} \cap C_i) / P(C_i); \qquad (2)$$

where:

$C = \{C_1, C_2 \dots C_n\}$ is a set of classes;

$A = \{A_1, A_2 \dots A_m\}$ is a set of descriptors;

$V_k = \{ V_{k1}, V_{k2} \dots V_{kl}\}$ is a set of values of descriptor A_k, and j = 1, 2, …l where l is number of quantization levels.

The 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 (QDE) filter [13] separately on each descriptor, and according misclassification error of the QDA algorithm we assigned more weights to more discriminated descriptors for a classification performance. Furthermore, we applied some kind of Inverse Document Frequency (IDF) principle [14] giving more weights to discrete values that occur rarely for particular descriptor.

To compute the probability of the spatial relationships among classes, we analysed 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:

$$P(C_j \mid C_i) = P(C_j \cap C_i) / P(C_i) \quad i \neq j \qquad (3)$$

In order to model possibilities when two or more equal classes appear in an image the probability $P(C_i|C_i)$ 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 – image classes, generalization classes, derived classes and scene classes. Image classes correspond to object labels which were directly identified in the image like "lion", "airplane", "grass", and "sky". Generalization classes include classes which were created by generalizing objects recognized in the image or in the case of high-level generalization by generalizing already generalized classes like: "lion" (image class) - "wildcat" (generalization of image class) – "animal"(high-level generalization). Same abstract classes that are "common" to human association based on 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 $D = C \cup I \cup V$. A subset C includes domain image classes, generalised classes and related classes as scene and derived classes. A subset I is used in case of some special instance of classes of interest. In this experiment instances were not used. A subset V represents class attributes and consists of descriptor values determinate by quantization of features of region 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. The descriptors A_8 – A_{16} are related to CIE Lab colour model.

A set of semantic concepts initially generated according segments' keyword for images from training model is:

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 set of instances is an empty set: $I = \emptyset$.

The set of descriptor values is:

$V = \{V_1$ = (size1, size2, ..., size10), V_2 = (xpos1, ..., xpos9), V_3 =(ypos1, ..., ypos6), V_4 = (hight1, …, hight7), V_5 = (width1, …, width6), V_6 = (boun1, ..., 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, ..., skewL6), V_{15} = (skewa1, skewa2, skewa3), V_{16} = (skewb1, skewb2, skewb3) }.

Relations from a set Σ are defined according to 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 relations 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_read, 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 Fig. 5 a part of knowledge base is presented, showing relations among particular class from set C and appropriate values of descriptors from set V defined by the former procedure. To every transition from set Σ_2 that models a relation among attribute values and class, probability (degree of truth) is assigned according to (2). For example, the degree of truth of relation between a particular class "Sky" and descriptor value "yPoz2" is 0.38 (Fig.5).

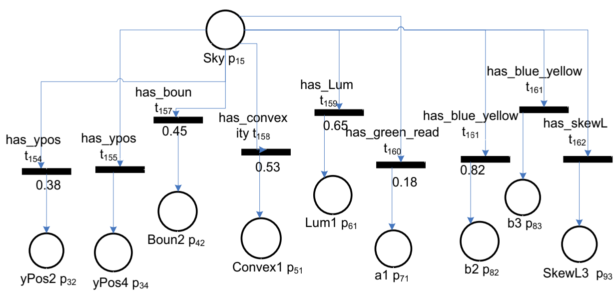

Figure 5. Example of relations among class and its attributes

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

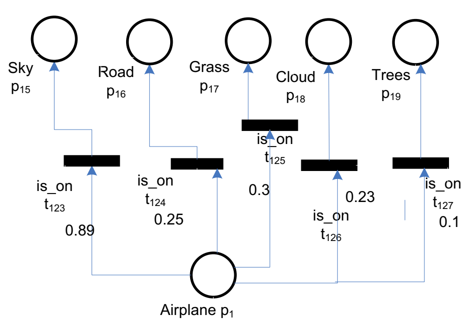

Figure 6. Part of spatial relationships among class "Airplane" and classes that can be found on the scene after the "Airplane" class is found

More accurate spatial relational descriptors (e.g., is_ below, on_top, on_bottom) among domain concepts can be defined to expend the set Σ_3 as presented in the Fig. 7.

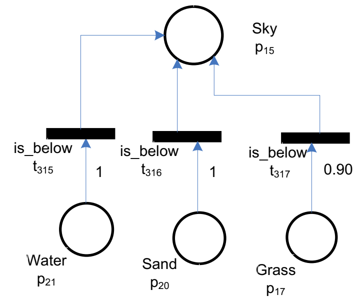

Figure 7. Part of spatial relations for class "Sky"

The set D of semantic concepts and the set Σ of relations are broad with generalization elements (e.g. generalizations for class tiger can be wild-cat, animal, natural object). Generalizations of concepts in the knowledge base are obtained by hierarchical relation as shown in Fig. 8.

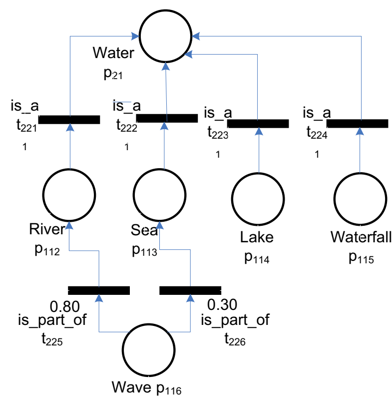

Figure 8. Part of hierarchy relations

Also, to improve the image annotation expanding the relations among words, particularly with synonymy and hierarchy relations among concepts, a lexical database like WordNet [7] can be used. Degree of truth of the transition form set Σ_1 is defined according to expert knowledge.

Furthermore, a set D can be expanded whether by joining and specializing class concepts identified in the image, (e.g. defining classes "leaves" and "branches" as parts of class "tree", or "locomotive" and "wagon" as parts of class "train") or completing descriptor list to corresponding low-level region features (e.g., texture) or some other kind of concepts or relations inferred by expert knowledge that can be useful in image retrieval.

Knowledge presented in Fig. 9 attempt to connect some words that can be use in query based on text with appropriate concepts that can be detected in the corresponding image.

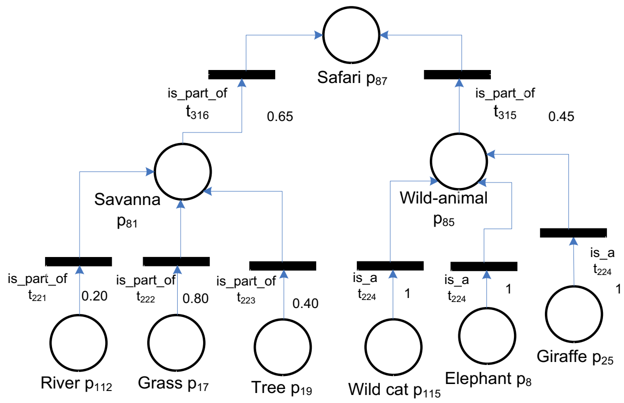

Figure 9. Including concepts of a higher semantic level into the knowledge database

In this way, by including concepts of a higher semantic level into the knowledge base, a concept organization in a natural language is transferred into the knowledge base to facilitate retrieval and manipulation of images.

IV. Image Anotation 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 [10].

Assumption is that unknown image is segmented and 16 feature descriptors are obtained. Thus, if there is a set of attributes assigned to a segment, that exist in knowledge base, they are mapped to places $\{p_1, p_3, p_7...., p_k\}$ with correspond token value $c(m_m)$ set according to the inverse value of a distance between a real value obtained from the image and a centre of quantization value. The initial marking of token distribution vector $\boldsymbol{\pi}_0$ is forming root nodes of the recognition trees [10].

By firing of enabled transitions new nodes on the following higher level of recognition tree are created and appropriate $c(m_i)*f(ti)$ values of tokens are obtained.

For all levels of each recognition tree represented by vector $\boldsymbol{\pi}^{\mathbf{i}}$ ($i \in 1, 2, ..., b$), the sum of nodes $\mathbf{z}^k$ is computed:

$$z^k = \sum_{i=1}^{p} \boldsymbol{\pi}_i^k, \quad k = 1, 2, \ldots, b, \tag{4}$$

where p is the number of nodes in the k-th recognition tree.

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

$$Z = \sum_{k=1}^{b} \mathbf{z}^k, \tag{5}$$

$Z = \sum_{k=1}^{b} z^k$,where b is the number of all recognition trees.

If there are some initial properties that include the relations from set Σ_3, than the recognition sub-trees with selective firing are constructed and all nodes with no enabled transitions (terminal nodes) are computed augmenting the total sum in (5). Accordingly, semantic concept assigned to a place that corresponds to the class with max argument of $\mathbf{Z}$ is selected as the best match for given set of properties:

$$i^* = \arg \max_{i = 1,..,n} \{\mathbf{Z}_i\}. \tag{6}$$

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 [10].

After recognition, image areas are classified into classes with best matching. Furthermore, obtained classes can be used as root nodes for next recognition process that will infer concepts from higher semantic levels.

Also, during inference process, the token value and the degree of trust into a certain transition is adjusted in order to model fuzzy facts correctly.

V. Experimental Results

The data set used for the experiments 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 10% 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 the 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 a human, are known, it is possible to determine which obtained image annotations are relevant to a particular image and thus calculate precision and recall. A recall is the ratio of correctly predicted keywords and all keywords for the image (ground-truth annotations), while a precision is the ratio of correctly predicted words, and the total number of words that were suggested.

Fig. 10 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.

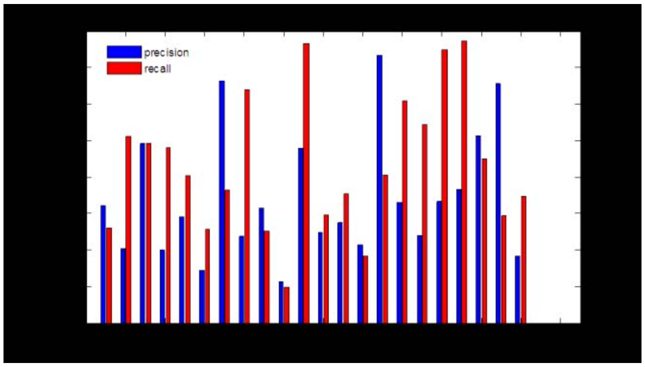

Figure 10. The average per word precision-recall graph

The results in Fig. 10 are obtained on small data set with frequencies less than 6% for most classes, so the model should be applied on more images for better model adjustments and more accurate parameters setting. Particularly, the critical moment in above explained knowledge representation schema is the initial relation among classes and its descriptor values. These relations are defined according to available samples, so for classes that occur with low frequency, the descriptor values might not be set correctly. Available solution to address this problem is using more samples and fine tuning of truth degree of the particular transitions.

Also, semantic similarity of class pairs such as “Cloud” and “Sky” are not taken into account to make the evaluations of results better. Furthermore, because of absence of test beds, higher semantic concepts, that could not be directly identified in the image, are not included in precision and retrieval calculation on Fig. 10, but are foreseen to have better results or at least equal to image classes.

VI. 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, which will correspond to keywords which users intuitively use during image retrieval.

It is hard 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 theory was used for knowledge representation. This representation uses 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.

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 generalisation, spatial and partial relationships, is presented.

The research is limited to specific domain of application, but we believe that our approach will scale well to larger databases containing similar images.

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] http://comminfo.rutgers.edu/conferences/mmchallenge/2010/02/10/yahoo-challenge-image/ [06.10.2010]

[7] 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.

[8] 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.

[9] 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.

[10] S. Ribarić and N. Pavešić, “Inference Procedures for Fuzzy Knowledge Representation Scheme”, Applied Artificial Intelligence, vol. 23, January 2009, pp. 16-43.

[11] 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.

[12] 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.

[13] G.J McLachlan, Discriminant Analysis and Statistical Pattern Recognition, John Wiley & Sons, Hoboken, New Jersey, 1992, 2004.

[14] S. Robertson, "Understanding inverse document frequency: on theoretical arguments for IDF", Journal of Documentation, vol. 60, 2004, pp.503 - 520