DEFINITION OF DESCRIPTORS FOR SEMANTIC IMAGE INTERPRETATION

Marina Ivašić-Kos, Patrizia Poščić, Mile Pavlić

Department for Computer Science, University in Rijeka, Omladinska 14, 51000 Rijeka, Croatia
E-mail: marinai@uniri.hr, patrizia@uniri.hr, mile.pavlic@ris.hr

ABSTRACT

A lot of effort has been put into researching image interpretation, but there is still no universally accepted approach to map low-level feature into high level image semantic interpretation [1]. In this paper, a method for continuous low-level features vector quantization is presented so as to define appropriate values for descriptive variables. The similarity among different concepts of the domain is examined and compared by using the measure of similarity which is based on the probabilistic model and the measure of distance. Also, an abstract image description vector suitable for image analysis is given.

Furthermore, formal explicit description of concepts and their properties as well as hierarchical relationship among concepts in an outdoor image domain will be presented.

Keywords **–** image representations, quantization, image classification.

1. INTRODUCTION

The main challenge of content-based image retrieval (CBIR) systems is to meet the user needs for semantic image retrieval. From a user's point of view, a CBIR system should enable apart retrieval of certain images by query by example (QBE) using only low level features directly extracted from an image, a textual queries which also include the semantic image interpretation. Examples of such queries are "find images of wild cats", "find images of outdoor landscape", etc.

Moreover, one should consider that user queries can consist of image tokens which are expected to be found in the wanted image, but usually these are formulated using semantic notions of a higher level than object labels, according to [1]. The problem of complexity, subjectivity and ambiguity of human image interpretation is mentioned as a semantic interpretation problem [2].

The effort of present CBIR systems is to use, apart from low level features like colour, texture and shape, the high level features which are the semantic interpretations of humans' visual perception.

Up to this point, the explored approaches and attempts to integrate semantics mostly relate to object detection, object recognition and automatic image annotation [2].

Current influential methods which link visual image features and corresponding concepts, i.e. denotations and concept keywords, are methods of annotation. Image analysis is essentially based on low-level features, and learning words for annotation is based on techniques of machine learning. Low-level features obtained as a result of algorithms for feature extraction are not sufficiently descriptive for determining image context [1]. By combining vectors of features, or some other kinds of representation using descriptive variables (abbrev. descriptors) appropriate for knowledge representation schemes, objects are recognized.

When objects are identified, they can get symbolic annotations, i.e. the name of the concept (class) which they belong to. Then, the labels of the concepts recognized in the image with the highest probability, like "trees, sky, wolf", are chosen to annotate the image.

Referent models mentioned in [1, 2] are e.g. CRM (Continuous-space Relevance Model) by Lavrenko, et al., then models which use Latent Semantic Analysis, as published by Monay and Gatica-Pereza, classificators as published by Chan in [12] that use SVM (Support Vector Machine), image retrieval systems like Alipr (http://alipr.com/) or Symplicity [11] etc. For viewing and analyzing high level semantics, ontology or description logic, as knowledge representation schemes, are often pointed out. Some examples are a SCULPTEUR system [9] that uses ontology to model contextual information about art objects in museum collections and a medical ontology like MIAKT (Medical Imaging and Advanced Knowledge Technologies, http://www.aktors.org/miakt/) for medical problem solving of breast cancer screening and diagnosis, or The Digital Anatomist Project, a complete ontology for biomedical concepts (http://sig.biostr.washington.edu/projects/da/). For solving the uncertain reasoning problems fuzzy ontologies or ontologies with extension of description logic are proposed as in [10].

In this paper, a method for continuous low-level features vector quantization is presented so as to define appropriate

values for descriptive variables. An abstract image

By using the measure of similarity which is based on the probabilistic model, the similarity among different classes of the domain is examined. Obtained results are compared to the measure of similarity which is based on the measure of distance.

Furthermore, formal explicit description of concepts and their properties as well as hierarchical relationship among concepts in an outdoor image domain will be presented.

2. CONTINUOUS FEATURES VALUE APPROXIMATION

Since image consists of image elements (pixels) which have no meaning, extracted features will, in a certain way, show one of the visual properties of the image or, more precisely, of the image segments. In this context, visual image properties are the content of the image which is usually shown using low level features, like colour, shape, texture, but can also be presented as any kind of information which can be derived from the image.

Without modification, a set of data from [3] was used, which relates to 400 outdoor images from Corel Stock Photo Library. Images include natural objects (animals, parts of landscape) and artificial objects.

In the learning set, images are segmented with normalized cut (n-cut) algorithm, so segments do not fully correspond to objects.

Each segmented area was associated with one or more keywords, i.e. concept label (class name). Segments from images that contain natural objects can be classified into animal and landscape classes. In mentioned domain we have considered bear, polar bear, bird, fox, wolf, lion, and elephant and tiger concepts. For landscape, cloud, sky, water, trees, grass, ground, rock, send, mountain and snow concepts were considered. The frequency of segments with mentioned concepts is shown in Fig. 1. The frequency of presented concepts is relatively small; only three concepts appear in more than 6% of cases.

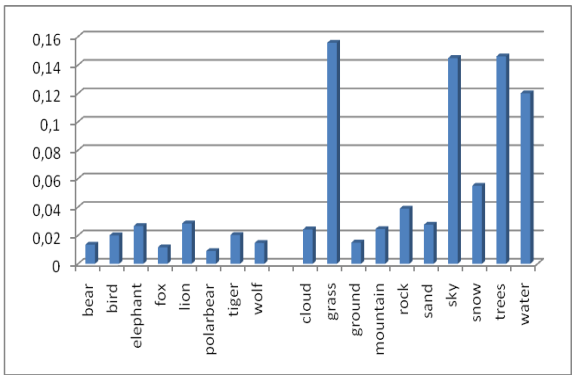

Figure 1. Frequency of natural objects' concepts

description vector suitable for image analysis is given.

Only four concepts of artificial object are used; plane, train, tracks and roads. Frequency of these concepts is also less than 0.05.

An important task for image retrieval is to choose relevant features shown using one or more corresponding feature descriptors, in order to form an abstract image description suitable for image retrieval and image analysis (so called signature) [1]. In retrieval, it is always wise to combine more features in order to generate a more robust image description.

For every segmented region of the above mentioned data, a set of 36 feature descriptors that compactly characterize each region's colour, position, texture, size and shape are calculated in [4]. As an alternative, a MPEG-7 standard format for visual descriptors of still images can also be chosen.

Hereunder, we have chosen 16 of them as relevant and sufficient for image descriptors as in [3]. The chosen feature descriptors are: size (normalized area of segment), position (horizontal and vertical position of barycentre, with their standard deviation), shape (convexity, boundary/area ratio, coefficient asymmetries of Lab components), and colour (luminance, green-red, blue-yellow corresponding to the average Lab components and standard deviation of Lab components).

For images from the outdoor domain, the precise information on the value of every feature does not play a crucial role in determining the class to which a certain segment belongs. Therefore, these are approximated with corresponding discrete variables in order to simplify the model.

Model simplification is, in this case, based on quantization of values which can be assumed by a certain feature of the image segment. In this way, the segment is no longer described with continuous values but with discrete ones or their corresponding linguistic descriptions.

For instance, in describing that a certain area belongs to the class '*Water'* from the given domain, the information that the area is big, that it is located at the bottom of the image and it is mostly blue, is as useful as the numerical features that the relative area size is 0.217433, with barycentre coordinates (0.769531, 0.735719), light intensity 82.2608, then -0.72716 in green and -10.6118 in blue colour intensity.

After the quantization, every image segment is described using an m-dimensional vector $[D_1\ D_2\ ...\ D_m]$.

Defined vector component are descriptors as follows: D_1 size, D_2 - horizontal position (x), D_3 - vertical position (y), D_4 - boundary/area ratio, D_5 – convexity, D_6 - luminance (L), D_7 - green-red intensity (a), D_8 - blue-yellow intensity (b) and D_9 - Lab skew coefficients.

More formally, for the given scheme $S = [D_1\ D_2\ ...\ D_m]$, the domain of the descriptor $Dom(D_i) = V_i$, for $i \in 1...m$, where $V = \cup V_i$, then:

$S \subseteq V_1 \times V_2 \times \ldots \times V_m = \{(v_1, v_2 \ldots v_m): v_i \in V_i\}$ and function f: S -> V so that $f(D_i) \in V_i$.

In other words, to every vector component D_i, $i \in 1...m$, correspond a descriptive variable with discrete values V_i, $i \in 1...m$.

Further on, every value of descriptive variable D_i can be given a descriptive meaning in order to improve the user interaction. For instance, the descriptor of size D_1 can be associated with values from the set V_1 = {low, middle, high} or V'_1 = {very low, low, middle, high, very high}.

Each value of these descriptive variables is mapped to an appropriate range of values of the corresponding low-level continuous features:

$$Q(V_i) \subseteq \{ \langle x_{ik}, x_{ik+1} \rangle : x_{ik} \in R_i \in \mathbf{R}; i=1..m; k \in \mathbf{N} \}.$$

It is not simple to determine how many values (clusters) will a certain descriptive variable have and what is the range of continuous features value that will be associated to it. Clusters are usually formed in such a manner that the intervals R_i of all possible values which a certain feature can assume are divided into disjunctive intervals of equal width.

In [5] the various value ranges for every low-level descriptor are chosen so that the resulting intervals are equally populated. Also, resulting intervals overlap. In [6] some low-level descriptors are grouped and presented with Gauss-mixture models.

In this paper authors have experimented with the irregular quantization which does not have the same period of quantization in the whole set of values of the data used for learning. In order to define the number and width of subintervals for possible values which will be associated to every descriptive variable, we used k-means algorithm with city block measure of distance, and Expectation Maximization algorithms with Euclidean measure of distance.

The achieved results are shown in Fig. 2.

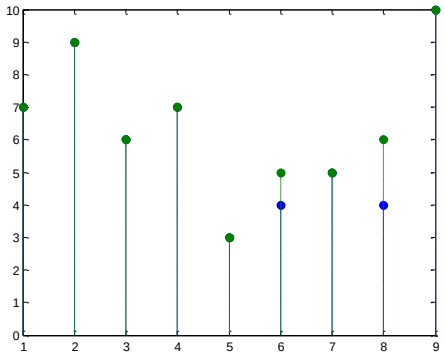

Figure 2. Clusters of descriptor values

The abscissa shows the descriptors D_i, and the ordinate shows the number of clusters into which their numerical values are partitioned. For instance, the values of features which correspond to descriptor D_5 are grouped into 3 groups, so D_5 will have 3 discrete values.

The results of quantization by using the above mentioned methods almost match, which shows that grouping is performed successfully.

Examples and text in the remainder of the paper will refer to quantization achieved through the k-mean method. For example, the range of values of continuous values which the feature *size* can assume is from 0 to 1, considering that the size of the segment is normalized. In the learning set, which is being used, 90% of features are in the interval from 0.02 to 0.65. Descriptive variable *'size'* has values {s1, s2, s3 … s7}. Each of the mentioned values is a representative of a cluster of continuous features with the centre in: {0.03, 0.07, 0.11, 0.16, 0.23, 0.34, 0.51}.

After the descriptive variables and cluster centres of their associated continuous values are defined, each sample is shown using these variables. Numerical features of the sample have been replaced with the value of the group whose centre is the closest to the given value.

For instance, for a random sample s, the vector below represents values of descriptive variables D_1, D_2 … D_9:

[s7 x5 y5 o2 c2 L2 a1 b3 k8].

Using the analysis of segments which belong to a certain class, i.e. based on the naive density estimation of the intersection of descriptive value occurrence and class occurrence, values of certain descriptive variables which are typical for a certain class have been chosen. For example, the attribute value of the class *'airplane'* for variable D_1 is s_1 with 56% of probability, for variable D_3 is y_3 with probability 74%, etc.

Because there are vast differences within the class to which the object belongs to, which include the difference in colour, area size the object takes, object's affine transformations, zoom differences, concept environment, overlapping and incomplete concepts, etc., the occurrences (samples) which correspond to one class are associated with different values of a descriptor. Therefore, it is foreseen for each of classes to have one or more associated values of certain descriptive variables, $f(D_i) \subseteq V_i$.

Below, attribute values of class descriptor *'Airplane'* is shown, following the signature described earlier:

({s6, s2}; {x2, x3, x6}; {y3, y4, y1}; {o7, o1}; {c1, c3}; {l2, l4}; {a4, a1}; {b1, b4}; {k10, k7})

Each of the specific value is associated with a degree of probability, based on the conditional probability formula:

$$P(D \mid C_i) = P(D \cap C_i) / P(C_i) \qquad (1)$$

i.e. its form for the function of multiple independent subsets of D (3):

$$P(\bigcup_k D_k \mid C_i) = \sum_k P(D_k \cap C_i) / P(C_i) \qquad (2)$$

where:
$\forall_i \; C_i \in C$, $C = \{C_1, C_2 \dots C_n\}$ is a set of classes;
$\forall_k \; D_k \in D$, $D = \{D_1, D_2 \dots D_m\}$ is a set of descriptors.

The values which have probability lower than the threshold are ignored and/or 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.

Each of the attribute values is also associated with a degree of reliability like (s6, 0.62), (s2, 0.38) in order to model fuzzy facts correctly.

3. THE COMPARISON OF SIMILARITIES AMONG CLASSES

In the previously described procedure, the continuous statistical marks of every feature were grouped and approximated in such a manner that all values from j-th cluster are approximated using the middle of the cluster. The middle of each cluster is associated with a discrete value of descriptor. Furthermore, for every class of the domain, descriptor vectors are determined which show the corresponding knowledge. By approximating feature values, a certain part of information is lost, but the possibility emerged to distinguish important features of continuous features from the irrelevant ones.

Below, the similarity among classes is compared using the measure of similarity presented in [13], applicable as long as there is a probabilistic model. In [14] the similarity between A and B is measured by the ratio between the amount of information needed to state the commonality of A and B and the information needed to fully describe what A and B are. By comparing the similarity among classes, it can be indirectly shown to what degree the approximation of data was successful. Furthermore, features can be detected, which are critical for the description of the class itself and for its differentiation from other classes. Such features are associated with weight in order to make the difference in relevant features more influential in the process of class comparison than the difference in less important features.

For the comparison of classes C_i and C_j, i.e. their vectors of descriptors D_i with discrete or ordinal values and known distribution of probability $P(D_i)$, we have applied the measure of similarity from [14] for words:

$$sim(C_i, C_j) = 2 \times I(D_i \cap D_j) / I(D_i) + I(D_j) \qquad (3)$$

where I(D) is the amount of information contained in a set D of features. Assuming that features are independent of one another and that P (Di) is probability of feature descriptor D_i, I (D) is calculated as:

$$I(D) = - \sum_{D_i \in D} \log P(D_i). \qquad (4)$$

When two classes have identical sets of features values, their similarity reaches the maximum value of 1. The minimum similarity 0 is reached when two classes do not have any common feature descriptor.

The results in figure 3 show, given the probability distribution, the similarity among class “Lion” and other classes in domain. The abscissa shows class marks, and the ordinate shows similarity. Each class is most similar to itself, like class “Lion”. Then, by similarity, follow classes “Ground” and “Wolf”. Class “Lion” is the least similar to class “Sky” and “Water”; sim (lion, water) = 0.0270; sim(lion, sky) = 0.0249.

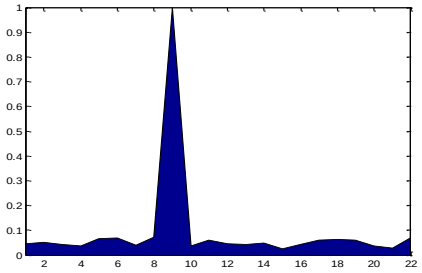

Figure 3. Similarity among class “Lion” and other classes in domain

Because of the comparison and estimation of information loss due to grouping data into clusters and quantisation, each class is also described using vectors of continuous values. Each component of the vector of a certain class is the median of corresponding features of samples which belong to the class. The median is chosen as the parameter due to its insensitivity to extreme values in the given data set. For the comparison of similarity among classes described in this manner, we used the measure of similarity:

$$sim_{dist} = 1 / (1 + dist(A,B)) \qquad (5).$$

The comparison of similarity using the stated measures of similarities is shown in Figure 5.

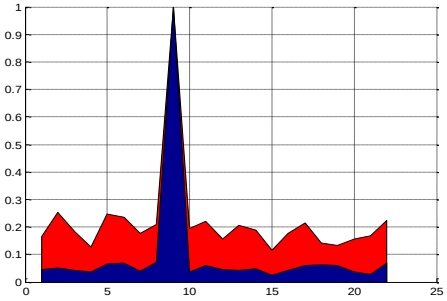

Figure 4. The comparison of similarities using different measures of similarity

The obtained percentages of similarity differ, but it is important that the ratios of similarity among classes match. In the mentioned measures, we did not include parameters of weight for certain descriptors. By adjusting the weight of certain descriptors, it is possible to influence the description of the class and increasing the difference among classes.

After the descriptors that describe classes are defined, and the measure of reliability is calculated and adjusted for every descriptor value, the knowledge on the domain needs to be included, in order to improve the classification of unknown segments in a-priori defined classes.

By connecting domain classes in taxonomy of a tree, semantic nets or ontologies, the semantic similarity among concepts can be determined, either based on the concept probability as in [14], or their distance in the taxonomy [13].

4. A KNOWLEDGE MODEL OF OUTDOOR IMAGE CLASSES

The problem outlined in this paper is how to determine a precise model for recording knowledge by which an image can be described or interpreted. During model creation, basic principles of knowledge organization were used, like: classification, generalization and hierarchy.

Fig. 5, by using Unified Modeling Language (UML) formalism [7], shows relation among image segments, descriptors and a class label to which the segment and/or image are associated with.

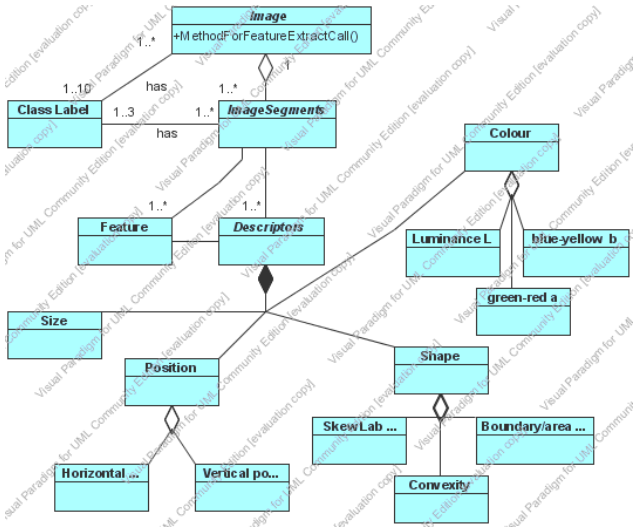

Figure 5. Relations among class and descriptors

These simplified models correspond to our domain and experiment, but can be expanded so as to include additional descriptors corresponding either to low-level region features (e.g., texture), relational descriptors (e.g., on the left of, on the right of, lower than, in front of) or to higher-level semantics which, in domain-specific applications, could be inferred either from the visual information itself or from associated information (e.g., annotation).

Classes chosen for image annotation in the former stage are arranged into a corresponding set of semantic concepts. Relations, mostly hierarchical and topological relations, are defined according to expert knowledge on relations between concepts in the domain.

Furthermore, to improve the image annotation expanding the relations among words, particularly with synonymy and hierarchy relations among concepts, a lexical database like WordNet [8] can be used.

What level of abstraction will represent a concept also depends on the database the image belongs to and user interest. Set C of initial classes for annotation can be broaden with elements which are obtained by generalizing (e.g. Wild-Cat, Vehicles), joining or distributing concepts (e.g. Leaves, Branches, Locomotive, Wagon) identified in the image. Topological relations can be defined among concepts; relations which describe the arrangement and co-occurrences of concepts on the scene. In this case, we only included co-occurrence relation.

In this way, by including concepts of a higher semantic level into the knowledge database, concept organization in a natural language is transferred into the database. Further on, linking images and concepts broadens image retrieval with visual image content to retrieval via text, i.e. keywords which describe and define the desired object more precisely.

4. CONCLUSION

The problem of automatic semantic image interpretation is complex, even when it relates only to images of similar type and the context of a specific domain.

The first step towards automatic semantic image interpretation is the definition of a model which is able to precisely, clearly, intuitively and visually show knowledge associated to the image interpretation, as illustrated in this paper.

The paper uses UML class diagram to model basic relationships between the classes and appropriated descriptors according to descriptor's vector selected to represent an image segment.

The paper shortly specifies the procedure for transformation of continuous values of features into discrete ones. The quantization of descriptor values is defined using the k-means and EM algorithm so the quantization intervals depend on the data. After the quantization and approximation of continuous features to discrete, descriptor values which are typical for a certain class are determined. Furthermore, due to ambiguity and incomplete information, it is necessary to adjust and fine-tune the reliability of descriptor values or descriptor values itself.

Furthermore, the impact of transforming numerical into descriptive variables on similarities among classes from the knowledge base has been analysed. The similarity of classes

described in discrete values is based on probability, and in the case of continuous values on distance. The results of similarity among classes obtained by different metrics vary, but their ratios match.

In further work, an analysis should also be conducted on how the adjustment of descriptor values and assignment of the weight parameters affects the results of classification and image annotation.

5. REFERENCES

[1] R. Datta, D. Joshi, and J. Li, "Image Retrieval: Ideas, Influences, and Trends of the New Age", ACM Trans. on Computing Surveys 2008; 20.

[2] J.S. Hare, et al., "Mind the Gap: Another look at the problem of the semantic gap in image retrieval", Multimedia Content Analysis, Management and Retrieval, San Jose, California, 2006; 6073: 1-12.

[3] P. Carbonetto, N. Freitas and K. Barnard, "A Statistical Model for General Contextual Object Recognition". In 8th European Conference on Computer Vision ECCV, Prague, Czech Republic, 2004(1): 350-362.

[4] P. Duygulu et al., "Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary", Proc. of the 7th European Conference on Computer Vision, London, UK, 2002; 97–112.

[5] M. Srikanth, J. VarnerJ, M. Bowden and D. Moldovan, "Exploiting Ontologies for Automatic Image Annotation", Proc. of the 28th Annual Inter. ACM SIGIR Conference on Research and Development in Information Retrieval, Salvador, Brazil, 2005;552-558.

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

[7] G. Booch, J. Rumbaugh and I. Jacobson, *The Unified Modeling Language User Guide*, 6nd Edn, Addison Wesley, NY, 2000

[8] C. Fellbaum, "WordNet: An Electronic Lexical Database", MIT Press, 1998.

[9] M. Addis, et al., "SCULPTEUR: Towards a new paradigm for multimedia museum information handling", 2nd Int. Semantic Web Conference, October 2003; 582–596.

[10] J.P. Schober, T. Hermes and O. Herzog, "Picturefinder: Description Logics for Semantic Image Retrieval", IEEE International Conference on Multimedia, Amsterdam, July 2005; 1571-1574

[11] J.Z. Wang, J. Li and G. Wiederhold, "Simplicity: Semantics-sensitive integrated matching for picture libraries", IEEE Trans PAMI, vol. 23, p. 947, 2001

[12] Y. Chen, X. Zhou and T.S. Huang, "One-class SVM for learning in image retrieval", Proc. IEEE ICIP, 2002

[13] P. Resnik, „Using information content to evaluate semantic similarity in a taxonomy", Proceedings of IJCAI, Montreal, Canada, p. 448–453, 1995

[14] D. Lin, „An Information-Theoretic Definition of Similarity", Proc. of the 15th International Conference on Machine Learning, pp. 296 – 304, 1998.

.