ISBN 978-83-60810-22-4

Automated annotation system for natural images

Gabriel Mihai, Liana Stanescu
University of Craiova,
Faculty of Automation,
Computers and Electronics,
Bvd. Decebal,
No.107, Romania.
{ mihai_gabriel, stanescu}@software.ucv.ro

Abstract—**Automated annotation of digital images remains a highly challenging task. This process can be used for indexing, retrieving, and understanding of large collections of image data. This paper presents an image annotation system used for annotating natural images. The proposed system is using an efficient annotation model called Cross Media Relevance Model for the annotation process. Image's regions are described using a vocabulary of blobs generated from image features using the K-means clustering algorithm. Using SAIAPR TC-12 Dataset of annotated images it is estimated the joint probability of generating a word given the blobs in an image. The annotation process of each new image starts with a segmentation phase. An original and efficient segmentation algorithm based on a hexagonal structure is applied to obtain the list of regions. Each meaningful word assigned to the annotated image is retrieved from an ontology derived in an original manner starting from the hierarchical vocabulary associated with SAIAPR TC-12 and from the spatial relationships between regions.**

Keywords—Image annotation, image segmentation, ontology, relevance models.

I. Introduction

THE automated task used to assign semantic labels to images is known as automatic image annotation. The importance of this task has increased with the growth of the digital images collections. It is a challenge that has been identified as one of the hot-topics in the new age of image retrieval [26]. Image annotation is a difficult task for two main reasons: semantic gap problem - it is hard to extract semantically meaningful entities using just low level image features and the lack of correspondence between the keywords and image regions in the training data.

Representing the content of the image using image features and then performing non-textual queries like color and texture is not an easy task for users. They prefer instead textual queries and this request can be satisfied using automatic annotation.

There are many annotation models proposed and each model has tried to improve a previous one. These models were splitted in two categories:

a) Parametric models: Co-occurrence Model [1], Translation Model [2], Correlation Latent Dirichlet Allocation [4]

b) Non-parametric models: Cross Media Relevance Model (CMRM) [3], Continuous Cross-Media Relevance Model (CRM) [10], Multiple Bernoulli Relevance Model (MBRM) [11], Coherent Language Model (CLM) [12]

The annotation process implemented in our system is based on CMRM. Using a set of annotated images [20] the system learns the joint distribution of the blobs and words. The blobs are clusters of image regions obtained using the K-means algorithm. Having the set of blobs each image from the test set is represented using a discrete sequence of blobs identifiers. The distribution is used to generate a set of words for a new image.

Each new image is segmented using an original segmentation algorithm [13] which integrates pixels into a grid-graph. The usage of the hexagonal structure improves the time complexity of the methods used and the quality of the segmentation results. An evaluation of this algorithm against other well know segmentation algorithms like Normalized Cuts segmentation algorithm [14], Efficient Graph-Based segmentation algorithm [15], Mean-Shift segmentation algorithm [16], Color set back-projection algorithm[17] is presented in [17][18]. This algorithm was also used for an image annotation system presented in [24].

The meaningful keywords assigned by the annotation system to each new image are retrieved from an ontology created in an original manner starting from the information provided by [20]. The concepts and the relationships between them in the ontology are inferred from the word's list, from the ontology's paths and from the existing relationships between regions.

The remainder of the paper is organized as follows: related work is discussed in Section 2, Section 3 provides details about the segmentation algorithm used, Section 4 contains a description of the annotation model, Section 5 presents the dataset used for experiments, Section 6 provides a description of the modules included in system's architecture, Section 7 contains the evaluation of the annotation system and Section 8 concludes the paper.

II. Related Work

Object recognition and image annotation are very challenging tasks. For this reason a number of models using a discrete image vocabulary have been proposed for the image annotation task. One approach to automatically annotating images is to look at the probability of associating words with image regions. Mori et al. [1] used a Co-occurrence Model

978-83-60810-22-4/$25.00 © 2011 IEEE

in which they looked at the co-occurrence of words with image regions created using a regular grid. To estimate the correct probability this model required large numbers of training samples. Each image is converted into a set of rectangular image regions by a regular grid. The keywords of each training image are propagated to each image region. The major drawback of the above Co-occurrence Model is that it assumes that if some keywords are annotated to an image, they are propagated to each region in this image with equal probabilities.

Duygulu et al [2] described images using a vocabulary of blobs. Image regions were obtained using the Normalized-cuts segmentation algorithm. For each image region 33 features such as color, texture, position and shape information were computed. The regions were clustered using the K-means clustering algorithm into 500 clusters called “blobs". The vector quantized image regions are treated as “visual words” and the relationship between these and the textual keywords can be thought as that between two languages, such as French and German. The training set is analogous to a set of aligned bitexts - texts in two languages. Given a test image, the annotation process is similar to translating the visual words to textual keywords using a lexicon learned from the aligned bitexts. This annotation model called Translation Model was a substantial improvement of the Co-occurrence model.

Jeon et al. [3] viewed the annotation process as analogous to the cross-lingual retrieval problem and used a Cross Media Relevance Model to perform both image annotation and ranked retrieval. The experimental results have shown that the performance of this model on the same dataset was considerably better than the models proposed by Mori et al. [1] and Duygulu et al. [2]. The essential idea is that of finding the training images which are similar to the test image and propagate their annotations to the test image. CMRM does not assume any form of joint probability distribution on the visual features and textual features so that it does not have a training stage to estimate model parameters. For this reason, CMRM is much more efficient in implementation than the above mentioned parametric models.

There are other models like Correlation LDA proposed by Blei and Jordan [4] that extends the Latent Dirichlet Allocation model to words and images. This model is estimated using Expectation-Maximization algorithm and assumes that a Dirichlet distribution can be used to generate a mixture of latent factors.

In [5] it is proposed the use of the Maximum Entropy approach for the task of automatic image annotation. Maximum Entropy is a statistical technique allowing predicting the probability of a label given test data. The image is represented using a language of visterms (visual terms) which are clusters of rectangular regions.

In [6][25] it is described a real-time ALIPR image search engine which uses multi resolution 2D Hidden Markov Models to model concepts determined by a training set. A computational efficiency is obtained in [25] due to a fundamental change in the modeling approach. In [6] every image was characterized by a set of feature vectors residing on grids at several resolutions. The profiling model of each concept is the probability law governing the generation of feature vectors on 2-D grids. Under the new approach, every image is characterized by a statistical distribution. The profiling model specifies a probability law for distributions directly.

In [7] Latent Semantic Analysis (LSA) [8] and Probabilistic Latent Semantic Analysis (PLSA) [9] are explored for automatic image annotation. A document of image and texts can be represented as a bag of words, which includes the visual words – vector quantized image regions and textual words. Then LSA and PLSA can be deployed to project a document into a latent semantic space. Annotating images is achieved by keywords propagation in this latent semantic space.

An improved model of CMRM is proposed in [10], the Continuous Cross-Media Relevance Model (CRM) which preserves the continuous feature vector of each region and this offers more discriminative power. A further extension of the CRM model called the Multiple Bernoulli Relevance Model (MBRM) is presented in [11]. The keyword distribution of an image annotation is modeled as a multiple Bernoulli distribution, which only represents the existence/nonexistence binary status of each word.

All the above mentioned methods predict each word independently given a test image. They can model the correlation between keywords and visual features but they are not able to model the correlation between two textual words. To solve this problem, in [12] it is proposed a Coherent Language Model (CLM) extended from CMRM. This model defines a language model as a multinomial distribution of words. Instead of estimating the conditional distribution of a single word it is estimated the conditional distribution of the language model. The correlation between words is explained by a constraint on the multinomial distribution that the summation of the individual words distribution is equal to one. The prediction of one word has an effect on the prediction of another word.

III. The Segmentation Algorithm

For image segmentation we have used an original and efficient segmentation algorithm [6] based on color and some geometric features of an image. The novelty of our algorithm concerns two main aspects:

a) minimizing the running time - a hexagonal structure based on the image pixels is constructed and used in color and syntactic based segmentation

b) using an efficient method for segmentation of color images based on spanning trees and both color and syntactic features of regions. A similar approach is used in [7] where image segmentation is produced by creating a forest of minimum spanning trees of the connected components of the associated weighted graph of the image.

In figure 1 it is presented the hexagonal structure used by the segmentation algorithm:

A particularity of this approach is the basic usage of the hexagonal structure instead of color pixels. In this way the hexagonal structure can be represented as a grid-graph G = (V, E) where each hexagon h in the structure has a corresponding vertex v ϵ V, as presented in Figure 1. Each

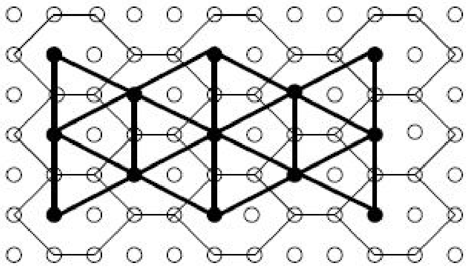


Fig. 1. The grid-graph constructed on the hexagonal structure of an image

hexagon has six neighbors and each neighborhood connection is represented by an edge in the set E of the graph. To each hexagon two important attributes are associated: the dominant color and the coordinates of the gravity center. For determining these attributes were used eight pixels: the six pixels of the hexagon frontier, and two interior pixels of the hexagon.

Image segmentation is realized in two distinct steps:

c) a pre-segmentation step - only color information is used to determine an initial segmentation. A color based region model is used to obtain a forest of maximum spanning trees based on a modified form of the Kruskal's algorithm. For each region of the input image it is obtained a maximal spanning tree. The evidence for a boundary between two adjacent regions is based on the difference between the internal contrast and the external contrast between the regions

d) a syntactic-based segmentation - color and geometric properties of regions are used. It is used a new graph which has a vertex for each connected component determined by the color-based segmentation algorithm. The region model contains in addition some geometric properties of regions such as the area of the region and the region boundary. A forest of minimum spanning trees is obtained using a modified form of the Boruvka's algorithm. Each minimum spanning tree represents a region determined by the segmentation algorithm.

IV. The Annotation Model

The Cross Media Relevance Model is a non-parametric model for image annotation and assigns words to the entire image and not to specific blobs – clusters of image regions, because the blob vocabulary can give rise to many errors. Some principles defined for the relevance models [22, 23] are applied by this model to automatically annotate images and for ranked retrieval. Relevance models were introduced to perform a query expansion in a more formal manner. Given a training set of images with annotations this model allows predicting the probability of generating a word given the blobs in an image. A test image I is annotated by estimating the joint probability of a keyword w and a set of blobs:

$$P(w,b_1,\dots,b_m) = \sum_{J\in T} P(J)P(w,b_1,\dots,b_m|J).$$

For the annotation process the following assumptions are made:

a) it is given a collection C of un-annotated images

b) each image I from C to can be represented by a discrete set of blobs I=$\{b_1 \dots b_m\}$

c) there exists a training collection T, of annotated images, where each image J from T has a dual representation in terms of both words and blobs: J=$\{b_1 \dots b_m; w_1 \dots w_n\}$

d) P(J) is kept uniform over all images in T

e) the number of blobs m and words in each image (m and n) may be different from image to image.

f) no underlying one to one correspondence is assumed between the set of blobs and the set of words; it is assumed that the set of blobs is related to the set of words.

$P(w,b_1,\dots,b_m|J)$ represents the joint probability of keyword w and the set of blobs (b_1,…,b_m) conditioned on training image J. An intuitive interpretation of this probability is how likely w co-occurs with individual blobs given that we have observed an annotated image J.

In CMRM it is assumed that, given image J, the events of observing a particular keyword w and any of the blobs $(b_1,\dots,b_m)$ are mutually independent, so that the joint probability can be factorized into individual conditional probabilities. This means thatP(b_1,…,b_m |J) can be written as:

$$P(w,b_1,\dots,b_m|J) = P(w|J)\prod_{i=1}^{m} P(b_i|J$$

$$P(w|J) = (1-\alpha_J)\frac{\#(w,J)}{|J|} + \alpha_J\frac{\#(w,\ T)}{|T|}$$

$$P(b|J) = (1-\beta_J)\frac{\#(b,J)}{|J|} + \beta_J\frac{\#(b,\ T)}{|T|}$$

where:

a) *P*(w|J) , P(w|J) denote the probabilities of selecting the word w, the blob b from the model of the image J.

b) #(w, J) denotes the actual number of times the word w occurs in the caption of image J.

c) #(w, T) is the total number of times w occurs in all captions in the training set T .

d) #(b, J) reflects the actual number of times some region of the image J is labeled with blob b.

e) #(b, T) is the cumulative number of occurrences of blob b in the training set.

f) |J| stands for the count of all words and blobs occurring in image J.

g) |T| denotes the total size of the training set.

h) The prior probabilities *P(J)* can be kept uniform over all images in *T*

The smoothing parameters α and β determine the degree of interpolation between the maximum likelihood estimates and the background probabilities for the words and the blobs

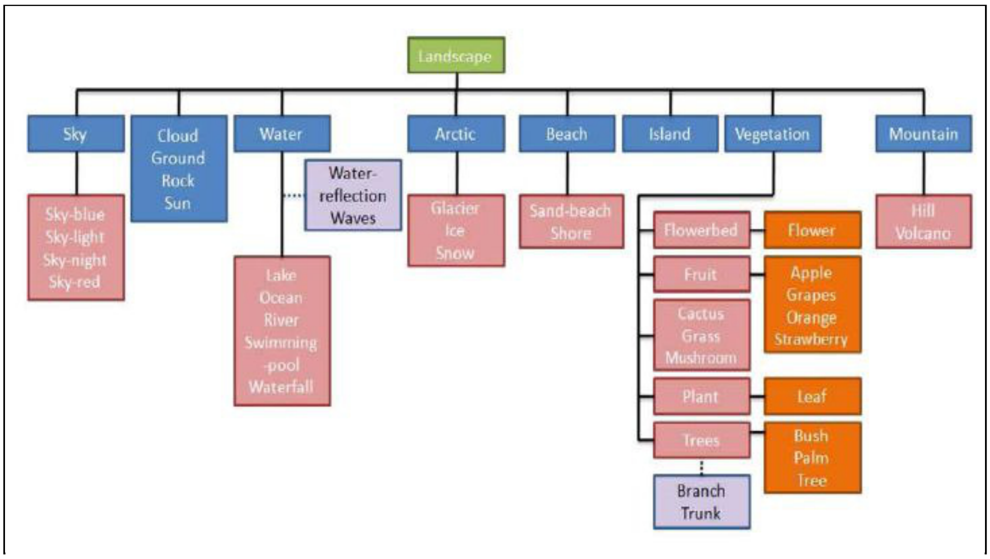


Fig. 2. The hierarchical structure of the Landscape-Nature branch.

respectively. The values determined after experiments for the Cross Media Relevance Model were α = 0.1 and β = 0.9.

V. Dataset

We have used for our experiments the segmented and annotated SAIAPR TC-12 [20][27] benchmark which is an extension of the IAPR TC-12 [21] collection for the evaluation of automatic image annotation methods and for studying their impact on multimedia information retrieval. IAPR TC-12 was used to evaluate content based image retrieval and multimedia image retrieval methods [28][29]. SAIAPR TC-12 benchmark contains the pictures from the IAPR TC-12 collection plus: segmentation masks and segmented images for the 20,000 pictures, region-level annotations according an annotation hierarchy, region-level annotations according an annotation hierarchy, spatial relationships information. Each image was manually segmented using a Matlab tool named Interactive Segmentation and Annotation Tool (ISATOOL). ISATOOL allows the interactive segmentation of objects by drawing points around the desired object, while splines are used to join the marked points, which also produces fairly accurate segmentation with much lower segmentation effort. Each region has associated a segmentation mask and a label from a predefined vocabulary of 275 labels. This vocabulary is organized according to a hierarchy of concepts having six main branches: Humans, Animals, Food, Landscape-Nature, Man-made and Other. In figure 2 it is presented the hierarchical structure of the Landscape-Nature branch.

For each pair of regions the following relationships have been calculated in every image: adjacent, disjoint, beside, X-aligned, above, below and Y-aligned. The following features have been extracted from each region: area, boundary/area, width and height of the region, average and standard deviation in x and y, convexity, average, standard deviation and skewness in two color spaces: RGB and CIE-Lab.

The dataset contains several folders of images, each folder having the structure presented in figure 3:

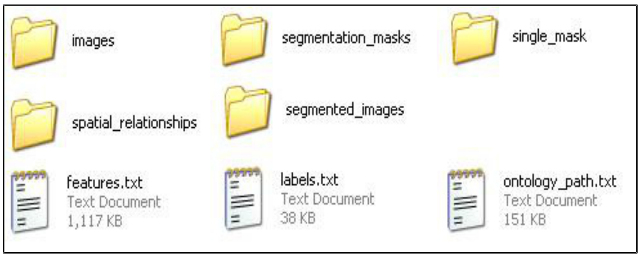

images
segmentation_masks
single_mask
spatial_relationships
segmented_images
features.txt
Text Document
1,117 KB
labels.txt
Text Document
38 KB
ontology_path.txt
Text Document
151 KB

Fig. 3. The structure of images' folder

where:

a) images folder contains the initial images that were manually segmented

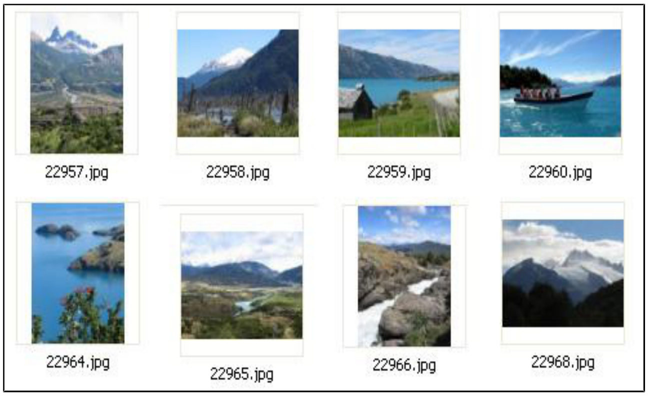


Fig. 4. Initial images

b) segmentation_masks folder contains files having the extension .mat (Matlab files). For each image's region a file is provided containing a segmentation mask which can be seen as a matrix with 0 and 1 values. A value of 1 in a matrix location means that the pixel having that position in the original image belongs to that region.

c) single_mask folder contains a single .mat file per image, representing the mask of the entire image.

d) spatial_relationships contains a file per image with information about the spatial relationships detected between each pair of regions

e) segmented_images folder contains manually segmented images having the regions shown with boundary

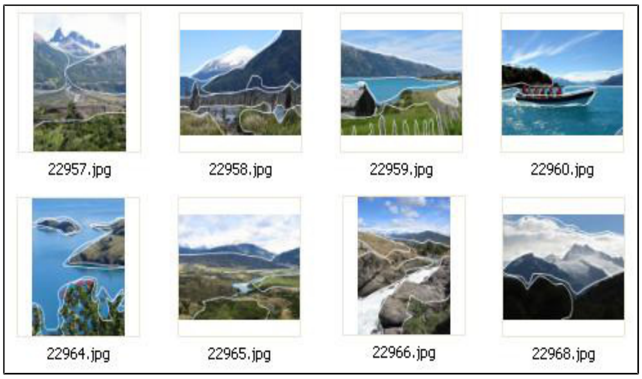
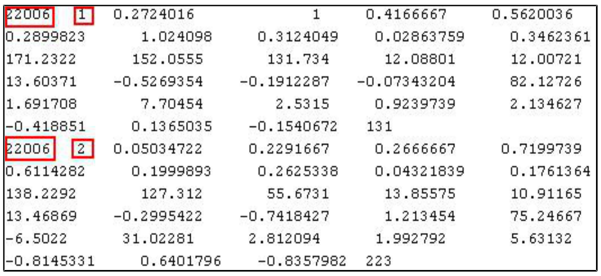


Fig. 5. Manually segmented images



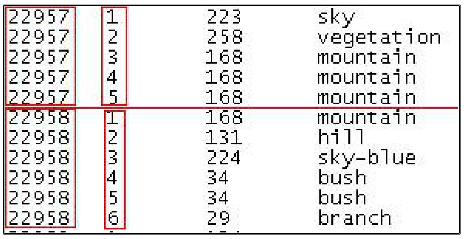
```
22006   1   0.2724016           1    0.4166667    0.5620036
0.2899823    1.024098    0.3124049    0.02863759    0.3462361
171.2322    152.0555    131.734    12.08801    12.00721
13.60371    -0.5269354    -0.1912287    -0.07343204    82.12726
1.691708    7.70454    2.5315    0.9239739    2.134627
-0.418851    0.1365035    -0.1540672    131
22006   2   0.05034722    0.2291667    0.2666667    0.7199739
0.6114282    0.1999893    0.2625338    0.04321839    0.1761364
138.2292    127.312    55.6731    13.85575    10.91165
13.46869    -0.2995422    -0.7418427    1.213454    75.24667
-6.5022    31.02281    2.812094    1.992792    5.63132
-0.8145331    0.6401796    -0.8357982    223
```

Fig. 6. Features' values for each region

f) features.txt contains the values of the extracted features from each region

g) 2206 identifies the picture (Figure 6 - 22006.jpg) , values 1 and 2 represent the index of each region and the rest of the values represent the values of the extracted features

h) labels.txt file contains the information needed to identify the words assigned to each image region, each word being indicated by his index. Using this information and the list of all words available in the wlist.txt file (being available for all folders at a higher level) having a pair (word index, word) on each line we can determine the words assigned to regions

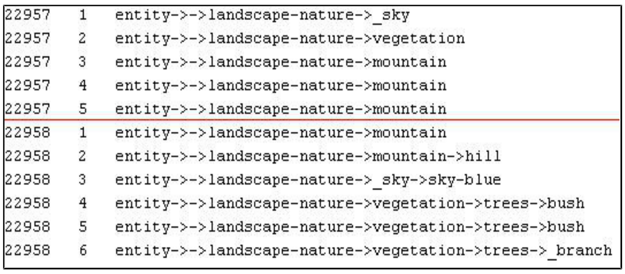


```
22957   1   223   sky
22957   2   258   vegetation
22957   3   168   mountain
22957   4   168   mountain
22957   5   168   mountain
22958   1   168   mountain
22958   2   131   hill
22958   3   224   sky-blue
22958   4   34    bush
22958   5   34    bush
22958   6   29    branch
```

Fig. 7. Words assigned to regions

i) where 22957 and 22958 represent images' identifiers, 1, 2…5 or 1,2 …6 represent the index of each region.

j) ontology_path.txt file contains the path in the ontology for each word associated to a region



```
22957   1   entity->->landscape-nature->_sky
22957   2   entity->->landscape-nature->vegetation
22957   3   entity->->landscape-nature->mountain
22957   4   entity->->landscape-nature->mountain
22957   5   entity->->landscape-nature->mountain
22958   1   entity->->landscape-nature->mountain
22958   2   entity->->landscape-nature->mountain->hill
22958   3   entity->->landscape-nature->_sky->sky-blue
22958   4   entity->->landscape-nature->vegetation->trees->bush
22958   5   entity->->landscape-nature->vegetation->trees->bush
22958   6   entity->->landscape-nature->vegetation->trees->_branch
```

Fig. 8. Ontology's paths assigned to regions.

VI. System's Architecture

System's architecture is presented in figure 9 and contains 6 modules:

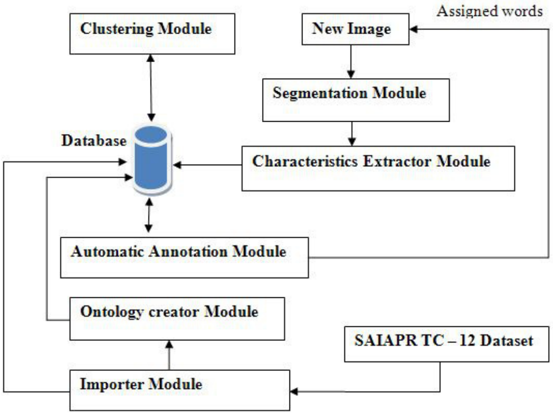


Fig.9. System's architecture

a) Importer module – this module is used to extract the existing information in the dataset. Having available segmentation's mask for each image's region this module detects the pixels that belong to that region. By parsing the content of the features.txt file the module extracts a list of feature vectors that are stored in the database. These feature vectors are clustered by the Clustering module for obtaining a list of blobs. The exiting information in the labels.txt and ontology_path.txt files about the words assigned to regions and the paths in the ontology is extracted and is made available to the Ontology creator module.

b) Ontology creator module - using the information provided by the Importer module and an original approach this module creates an ontology that is used for annotating new images. The existing ontology's paths are used to establish the hierarchical structure of the ontology. Each path is converted to several hierarchical relationships of parent-child type. The information contained in the spatial_relationships folder is used to generate several relationships in the ontology having spatial-relationship type. Each word is represented as a concept in the ontology having as unique identifier his index in the wlist.txt file. The ontology is represented as a Topic Map [30] using the XTM syntax [31]. In the bellow table are presented two ontology concepts (Mountain and Landscape) modeled as topics in the Topic Map and a hierarchical relationship between them modeled as an association:

Topics	<topic id= ‟168”> <instanceOf> <topicRef xlink:href="#semantic-class"/> </instanceOf> <baseName> <baseNameString>Mountain</baseNameString> </baseName> </topic>
	<topic id= ‟148”> <instanceOf> <topicRef xlink:href="#semantic-class"/> </instanceOf> <baseName> <baseNameString>Landscape</baseNameString> </baseName> </topic>
Association	<association id=”148-168”> <instanceOf> <topicRef xlink:href="#parent-child"/> </instanceOf> <member> <roleSpec> <topicRef xlink:href="#parent"/> </roleSpec> <topicRef xlink:href="#148"/> </member> <member> <roleSpec> <topicRef xlink:href="#child"/> </roleSpec> <topicRef xlink:href="#168"/> </member> </association>

Segmentation module – this module is using the segmentation algorithm described in Section 3 to obtain a list of regions from each new image. The segmentation algorithm is using some methods during the segmentation process:

SameVertexColor – used to determine the color of a hexagon

ExpandColorArea – used to determine the list of hexagons having the color of the hexagon used as a starting point and has O(n) as running time where n is the number of hexagons from a region with the same color.

ListRegions – used to obtain the list of regions and has O(n^2) as running time where n is the number of hexagons from the hexagonal network.

ContourRegions – used to obtain the contour of each region and has O(n) as running time where n is the number of hexagons from a region with the same color

Characteristics extractor module - this module is using the regions detected by the Segmentation module. For each segmented region it is computed a feature vector that contains visual information of the region such as area, boundary/area, width and height of the region, average and standard deviation in x and y, convexity, average, standard deviation. All feature vectors obtain are stored in the database in order to be accessible for other modules.

Clustering module - we have used K-means algorithm to quantize the feature vectors obtained from the training set and to generate blobs. After the quantization, each image in the training set was represented as a set of blobs identifiers. For each blob it is computed a median feature vector and a list of words that were assigned to the test images that have that blob in their representation.

Automatic annotation module - for each region belonging to a new image it is assigned the blob which is closest to it in the cluster space. The assigned blob has the minimum value of the Euclidian distance computed between the median feature vector of that blob and the feature vector of the region. In this way the new image will be represented by a set of blobs identifiers. Having the set of blobs and for each blob having a list of words we can determine a list of potential words that can be assigned to the image. What needs to be established is which words better describe the image content. This can be made using formulas (3) and (4) of the Cross Media Relevance Model. For each word it is computed the probability to be assigned to the image and after that the set of words having a probability greater than a threshold value will be used to annotate the image.

VII. Evaluation of the Annotation System

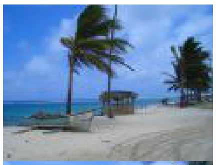
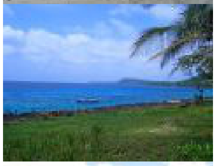
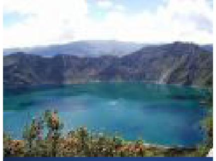
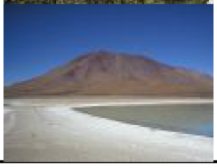
In order to evaluate the annotation system we have used a testing set of 400 images that were manually annotated and not included in the training set used for the CMRM model. This set was segmented using the original segmentation algorithm described above and a list of words having the joint probability greater than a threshold value was assigned to each image. Then the number of relevant words automatically assigned by the annotation system was compared against the number of relevant words manually assigned by computing a recall value. Using this approach for each image we have obtained a statistic evaluation having the structure presented in Table 1.

After computing the recall value for each image it was obtained a medium recall value equal to 0.73.

VIII. Conclusions and Future Work

In this paper we described a system that can be used for annotating natural images. The CMRM annotation model implemented by the system was proven to be very efficient by several studies. This model learns the joint probability of words and blobs based on a well know benchmark: SAIAPR TC-12. This benchmark contains a large-size image collection comprising diverse and realistic images, includes an annotation vocabulary having a hierarchical organization, well defined criteria for the objective segmentation and annotation of images. Because the quality of an image region and the running time of the segmentation process are two important factors for the annotation process we have used a segmentation algorithm based on a hexagonal structure which was proved to satisfy both requirements: a better quality and a smaller running time. Each new image was annotated with

Table 1. Statistic evaluation of the system

Index	Image	Relevant words automatically assigned (RWAA)	Words manually assigned (WMA)	Recall = RWAA/ WMA
0		sky-blue, sand-beach, ocean	sand-beach, ocean, boat, palm, hut, sky-blue	3/6 = 0.50
1		sky-blue, grass, ocean, cloud	grass, ocean, boat, cloud, sky-blue, branch	4/6 = 0.66
2		sky, mountain, lake	lake, vegetation, mountain, cloud, sky	3/5 = 0.60
3		mountain, sky-blue, sand-dessert	mountain, lake, sand-dessert, sky-blue	3/4 = 0.75

words taken from an ontology created starting from the information provided by the benchmark: the hierarchical organization of the vocabulary and the spatial relationships between regions. The ontology created in an original manner was represented using the Topic Map standard, each concept being modeled as a topic item and each relationship as an association having a specific type.

Further extensions of the system will include the two models of image retrieval provided by CMRM: Annotation-based Retrieval Model and Direct Retrieval Model.

Acknowledgment

The support of The National University Research Council under Grant CNCSIS IDEI 535 is gratefully acknowledged.

References

[1] Y. Mori, H. Takahashi, R.Oka: Image-to-word transformation based on dividing and vector quantizing images with words. In: MISRM'99 First Intl. Workshop on Multimedia Intelligent Storage and Retrieval Management (1999)

[2] P. Duygulu, K. Barnard, N. de Freitas, D. Forsyth: Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary. In Seventh European Conf. on Computer Vision, pp. 97–112 (2002)

[3] J. Jeon, V. Lavrenko, R. Manmatha: Automatic Image Annotation and Retrieval using Cross-Media Relevance Models. In: Proceedings of the 26th Intl. ACM SIGIR Conf., pp. 119–126 (2003)

[4] D. Blei, Michael, and M. I. Jordan. Modeling annotated data. To appear in the Proceedings of the 26th annual international ACM SIGIR conference

[5] J. Jeon and R. Manmatha, "Using maximum entropy for automatic image annotation." in CIVR, pp. 24–32, 2004

[6] J. Li, J. Wang,: Automatic linguistic indexing of pictures by a statistical modeling approach. IEEE Transactions on Pattern Analysis and Machine Intelligence 25 (2003)

[7] F. Monay and D. Gatica-Perez. Plsa-based image auto-annotation: constraining the latent space. In Proceedings of ACM International Conference on Multimedia (ACM MULTIMEDIA), pages 348–351, 2004.

[8] S. C. Deerwester, S. T. Dumais, T. K. Landauer, G. W. Furnas, and R. A. Harshman. Indexing by latent semantic analysis. Journal of the Society for Information Science, 41(6):391–407, 1990.

[9] T. Hofmann. Unsupervised learning by probabilistic latent semantic analysis. Machine Learning, 42(1-2):177–196, 2001.

[10] V. Lavrenko, R. Manmatha, and J. Jeon. A model for learning the semantics of pictures. In Proceedings of Advances in Neural Information Processing Systems (NIPS), 2004.

[11] S. L. Feng et al. Multiple bernoulli relevance models for image and video annotation. In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 1242–1245, 2004.

[12] J. Rong, J. Y. Chai, and L. Si. Effective automatic image annotation via a coherent language model and active learning. In Proceedings of ACM International Conference on Multimedia (ACM MULTIMEDIA), pages 892–899, 2004.

[13] D. Burdescu, M. Brezovan, E. Ganea, and L. Stanescu, "A New Method for Segmentation of Images Represented in a HSV Color Space", Lecture Notes in Computer Science, 5807, 606-617, 2009.

[14] J. Shi and J. Malik, "Normalized Cuts and Image Segmentation", IEEE Transactions on pattern analysis and machine intelligence, Vol. 22, No. 8, 2000.

[15] P. Felzenszwalb and D. Huttenlocher, "Efficient Graph-Based Image Segmentation", Intl J. Computer Vision, vol. 59, no. 2, 2004.

[16] D. Comaniciu and P. Meer, "Mean Shift: A Robust Approach toward Feature Space Analysis", IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 24, pp. 603-619, 2002.

[17] J. R. Smith, S. F. Chang.: "Tools and Techniques for Color Image Retrieval", Symposium on Electronic Imaging. In: Science and Technology - Storage & Retrieval for Image and Video Databases IV, volume 2670, San Jose, CA, February 1996. IS&T/SPIE. (1996)

[18] G. Mihai, A. Doringa, L. Stanescu, A Graphical Interface for Evaluating Three Graph-Based Image Segmentation, 2010,Proceedings of the International Multiconference on Computer Science and Information Technology pp. 735–740,2010

[19] 19. A. Iancu, B. Popescu, M. Brezovan , E. Ganea, "Region-based Measures for Evaluation of Color Image Segmentation", Proceedings of the International Multiconference on Computer Science and Information Technology pp. 717–722, 2010

[20] "Segmented and Annotated IAPR TC-12 dataset", http://imageclef.org/SIAPRdata

[21] "IAPR TC-12 Benchmark", http://imageclef.org/photodata

[22] V. Lavrenko and W. Croft. "Relevance-based language models". Proceedings of the 24th annual international ACM SIGIR conference, pages 120-127, 2001.

[23] V. Lavrenko, M. Choquette, and W. Croft. "Cross-lingual relevance models". Proceedings of the 25th annual international ACM SIGIR conference, pages 175-182, 2002.

[24] E. Ganea, M. Brezovan , "An Hypergraph Object Oriented Model for Image Segmentation and Annotation", Proceedings of the International Multiconference on Computer Science and Information Technology pp. 695–701, 2010

[25] J. Li, J.Z.Wang, "Real-time computerized annotation of pictures", IEEE transactions on pattern analysis and machine intelligence, Vol. 30, No. 6. (June 2008), pp. 985-1002

[26] R. Datta, D. Joshi, J. Li, J. Z. Wang, Image retrieval: ideas, influences, and trends of the new age, ACM Computing Surveys 40 (2) (2008) 1–60.

[27] H. J. Escalante, C. A. Hernández, J. A. Gonzalez, A. López-López, M. Montes, E. F. Morales, L. Enrique Sucar, L. Villaseñor and M. Grubinger, "The segmented and annotated IAPR TC-12 benchmark ",Computer Vision and Image Understanding, Volume 114, Issue 4, April 2010, Pages 419-428

[28] P. Clough, M. Grubinger, T. Deselaers, A. Hanbury, H. Müller, "Overview of the ImageCLEF 2006 photographic retrieval and object annotation tasks", Evaluation of Multilingual and Multimodal Information Retrieval – 7th Workshop of the CLEF, LNCS vol. 4730, Springer, Alicante, Spain, 2006 , pp. 579–594.

[29] M. Grubinger, P. Clough, A. Hanbury, H. Müller, "Overview of the ImageCLEF 2007 photographic retrieval task", Advances in Multilingual and Multimodal Information Retrieval – 8th Workshop of CLEF, LNCS vol. 5152, Springer, Budapest, Hungary, 2007, pp. 433–444.

[30] Topic Maps, http://www.topicmaps.org/

[31] XTM syntax, http://www.topicmaps.org/xtm/