Semantic Modelling of Unshaped Object: An Efficient Approach in Content Based Image Retrieval

Md. Shamsujjoha†, Md. Shakil Ahmed‡, Fahmida Hossain‡, Taskeed Jabid*
†‡*Department of Computer Science and Engineering, East West University, Dhaka-1212, Bangladesh
‡Department of Computer Science and Engineering, University of Dhaka, Dhaka-1000, Bangladesh
Email: †dishacse@yahoo.com, ‡Shakil0304003@yahoo.com, ‡fahmidacsedu@gmail.com, *taskeed@yahoo.com

***Abstract*— This paper presents an efficient image exploration scheme for the unshaped object using semantic modelling. The local regions of an image have been classified with respect to the frequency of occurrences. The semantic concept is evaluated using RGB histogram dissimilarity factor, overall dissimilarity factor and regional dissimilarity factor. The dissimilarities determine the local concept with accuracy up to 89.86% which is much higher than the existing techniques. The proposed algorithm also allows to ranks the unshaped objects according to their semantic similarity.**

Keywords— Content based image retrieval; Dissimilarity factors; Histogram; Semantic modeling; Unshaped object.

I. Introduction

Automatic object detection in the images is one of the central challenges in computer vision and pattern analysis [1]. The content based image retrieval (CBIR) output images which are visually similar to the user request. However, the user expectation or queries can be unpredictable. Thus, a common approach is to take a relevance feedback on the output and then get a rough idea of the search target. This is time consuming [2] and also laborious for the user [3]. Due to the significant amount of variation between the images of the same category, the object detection becomes much harder. On the other hand, changes in the viewpoint, scale, illumination, partial occlusions and multiple instances further complicate this detection problem [4]. The object detection in images and videos has found its application in surveillance systems and airport security [5], automatic driving and driver assistance systems in high-end cars [6], human-robot interaction and immersive [7], interactive entertainments [8], smart assistance for the senior citizens [9], military applications [10].

In these consequences, this paper presents an efficient semantic modeling for content base image retrieval of unshaped objects like sea, sky, sand, soil, grass, ice, rock etc. The semantic understanding of scenes is an important research challenge for the image and video retrieval community on its own. Researchers indicate the urgency of semantic modelling to gain access to the content of still images [11-14]. The existing techniques involved in organizing, indexing and retrieving digital images are too inefficient compare to the exponential growth of the data. Moreover, the semantic gap between the users understanding and computers representation of images hinders fast progress in modelling high-level semantic content for image browsing and retrieval. Early retrieval systems [15] have been based on the extraction of low level, often global pictorial features. Here, scene classified in the definition of the classes like indoor versus outdoor, waterfalls versus mountains with overall classification accuracy 71.7%.

Thus, our main objective is to reduce the semantic gap between the human and computer image representation to gain more accuracy. In this paper we develop an image representation method that is more intuitive for the user in addition to the local image description *i.e.*, a global image representation based on local information. The supervised semantic modeling has been used for this purpose as in the unsupervised or semi-supervised methods the extraction of semantics can be incidental and the annotation accuracies become undesirably low [16-17]. Recently, several methods have notice the addressing of global as well as local image annotation as shown in [15]. In general, the work [15] learns the correspondence between global annotations and images or the image regions. But the global annotations are more general than the pure region naming, and consequently a semantic correspondence between keywords and image regions does not necessarily exist and is not considered. This is especially true for the correspondence between category labels and category members [18-19]. The proposed semantic modelling allows us to rank unshaped object according to their semantic similarity. Based on ranking data, we learn the perceptually plausible distance measurement techniques that lead to a high correlation between the human and the automatic ranking. This result is especially valuable for content-based image retrieval where the goal is to present output in descending semantic similarity from the queries.

II. Proposed Semantic Modeling of Unshapped Object

Semantic model is a conceptual data model that describes the meaning of its instances. In other words, a semantic data model is an abstraction that defines how the stored symbols relate to the real world. This is the main reasons for uses of semantic model for the unshaped object. The proposed semantic image description divides the entire image analysis process into following five stages:

Figure 1. Learned semantic concepts

Stage 1: Learns RGB histogram from stored and classified images.

Stage 2: Identifies of local image regions from the classified images through dissimilarity factors. In order to be independent on the largely varying quality of an automatic segmentation, the local image regions are extracted on a regular grid of $n \times n$ regions.

Stage 3: Calculates the overall dissimilarity factor for each local regular grid.

Stage 4: Determines regional dissimilarity factor of each local regular grid. The regional dissimilarity factor depends on overall dissimilarity factor and its neighbor's overall dissimilarity factor. Then, category of each local regular grid is being calculated with respect to lowest regional dissimilarity factor. If the lowest regional dissimilarity factor of a local regular grid crossed a predefined threshold, then we consider its category as unknown.

Stage 5: Finally, local (region based) concept is combined for a global representation. For each global semantic concept, its frequency of occurrence is determined, which represent the probability to present a category globally.

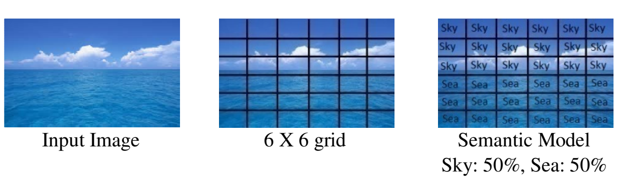

Figure 2. Image representation through semantic modeling

Fig. 1 shows the semantic concept of the local image regions. As can be seen the semantic content of the local image regions is much simpler than that of full image. Thus, it makes the acquisition of ground required for training and is much easier for testing. Since, the local semantic concepts correspond to real-world concepts, the method can also be used for the discretional image retrieval. Below we detailed the working procedure of the proposed algorithm with examples.

A. Descriptiong of the Proposed Concept Classifiers

The proposed concept classifier is the semantic classification of local image region that extracts the image regions into $n \times n$ a regular grid as shown in Fig. 2. In this figure, each local grid corresponds to the dissimilarity from the learned classified images. Here, the RGB histogram dissimilarity factor D is defined as follows:

$$D = \sum_{i=0}^{255} (|(R_i - R_{Li})| + |(G_i - G_{Li})| + |(B_i - B_{Li})|) \ldots (1)$$

Where, R_i, G_i and B_i are the percentage of i^{th} value in Red, Green and Blue color of local regular grid, respectively. On the other hand, R_{Li}, G_{Li} *and* B_{Li} are the percentage of i^{th} value of learned classified images in similar colors, respectively.

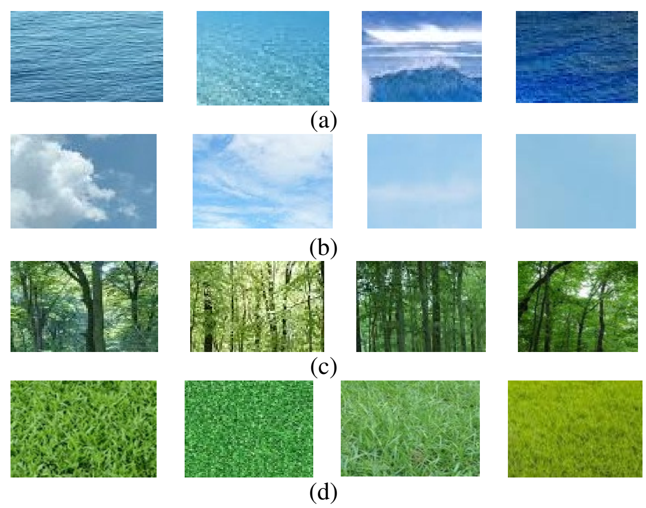

Figure 3. Training images (a) Sea (b) Sky (c) Forest (d) Grass

In addition, we calculate the overall dissimilarity factor with respect to semantic concept. Overall dissimilarity factor defined the dissimilarity among an image block and all trained image block of a category. Let, N the number of trained image block of a single category. Also let the category's RGB histogram dissimilarity factor be D_1, D_2, D_3 …, D_n(here dissimilar factor are sorted in ascending order). Then the overall dissimilarity factor OD for a single category of an image block is defined as follows:

$$OD = D_1 - \sum_{i=2}^{n} (3/(N + (D_i - D_1) \times N)) \ldots (2)$$

Searching image   A region of the search image

Figure 4. Local Concept

For example, considering an image blocks' dissimilarity factors for the grass as 0.45, 2.56, 3.45, 3.55, 4.56, … … …; and the dissimilarity factors for the forest as 0.56, 0.65, 0.98, 1.56, 1.66,… … …; In this example the lowest image block dissimilarity factor is smaller for the category grass (0.45<0.56) but the other dissimilarity factors for the category grass are much higher (*e.g.*, 2.56>0.65 or 3.45>0.98 or 3.55>1.56 etc). Therefore, others dissimilarity factors pushes its probability to become category forest rather than grass. To resolve this ambiguity, we determine overall dissimilarity factor. For this example, the overall dissimilarity factor for grass and forest are 0.39 and 0.32, respectively. So, the probability to become forest is larger than to become grass.

Finally, we calculate the regional dissimilarity factor of each image block. Regional dissimilarity factor do adjustment between its overall dissimilarity factor and its neighbors overall dissimilarity factor. Let, an image block overall dissimilarity factors be, sea = 0.82551, sky=0.797003 and ice=0.86. The summation of all its neighborhood's overall dissimilarity factors be sea=7.36, sky=16.61 and ice=18.92. Then, according to our proposed algorithm for this region, the sky category gets the minimum overall dissimilarity factor, but according to its neighbors probability it become sea as this much higher.

Original Image   After regional classification

Figure 5. Regional classification

The regional dissimilarity factor RD for any category is defined as follows:

$$RD = OD - \left(\frac{16}{16 + 16 \times \sum_{neighbor} OD}\right) \ldots (3)$$

According to Eq. 3, the regional dissimilarity factor for the example we are continuing are, sea=0.7055, sky=0.74 and ice=0.8098. Thus, the probability to become sea for the region is much higher. The category of a region is selected by the lowest regional dissimilarity factor, but if the lowest regional dissimilarity factor is larger than a certain threshold, we consider that the category of the region is unknown.

The regional category decides about the categories that the image contains. Let us consider this as global category. Also let, we want to find the probability of category X in an image. Then, if A is the number of regional category found in the image for X and B is the number of regional categories that are known in the image. Then, the probability that the image contain the category X is,

$$P(X) = \left(\frac{A}{B}\right) \times 100\% \ldots (4)$$

According to Eq. 4, the probability of rocks shown in Fig. 5 is, (36/36)*100% = 100%.

III. Performacne Analysis Of The Proposed Scheme

This section presents the overall simulation results of the proposed semantic algorithm. The semantic modeling for content-based image retrieval are often found in natural scenes such as sky, water, grass, trunks, foliage, field, rocks, flower, send etc., with maximum overall classification accuracy 71.7% [15]. The proposed semantic modeling of unshaped objects results an overall accuracy of 89.86% which is much higher than the existing method. The existing approaches extract local image regions on a regular grid of 10x10 regions and used HIS color histogram with value of hue as 36 bins, saturation as 32 bins and intensity as 16 bins. However, the proposed scheme used RGB histogram and extracted local image regions on a regular grid of 6x6 regions because its accuracy is maximum as shown in row 3 of Table I.

TABLE I. Overall Accuracy for Different Grid Size

Grid Size	No of Experiment Images	Accuracy
4 x 4	2000	50.43%
5 x 5	2000	62.37%
6 x 6	2000	89.86%
7 x 7	2000	85.34%
8 x 8	2000	81.43%
9 x 9	2000	80.23%
10 x 10	2000	78.96%

TABLE II. Execution of the Proposed Algorithm Accuracy (A) Training Images (B) Unshapped Object

(A)

Training Images	No of Experiment Images	Accuracy
10	3000	30.35%
20	3000	56.54%
30	3000	65.72%
40	3000	78.82%
50	3000	89.86%
60	3000	89.85%

(B)

Object	No of Experiment Images	Accuracy
1	500	100%
5	500	97%
10	1000	93.56%
15	1500	90.19%
20	2000	89.86%

Uses of overall dissimilarity factor and regional dissimilarity factor increased the classification accuracy. The proposed test simulation contains 20 unshaped objects and for each objects, we consider 50 learned classified images. From Table 2(a), we find that accuracy increases with number of training image of a category. But, no. of training image greater than 50 results very small increase in accuracy and thus we choose 50 training image of each category for test simulation. Table 2(b) represents that the increases in the number of unshaped objects leads to little decrease in the accuracy.

Tables III, to IV and Fig. 6 summarized the test simulation results of the proposed semantic model. As we discussed earlier, the simulation will divided the image in 36 regions. Test simulation for Fig. 6(a), 19 regions is detected as forest and 17 regions are detected as grass among the 36 regions and numerically, 52.78% is forest and 47.22% is grass. Here, initially regions dissimilarity factor from the learned classified images is calculated. To increase the overall retrieval accuracy, the overall dissimilarity factor and regional dissimilarity factor is calculated. For example, in Fig. 6(b), if the minimum among dissimilarity factors is taken, the result become, sea=62.06%, sky=24.13%, ice=10.34%, rocks=3.44%. But if we take minimum of the overall dissimilarity factor, then result becomes: sea=72.41%, sky=20.68% ice=6.89%, but if we take regional dissimilarity, the result is, sea: 79.30%, sky: 20.68%. Table 6 presents overall performance of the proposed method with respect to existing one [15]. Form this table we find that the proposed method performs much better than the existing design.

TABLE III. ACCURACY ON UNSHAPPED IMAGE CATEGORY ACCURACY

Object Name	No of experiment image	Correct region count (6 x 6)	Accuracy Percentage
Sea	201	6704	92.65%
Sky	197	6391	90.12%
Forest	213	6966	90.85%
Grass	231	7482	89.97%
Apple Garden	167	5243	87.21%
Dark Sky	187	6080	90.31%
Fire	205	6519	88.33%
Ice	162	5176	88.75%
Orange Garden	197	6201	87.44%
Red Rose Garden	175	5554	88.16%
Rice	169	5386	88.53%
Rice Plant	209	6714	89.23%
Rock	211	6834	89.97%
Sand	189	6092	89.54%
Soil	203	6571	89.92%
Star Sky	156	4997	88.98%
Sunflower Garden	149	4804	89.56%
Sunshine	162	5187	88.94%

TABLE IV. ACCURACY ON DIFFERENT COLOR SYSTEM

Object	No of Experiment Image	Using HIS color histogram	Using gray scale color histogram	Using RGB color histogram
5	500	88.09%	70.54%	97%
10	1000	85.11%	66.71%	93.56%
15	1500	82.79%	61.29%	90.19%
20	2000	80.21%	58.62%	89.86%

TABLE V. ACCURACY OF MINIMUM OF THE DISSIMILAR FACTOR, OVERALL DISSIMILAR FACTOR AND REGIONAL DISSIMILAR FACTOR

Name	No of image	Using minimum of the dissimilar factor	Using overall dissimilar factor	Using regional dissimilar factor
Sea	201	72.43%	84.19%	92.65%
Sky	197	73.42%	83.94%	90.12%
Forest	213	71.56%	84.25%	90.85%
Grass	231	71.92%	81.39%	89.97%
Apple Garden	167	71.34%	79.16%	87.21%
Dark Sky	187	72.93%	80.61%	90.31%
Fire	205	69.45%	79.74%	88.33%
Ice	162	68.97%	79.13%	88.75%
Orange Garden	197	68.34%	78.49%	87.44%
Red Rose Garden	175	68.61%	80.68%	88.16%
Rice	169	69.58%	79.80%	88.53%
Rice Plant	209	69.16%	80.06%	89.23%
Rock	211	72.32%	81.18%	89.97%
Sand	189	71.19%	80.95%	89.54%
Soil	203	71.21%	80.21%	89.92%
Star Sky	156	69.83%	79.92%	88.98%
Sunflower Garden	149	71.17%	80.55%	89.56%
Sunshine	162	69.97%	78.93%	88.94%
Tomato Garden	165	69.47%	79.29%	89.23%
Tuberose Garden	154	69.62%	78.86%	88.67%

TABLE VI. COMPARATIVE STUDY OF PROPOSED METHOD WITH EXISTING SCHEME [15]

Object Name	No. of Images	Existing Design [15]	The Proposed Method
Sky	3000	72.13%	90.12%
Grass	3000	69.42%	89.97%
Rocks	3000	71.92%	89.97%
Sand	3000	72.56%	89.54%
Forest	3000	68.67%	90.85%

IV. Conclusion

The paper describes an efficient approach for content based image retrieval using local region based on semantic modeling. The proposed semantic modeling of unshaped object reduces the semantic gap between the image understanding of humans and computers. The proposed method shows how effectively local semantic content can be utilized for image categorization. In the experiment, we achieved 89.86% classification accuracy which higher than other state-of-the-art methods [15]. This experimental data further verified the suitability of the proposed method compare to other existing approaches. An interesting future work can be enhancement of this idea for the shaped object, specifically to detect the dissimilarities among the edges of the different shapes and then the classification categories of objects can be determined accordingly.

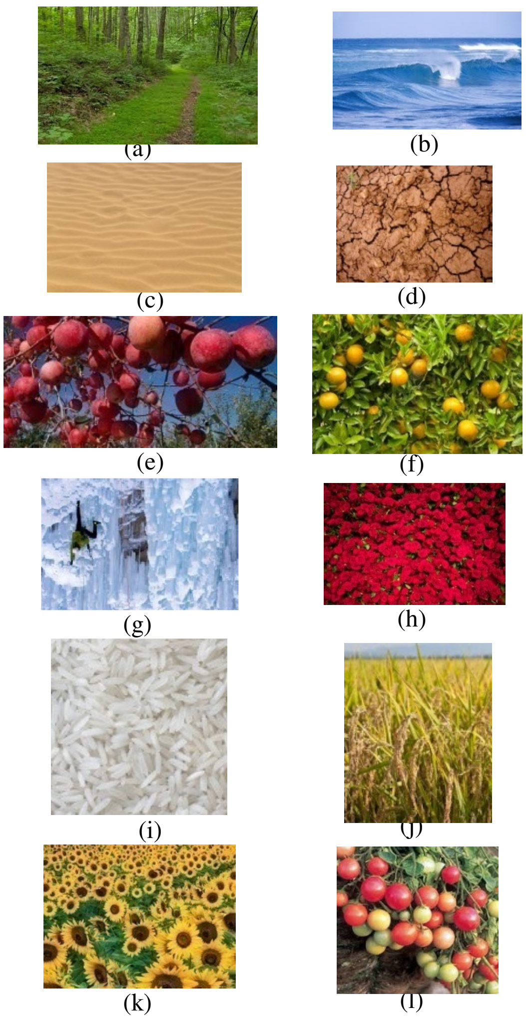

Figure 6. (a) Forest: 52.78%(b) Sea: 62.06% (c) Sand: 100% (d) Soil: 94.44% (e) Apple garden: 80.56% (f) Orange garden: 80.56% (g) Ice: 69.44% (h) Red rose garden: 100% (i) Rice: 100% (j) Rice plant: 63.89% (k) Sunflower garden: 91.67% (l) Tomato garden: 61.11%

References

[1] Agarwal, S., Awan, A., Roth, D., "Learning to detect objects in images via a sparse, part-based representation," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol.26, no.11, pp.1475,1490, 2004.

[2] R. W. White, I. Ruthven and J. M. Jose, "A study of factors affecting the utility of implicit relevance feedback.," in 28th annual international ACM SIGIR conference on Research and development in information retrieval (SIGIR), New York, NY, USA, 2005.

[3] S. Feng, R. Manmatha and V. Lavrenko, "Multiple bernoulli relevance models for image and video annotation," in Conference on Image and Video Retrieval (CIVR), Dublin, Ireland, 2004.

[4] Shahbaz Khan, F., Anwer R.M., van de Weijer J., Bagdanov A.D., Vanrell, M., Lopez, A.M., "Color attributes for object detection," Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on , vol., no., pp.3306,3313, 16-21 June 2012.

[5] Ziyan Wu; Radke, R.J., "Real-time airport security checkpoint surveillance using a camera network," Computer Vision and Pattern Recognition Workshops (CVPRW), 2011 IEEE Computer Society Conference on , vol., no., pp.25,32, 20-25 June 2011

[6] Sayanan Sivaraman and Mohan M. Trivedi, Towards Cooperative, "Predictive Driver Assistance", Proceedings of the 16th International IEEE Annual Conference on Intelligent Transportation Systems (ITSC 2013), The Hague, The Netherlands, October 6-9, 2013.

[7] Sandygulova, A.; Campbell, A.G.; Dragone, M.; O'Hare, G. M P, "Immersive human-robot interaction," Human-Robot Interaction (HRI), 7th ACM/IEEE International Conference on, pp.227,228, 2012.

[8] Jaeyong Chung; Jin Ryong Kim; Kwanghyun Shim, "Vision Based Motion Tracking System for Interactive Entertainment Applications," TENCON 2005 2005 IEEE Region 10 , vol., no., pp.1,6, 21-24 Nov. 2005

[9] Fong, A. C M; Fong, B.; Li, C. K., "Assistive independent living for senior citizens with special cognitive needs," Consumer Electronics (GCCE), 2012 IEEE 1st Global Conference on , vol., no., pp.448,449, 2-5 Oct. 2012

[10] Jianxin Wu; Nini Liu; Geyer, C.; Rehg, J.M., " C4 : A Real-Time Object Detection Framework," Image Processing, IEEE Transactions on , vol.22, no.10, pp.4096,4107, Oct. 2013.

[11] Ruofei Zhang; Zhongfei Zhang; Zhongyuan Qin, "Semantic repository modeling in image database," ICME '04, vol.3, no., pp.2079,2082 Vol.3, 27-30 June 2004.

[12] Passino, G.; Patras, I.; Izquierdo, E., "Pyramidal Model for Image Semantic Segmentation," Pattern Recognition (ICPR), 2010 20th International Conference on , vol., no., pp.1554,1557, 23-26 Aug. 2010.

[13] Xu Zhiping; Shiyong Zhang, "Artificial Visual Cortical Responding Model in Image Semantic Processing," Computer Systems and Applications, 2007. AICCSA '07. IEEE/ACS International Conference on , vol., no., pp.722,725, 13-16 May 2007.

[14] V. Lavrenko and R. Manmatha and J. Jeon,"A model for learning the semantics of pictures", in NIPS, MIT Press, 2003.

[15] Julia Vogel and Bernt Schiele, "Semantic Modeling of Natural Scenes for Content-Based Image Retrieval", Int. J. Comput. Vision 72, 2 pp-133-157, 2007).

[16] Yuxin Chen; Sampathkumar, H.; Bo Luo; Xue-wen Chen, "Bridging the Semantic Gap in Vertical Image Search by Integrating Text and Visual Features," *Knowledge and Data Engineering, IEEE Transactions on* , vol.25, no.10, pp.2257,2270, Oct. 2013.

[17] Eugene Agichtein, Eric Brill, and Susan Dumais, "Improving web search ranking by incorporating user behavior information," In *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval,* pp 19-26, 2006.

[18] J. Jain and C. Singh., "Ad you like it: advertisement sourcing and selection technique across multiple heterogeneous applications.," *SIGSOFT Softw. Eng. Notes,* vol. 37, no. 4, pp. 1-6, 2012.

[19] Lei Wu, Mingjing Li, Zhiwei Li, Wei-Ying Ma, and Nenghai Yu., "Visual language modeling for image classification," In *Proceedings of the international workshop on Workshop on multimedia information retrieval* pp 115-124, 2007.