IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 18, Issue 1, Ver. II (Jan – Feb. 2016), PP 43-60
www.iosrjournals.org

A Content-Based Image Retrieval Semantic Model for Shaped and Unshaped Objects

Md. Sahmsujjoha[1],Touhid Bhuiyan[2]
[1]*Department of Computer Science and Engineering, East West University, Dhaka, Bangladesh.*
[2]*Department of Software Engineering, Daffodil International University, Dhaka, Bangladesh.*

Abstract: *This paper presents an efficient content based image retrieval scheme for both the shaped and unshaped objects. The local regions of an unshaped image have been classified with respect to the frequency of occurrence. Then the semantic concept is evaluated throughRGB histogram dissimilarity factor, overall dissimilarity factor and regional dissimilarity factor. These dissimilarities cooperativelydetermine the local concept for theunshaped object. In addition, the semantic concept for shaped objects is measured through the normalized color findings, synchronized edge detection, small unnecessary particleremotion, and shape similarity checking. All these measurements mutually rank the shaped objects according to their probability of occurrences. In addition, several algorithms and theoretical explanations of the proposed semantic models have been presented. The corresponding examples and simulations prove that the proposed methods work accurately. The comparative results show that the proposed models have significantly better scalability than the existing approaches.*

Keywords:*Content Based Image Retrieval, Dissimilarity, Normalization, Semantic Modeling, Shaped and Unshaped Object.*

I. Introduction

Automatic object detection in the images is one of the central challenges in computer vision, pattern analysis and Content Based Image Retrieval (CBIR) [1]. Generally CBIR output images are visually similar to the user request. However, the user expectation or queries can be unpredictable. Thus a common approach is to take the relevant 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]. Dueto 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 thedetection problem[4]. In these consequences, this paper presents an efficient semantic model for the content base image retrieval. The proposed method classifies all objects according to the following two categories:

i. **Shape object:** No fixed shape *e.g.*, sea, sky, sand, soil, grass, ice etc.
ii. **Unshaped object:** Fixed shape *e.g.*, tiger, lion, dog, chicken, mango, plane 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 as well [5-8]. The existing techniques involved in organizing, indexing and retrieving digital images are too inefficient compared 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 both in browsing and retrieval. Thus, the main objective of the proposed model is to reduce the semantic gap between the human and computer image representation. The proposed image representation method is more intuitive for the user in addition to the local image description *i.e.*, a global image representation based on local information. Thus, the proposed method can be used in surveillance systems and airport security [9], automatic driving and driver assistance systems in high-end cars [10], human-robot interaction and immersive [11], interactive entertainments [12], smart assistance for the senior citizens [13], military applications [14] etc. Recently, several methods noticed the global as well as local image annotation, which learns the correspondence between global annotations and images or the image regions [15]. But the global annotations are more general than the pure region naming. 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. Oliva*et al.* present a local information based model for scene classification by organizing images along three semantic axes [16]. The semantic axes have been determined through psychophysical experiments. Mojsilovic*et al.* also used psychophysical experiments to obtain a set of semantic categories relevant to the humans as well as verbal descriptions [17]. In [18], a semantic model in addition to low-level features to increase indoor-outdoor classification performance has been proposed. A framework to describe natural scenes through multiple labels

is proposed in [19]. But the region labels are not combined to a global image representation, whichcan further be employed to content-based image retrieval. This problem is resolved in the proposed model by attachinga set to the semantically meaningful labels in local image regions for the subsequent retrieval step.The supervised semantic modeling has been used for this purpose as in the unsupervised or semi-supervised methods, extraction of semantics can be incidental and the annotation accuracies become undesirably low [20, 21]. The proposed supervised semantic model ranks the unshaped object according to their semantic similarity. Based on this ranking, we evaluate the perceptually plausible distance,whichresults a high correlation between the human and the automatic ranking. The result is especially valuable for CBIR techniques, where the output is presented in descending semantic similarity based on user queries.

Most state-of-the-art approaches for the shaped object ignoringcolor information and relying on intensity-based features due to the variations caused by the changes of illumination, compression, shadows and highlights [22-25]. These variations make the task of robust color description more difficult but color yields excellent result in combination with shape features in contrast to object detection. Hence the proposed semantic model used normalize color finding factor for the shaped object in addition to the edge detection, edge synchronization, small particle removal, histogram and shape similarity percentage, whichultimately develop an image representation method that is more intuitive for the user to CBIR for both the shaped and unshaped objects.

II. Basic Definition And Literature Review

This section presents the necessary background, which is required to understand the proposed work. The section starts with the introduction to the existing semantic models. Then we briefly present different digital image representation methods and corresponding computational requirements. The section ends with a discussion on the biometrical importance of the CBIR.

1.1 Existing Content-Based Image Retrieval Techniques

Today's computational world creates huge amount of digital images in every moment[26]. A large portion of these collections consists of transformed analogue photographs, diagrams, drawing, paintings, prints etc. Generally, object retrieval from these collections is quite difficult because of innumerable shape, color, size, brightness, sharpness. Thus, the regaining task is carried by the keyword indexing [27]. However, digital images databases open the door for the CBIR as well. Keywordbased search engines indicate that the media type must be images, which can be done in many ways, *e.g.*, browsingin database, fixingimage in terms of keywordsand features [28]. Yet another way is to provide an image or sketch from which features of the same type must be extracted [29, 30]. Several classes of features can be used to specify queries*e.g.*,color, textures, shape, spatial layout, faces etc. Color features are often easily obtained from the pixel intensities *i.e.*, color histograms over the entire image, or fixed sub-image, or segmented region [31]. Below we discuss the few popular CBIR schemes.

i. **Alta Vista Photofinder**: In alta vista search engine, CBIR similarity is defined with respect to dominant colors, shapes and textures. The keyword(s) are used to retrieve images tagged with it. If a retrieved image is shown, the link gives images that are visually similar to the selected image but the relative weights of the features is not shown [32].
ii. **Blobworld:** Blobworld used color, texture, location, shape and background for querying. The color is described using histogram of 218 bins of coordinates. Texture is represented by mean contrast and anisotropy over the region. Shape is represented by (approximate) area, eccentricity, and orientation. Here, user selects a category. The category limits the search space. In an initial image, the user selects a region. Next the user indicates the importance of the regions color, texture, and shape. More than one region can be used for querying [33].
iii. **C-bird:**C-bird is one of the CBIR methodsthat used digital libraries. Whenever an image is collected, a feature description and a layout description are computed. The feature description is a set of four vectors*i.e.*,color vector, most frequency color vector, orientation vectorand chromaticity vector. A 512-bin RGB histogram is stored in the color vector [34].
iv. **Excalibur Visual RetrievalWare:** Visual RetrievalWare toolkit is used to manipulate digital image files and their visual content. The toolkit contains C++ and Java API for image processing, feature extraction, indexing and content-based retrieval. It also includes sample programs, which might be used directly or can serve as templates for building more complex applications. One of these sample programs is the CST (Color, Shape, and Texture) demo. The CST demo allows queries by example based on HSV color histograms, relative orientation, curvature and contrast of lines in the image, and texture attributes, that measure the flow and roughness in the image [35].

v. **ImageRETRO:** ImageRETROinitially presents several images to the user. Selecting one of these images, the user actually selects a part of the database, which is then clustered, and another set of representative images is shown to the user. By this successive filtering, the user obtains a small set of images, which can be browsed manually [35, 36].

1.2 Existing Semantic Image Model

Semantic model includes semantic information *i.e.*, it describes the meaning of its object. Generally, semantic model for CBIR describes the contents of an image by a set of keywords [37]. Major requirements in semantic-based image retrieval are as follows:

i. The semantic gap in image representation between the human and the machine should as minimum as possible. The image representation should be more intuitive for the user [15]. Therefore, a vocabulary-supported access to images that replaces the common query-by-example paradigm with a query-by-keyword paradigm is needed.
ii. General image segmentation algorithms lead to both over and under segmentation of semantically contiguous regions. Thus, automatic image segmentation should be avoided and be substituted by a regular subdivision of the images [38].
iii. In addition to the local image description, a global image representation (based on local information) is needed. The global representation should allow a global comparison [39]. Moreover, the model should guide the system about the human perception of natural scenes.
iv. The image representation has to be evaluated quantitatively, especially with respect to its semantic representativeness. This assessed the semantic applicability, robustness, strengths, and weaknesses of the image representation through clearly defined and quantifiable tasks.
v. The last requirement is closely connected with the question of whether to employ supervised or unsupervised learning methods. The drawback of unsupervised or semi-supervised methods is that extraction of semantics can be incidental. Also, the annotation accuracies can become undesirably low [20, 21] in word region co-occurrences as describe earlier in Sec. 1.

First two requirements results better scalability, whereas the remaining requirements produces better accuracy. As shown in the Secs. 3 and 4, the proposed semantic model for CBIR tries to cover the above requirements as much as possible in its low level.

1.3 Digital Image Representation

Digital image representation means numeric representation of an image. It may be of vector or raster type.In other words, a digital image is considered as a matrix, whose row and column indices identify a point and the corresponding matrix element value identifies the gray levels at that point. The elements of digital array are called image elements, picture elements, pixels to be specific. The basic nature *o*fa digital image can be characterized by two components, first, the amount of source light incident and the amount of light reflected by the object [1, 3, 15, 34]. There are different representations of digital images such as RGB, Binary, and Histogram etc. A brief discussion about these representations is given below.

i. **RGB Image:** The RGB image has three channels Red, Green, and Blue. RGB channels roughly follow the color receptors in the human eye, and are used in computer displays and image scanners. If the RGB image is 24-bit, each channel has 8 bits. In other words, the image is composed of three grayscale images, where each grayscale image can store discrete pixels with conventional brightness intensities between 0 and 255 [24, 31].
ii. **Binary Image:** A binary image is a digital image that has only two possible values for each pixel namely black and white. Binary images often arise in digital image processing as the masks or as the result of certain operations such as segmentation, threshold, and dithering. Binary imageis used for noise removal [24].
iii. **Histogram:** A histogram is used to graphically summarize and display the distribution of a process image data set. A histogram can be constructed by segmenting the range of the data into equal sized bins generally known as segments, groups or classes [30].
iv. **Edge Detection:** Edges in images are areas,where contrasts jump in intensity occurs (from one pixel to the next). Edge detection in CBIR significantly reduces the amount of effort required while preserving the important structural properties[18, 20]. There are several edge detection mechanisms. These can be categorized into gradient based approaches and Laplacian based approaches [39].

1.4 Biometrical Importance of CBIR

Generally, the term *biometric*isused to describe a measurable physiological and behavioral characteristic. A biometric system provides an automated method of recognizing based on the biometric characteristics. The operation of a biometric based CBIR system can be described by a three-step process. The

first step in this process involves an observation or collection of the biometricdata. The second step converts and describes the observed data using a digital representation. In the third step, the newly converted data is compared with one or more previously generated data stored in a database. The result of this comparison is a "match" or a "non-match" and is used for the corresponding actions.A threshold determines the degree of similarity required to result in a match declaration. The acceptance or rejection ofbiometric data is dependent on the match score falling above or below the threshold. The threshold is adjustable so that the biometric system can be more or less strict, depending on the requirements of any given biometric application.Some biometric systems employ liveness detection. CBIRis very important in biometrical identification. If we do not retrieve the object from an image then, we cannot apply the individual's biometric characteristics in it [40].

III. Proposed Semantic Modeling fortheUnshaped Object

This section presents the proposed semantic model for the unshaped object. In the next section, we enhanced this work to the design of anefficient and cooperative semantic model for both the shaped and unshaped object.

Proposed Model

As discussed earlier, semantic 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 analysis process in the following five stages:

i. Stage one learns RGB histogram from stored and classified images.
ii. Stage two 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.
iii. Stage three calculates the overall dissimilarity factor for each local regular grid.
iv. Stage four 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.
v. Finally, stage five combinedregion based concept 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 1. Learned Semantic Concepts

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.

1.5 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:

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

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.

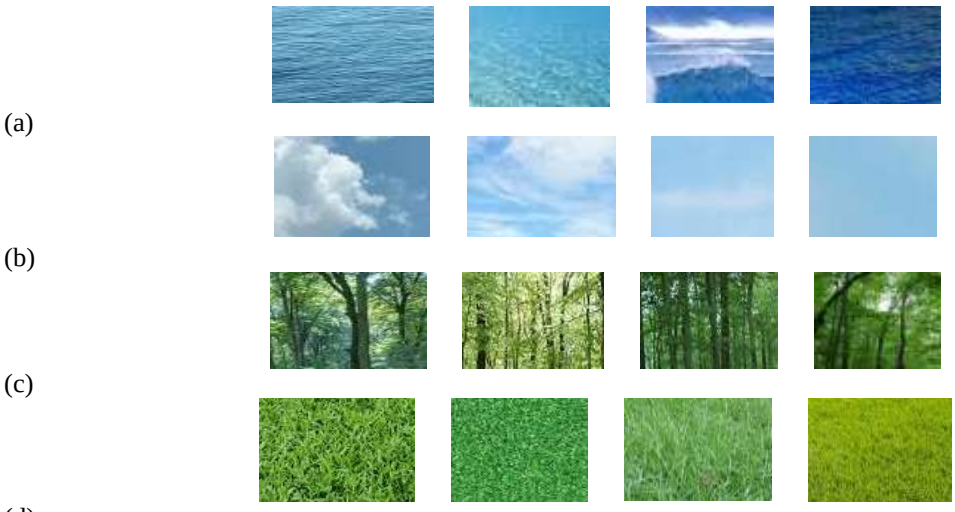


Figure 3. Training Images (a) Sea (b) Sky (c) Forest (d) Grass

Searching image

A local region of the search image

Figure 4. Local Concept

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[2]. 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 O*D* for a single category of an image block is defined as follows:

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

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 the grass.

Finally, we calculate the regional dissimilarity factor of each image block. Regional dissimilarity factors adjustments between its overall dissimilarity factor and its neighbors overall dissimilarity factor. Let, an image block's 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 becomes sea as this much higher.

[2]Sample training images and local concept are shown in Figs. 3 and 4, respectively

Original Image

After regional classification

Figure 5. Regional Classification

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

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

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 contains the category X is,

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

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

1.6 Performance Analysis of the Proposed Unshaped Model

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, sand 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. The proposed scheme used RGB histogram and extracted local image regions on a regular grid of 6x6 (total of 36 bins) regions as its accuracy is maximum, which is shown in Table I (row 3).

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 images of a category. We choose 50 training image of each category for test simulation as training image greater than 50 results very small increase in accuracy. Table 2(b) represents that the increases in the number of unshaped objects leads to little decrease in the accuracy. Tables III, 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 becomes, 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%

(a)

(b)

(c)

(d)

(e)

(f)

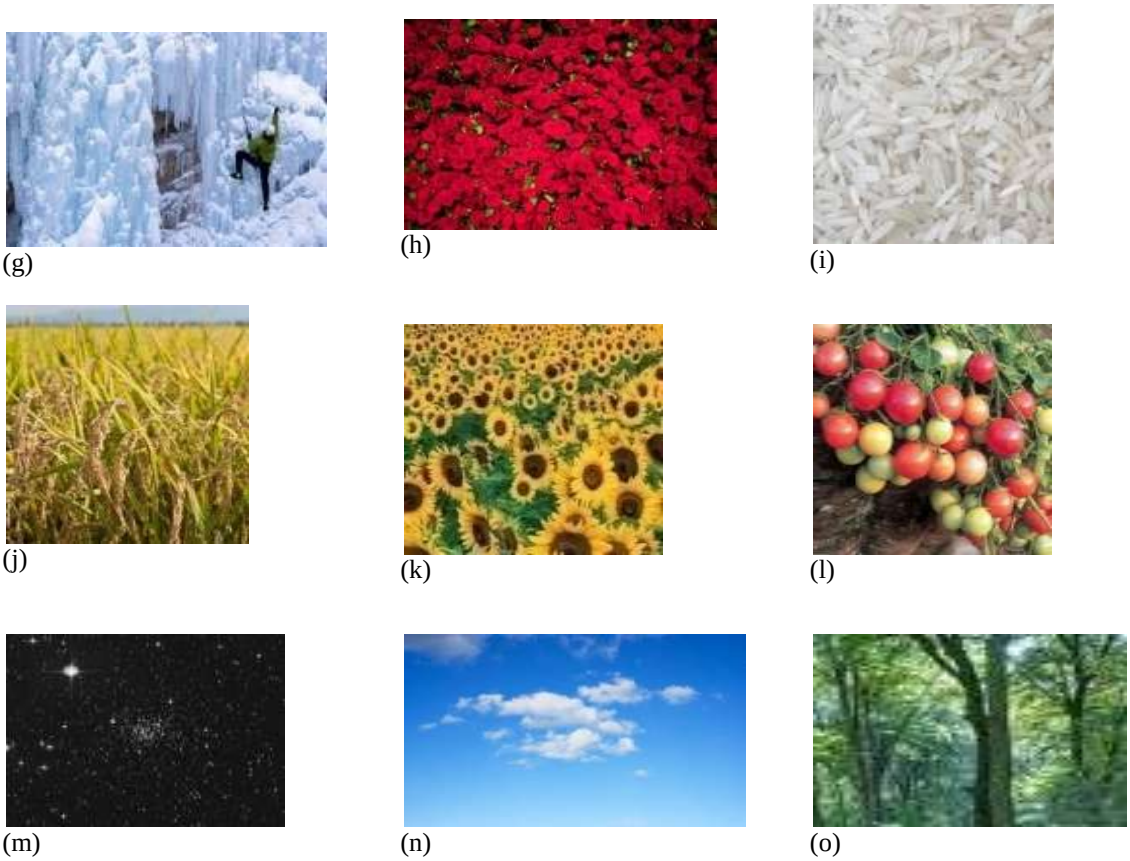


Figure 6. (a)Forest:52.8%(b)Sea:62% (c)Sand:100% (d) Soil:94.4% (e) Apple Garden:80.6% (f)Orange Garden:80% (g)Ice:70% (h)Red Rose Garden:100% (i) Rice:100% (j) Rice Plant:63.89% (k)Sunflower Garden: 92% (l)Tomato Garden:61%(m)Star Sky:91%, (n)Sky:70% (o)Forest:89.37%

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%

This section describes an efficient approach for content based image retrieval using local region based on semantic modeling only for unshaped object. In the next section, we enhanced this proposed model such that it can work both with shaped and unshaped objects. 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].

IV. Proposed Semantic Modeling fortheShaped Object

The section initially illustrates all the modifications required in the proposed unshaped model to detect the shaped objects. Throughout the paper, we denote this new model as integrated model. All explanations of the proposed integrated models are then discussed. The performance study of the proposed integrated model concludes the section.

1.7 Proposed Model

The proposed integrated model consists of extratwo steps in addition to the five steps of the proposed unshaped model. The existing five steps also require several adjustments which are given below:

i. Stage one calculates normalized color finding factor for each learned object. Unique normalize color percentage of the learned object in the search image is also calculated in this stage. If the percentage is greater than a threshold, then it can be presented in the search output depending on the remaining stage calculation; but if the percentage is lower than the threshold computation stops, as its probability to become the actual search output is very low.
ii. Stage two detects all the edges fromthe search image.
iii. All the edges found in stage two are then synchronized. The synchronization removes all the edges from search image that has a mismatch with the normalized color of the learned object.
iv. Stage four removes remainingsmall particles from the edge image found in the previous stage. The connected region population is computed for this purpose. Here, if a region's population is less than the threshold, then it is considered as noise andneed to be discarded from the edge image.
v. Stage five calculates histogram dissimilarity factor. This step is identical to the unshaped model proposed in Sec. 3.
vi. Stage six determinesthe shape similarity percentage between the learned object and the search object edge area. Similarity percentage is used for the final stage.
vii. The final stage (stage seven)calculates the occurrence probability of the learned object in the search image. The histogram dissimilarity factor and shape similarity percentage are used for these purposes as well. If the probability is less than one-half (50%), then the learned object is not presented in the search image. However, if there are multiple learned objects in the search image whose probability is more than or equal to one-half (50%), the learned object with highest probability is presented as the find output.

In next section, each of the above steps is discussed in detail along with their theoretical explanation, working procedures and corresponding examples. Few sample learned objects shown in Fig. 7 and a summarized simulation of the proposed integrated model is shown in Fig. 8. These figures are used in next section for the explanation and exemplification of the proposed integrated model.

Tiger

Lion

Elephant

Zebra

Cat

Chicken

Cow

Deer

Plane Ship Mango Car

Figure 7. Sample Learned Shaped Objects

Stage 1: User input image

Stage 2: Edge detection

Stage 3: Edge synchronize

Stage 4: Removed small particles

Stage 5: Histogram dissimilarity

Stage 6: Learned object

Stage 7: Curve similarity
(Tiger: Probability: 78.10%, Histogram Difference: 0.72, Curve Similarity: 63.38%)

Figure 8.Shape Object Detection Execution Steps

Algorithm 1: Normalized color finding factor***NCFF***(*L*, *S*) for the proposed integrated sematic model

Input: Learned object's unique normalized color set $L(L_1,L_2,\ldots\ \ldots\ \ldots,L_n)$ and Search image normalized color set $S(S_1,S_2,\ldots\ \ldots\ \ldots,S_n)$
Output: Unique normalized color percentageof the learned object present in the search image.
begin
1. *match*← 0
2. ***for***EachNormalizedColor in *L****do***
3. ***if***Scontainsthe Normalized Color ***then***
4. *match*++
5. ***end if***
6. ***end for***
7. ***return****match*/|*L*|×100

end

1.7.1 Proposed normalized color finding factorfor the proposed integrated model

Algorithm for the normalized color finding factor procedure of the proposed integrated semantic model is shown in Algorithm 1. Algorithm 1 uses equation (7) for the RGB images, where*ColorRColorG* and*ColorB* represent the Red, Green and Blue color of a pixel; *NormalizeR*, *NormalizeG*, and *NormalizeB*represent the Normalized Red, Green and Blue color of that pixel respectively.

$$NormalizeR = \left(\frac{ColorR}{16}\right) * 16 + \frac{16}{2}$$
$$NormalizeG = \left(\frac{ColorG}{16}\right) * 16 + \frac{16}{2} \ldots\ldots\ldots (7)$$
$$NormalizeB = \left(\frac{ColorB}{16}\right) * 16 + \frac{16}{2}$$

From Algorithm 1, we find that the initial value to the *match* is zero and then it is incremented whenever the search image set contains the unique normalized color of the learned object. The computation continues for each normalized color. Finally, line 7 of the Algorithm 1 returns the percentage of the unique normalized color of the learned object exists in the search image. Fig. 9 shows two sets of sample images which is used during simulation for the execution of Algorithm 1.

(a)

(b)

Figure 9.Sample Training ImagesSet (a) Tiger (b) Car

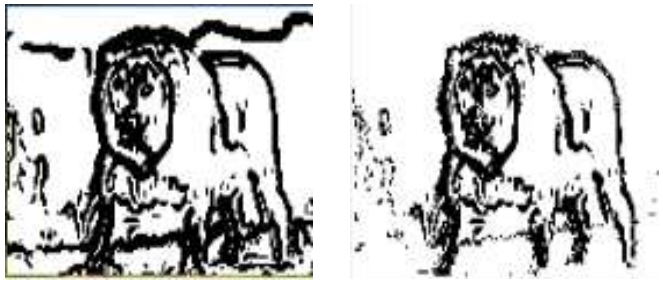
Original Image Smooth Image Edge of the Image

Figure 10. Edge Detection of the Search Image

1.7.2 Edge detection and synchronization for the proposed integrated model

The next step of the proposed algorithm detects edges of the input image. Improved Sobledge detection mechanism [41] has been used for the purpose asit restrainsnoises from the input image. In the proposed scheme, a simple image smoothing process is applied before the Sobelprocedure to reduce supplementary noises. The effect of smoothing in the input image is shown in Fig. 10.

From Fig. 10 we find that some edges actually don't belong to the learned object such as background edges. These edges need to removed, otherwise it will directly affect the output accuracy[2]. Thus we used edge synchronization to remove these extra edges. Design procedure of the proposed edge synchronization mechanism is shown in Algorithm 2. According to the proposed algorithm, any pixel in the input image whose normalize color doesn't exists in the learned objects is discarded (lines 4 and 5 of Algorithm 1). The complexity of this algorithm is directly proportional to size of the input image. The working procedure of the Algorithm 2is shown in Fig. 11. This figure shows that the proposed synchronization removes lots of background noises and hence increases the search image's probability to match one of the learned objects (lion in the example).



Edge the Image Edge Synchronization

(a) (b)

Figure 11. Edge Synchronization

[2]Since the edge pixels RGB normalized color does not exists in the learned object normalized.

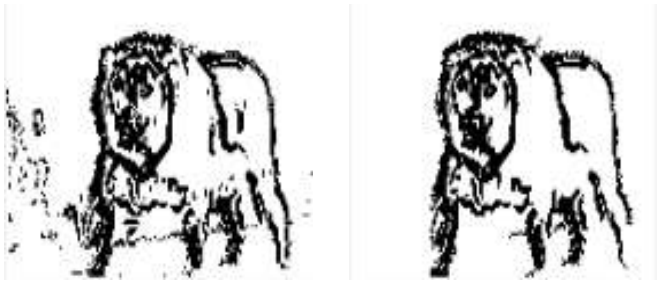
Algorithm 2: Edge synchrone ***ES***(*E*, *L*) for the proposed integrated sematic model

Input: Search image edge *E*, Learned object unique normalize color set $L(L_1,L_2,\ldots\ \ldots\ \ldots, L_n)$
Output: Synchronized search image'sedges.
begin
1. ***for****i*← 0 to |*E.width*|-1***do***
2. ***for****j*← 0to |*E.height*|-1 ***do***
3. ***if****E.pixel*(*i*,*j*)isBlack***then***
4. ***if****L*doesn't contain the Normalized RGB Color of*E.pixel*(*i*,*j*)***then***
5. *E.setPixel*(*i*,*j*,*White*)
6. ***end if***
7. ***end if***
8. ***end for***
9. ***end for***
10. ***return****E*

end

1.7.3 Small particle removal for the proposed integrated model

Fig. 11(b) shows that there are few small particles in the synchronizedoutput image. These particles represent noises and need to be removed for more accurate retrieval. According our design procedure, an integrated Breadth First Search (BFS) and Depth First Search (DFS) algorithm is used to remove these small particles with respect to connected regions[3]. As shown in Sec. 2, the length of these particles is very small. Thus, the proposed method removed any connected region that doesn't satisfy the requiredlength. Algorithm 3 presents the pseudocode of the proposed small particle removal procedure. Input to algorithm is search synchronize imageand thethreshold of the length (limit). Output of the algorithm is an image without any samll perticle.The effect of Algorithm 3 is shown in Fig. 12. From this figure, we find that theno. of small particles is much lower here. Hence, the search image's probability to become lion increases.



Edge Synchronization Small particle remove

Figure 12.Small Particle Removal Process

[3] A brief introduction to the connected region is already discussed in Sec. 2.4. Elaborate discussion on integrated BFS, DFS and connected region is beyond the scope of the paper. Interested readers are asked to see [15, 24, 27] for more detail.

Algorithm 3: Small Particle Removal ***RSP***(*SE, Limit*) for the proposed integrated sematic model

Input: Search (edge) synchronized image (*SE*), Minimum length(*Limit*).
Output: Search (edge) image without the small paricles.
begin
1. ***for****i*←0to |*SE.width*| - 1***do***
2. ***for*** j←0to |*SE.height*|-1***do***
3. ***if****SE.pixel*(*i*,*j*)is*Black****then***
4. ***if****SE.pixel*(*i*,*j*)length <*Limit****then***
5. *SE.setPixel*(*i*,*j*,*White*)
6. ***end if***
7. ***end if***
8. ***end for***
9. ***end for***
10. ***return****SE*

end

1.7.4 Histogram dissimilarity calculation for the proposed integrated model

The fifth stage of the proposed semantic model for the shaped object calculates histogram dissimilarity factor

between the learned object and the search image (edges) RGB regions. A threshold is used to determine whether the learned object is absent in the search image or not. Histogram Dissimilar Factor for the shaped object is defined as follows:

$$F = \sum_{i=0}^{255} (|R_i - RL_i| + |G_i - GL_i| + |B_i - BL_i|) \dots\dots\dots (8)$$

In Eq. (8), R_i, G_i and B_i repesent the percentage of i^{th}value in the red, green and blue coloramong the edges of the search image whereas, RL_i, GL_i and BL_i repesent the percentage of i^{th} value in the red, green and blue color of learned objects, respectively. Fig. 13 presents the sample simulation result of the proposed histogram dissimilarity calculation.

RGB Image Image Edge RGB region between the edge

Figure 13. RGB Region between the Edges of the Search Image

(a) Search Image (b) Search Edge (c) Combined Edge (d) Learned Image (e) Learned Edge

Figure 14. Shape Similarity Percentage

1.7.5 Shape similarity percentagefor the proposed integrated model

In the proposed model, shape similarity percentage between the learned objects and the search image edge area is calculated according to Eq. 9 where, *SC* is similar edge pixel count, *DC* is dissimilar edge pixel count and *CS* is shape similarity percentage.

$$CS = \left(\frac{SC}{SC+DC}\right) \times 100 \ \% \quad \dots\dots\dots (9)$$

As shown in Fig. 14, the proposed shape similarity percentage placed the learned image edge and search image edge in the same frame. Here, the black color combined edge indicates that these edges exists in both shapes (Fig. 14 (b)), but red color combined edges indicate that these edges exists only in the learned shape and yellow color combined edge indicates that these edges exists only in search shape (Fig. 14 (c)).

1.7.6 Probability calculation for the proposed integrated model

The final stage of the proposed semantic shaped model determines probability of the learned object existence in the search image using histogram dissimilarity factor and shape similarity percentage. If the probability is less than half (50%) then we consider that the learned image is absent in the search image. However, if there are multiple learned objects then the learned image with highest probability is shown as the final output.Eq. (10) defines this probability calculation and Fig. 15 presents a sample execution of the final step.

$$P(x) = \left(\frac{4-F}{4}\right) \times 80 + \left(\frac{CS}{100}\right) \times 20 \ \% \quad \dots\dots\dots (10)$$

Search Image Learned Object

Figure 15. Probability Calculation (65.07%:*F*=1.36901, *CS*=62.27642%)

Search Image Learned Image

Figure 16. Cow (Learned Cow: 64.43%, Histogram Dissimilarity: 1.19, Shape Similarity: 41.59%).

1.8 Performance Evaluation of the Proposed Integrated Semantic Model

The proposed integrated algorithm used normalized color finding factor, which reduces lot of the searching space. Specifically, if an image does not contain any of the learned objectsthen the normalize color finding factordetectsit with an average of 80% accuracy. Moreover, the edge synchronization removes the unwanted (object's) edge from the input image. Then the small particle removal procedure removes noises from the input image. In addition, histogram dissimilarity factor matches the color, whereas shape similarity finds the shape similarity between the search image object and learned object. All these steps together increase accuracies of the proposed model. Two sample execution results are shown in Figs. 16 and 17. Tables VII to IX represent the overall performance of the proposed integrated semantic model for the shaped object. From TableVII(a), we find that the accuracy of the proposed model increases along with the number of training images and tends to become stable after twenty training images. Therefore, we consider twenty training images for the performance evaluation. On the other hand, Table VII (b) shows that the accuracy of the proposed method is slightly decreased when number of learned objects and experiment images increase. Here, the accuracy also tends to become stable after twenty learned objects. These tables also show that the proposed integrated method achieves 76.28% accuracy which is much higher than the existing shaped approaches [4] and [42], which achieve 34.8% and 68.78% accuracies respectively.

TABLE VII. PROPOSED INTEGRATED MODEL ACCURACY ON(A) TRAINING IMAGES(B) OBJECTS

(A)

Training Images	No. of Experiment Images	Accuracy Percentage
1	2500	7.571%
3	2500	16.350%
7	2500	39.743%
10	2500	55.829%
15	2500	68.822%
20	2500	76.285%
30	2500	76.286%
40	2500	76.287%
50	2500	76.287%
500	2500	76.287%
1000	2500	76.287%

(B)

No. of Object	No. of Experiment Images	Accuracy Percentage
1	500	89.11%
5	500	86.54%
10	1000	82.56%
15	1500	79.19%
20	2500	76.28%
30	3000	76.14%
40	4000	75.89%
50	5000	75.67%
100	10000	75.44%
500	100000	75.29%
1000	5000000	75.13%




(a) (b) (c) (d)

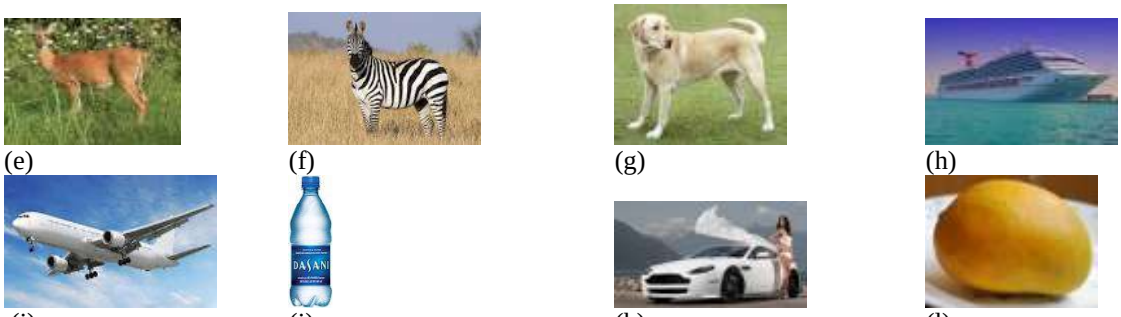


Figure 16. Search image simulation results (a) Cow: 79.08% (b) Tiger: 89.22% (c) Lion: 86.52% (d) Elephant: 80.35% (e) Deer: 76.55% (f) Zebra: 79.05% (g) Dog: 80.66% (h) Ship: 84.62% (i) Airplane: 79.37% (j) Bottle: 76.85% (k) Car: 76.63% (l) Mango: 88.83%.

TABLE VIII. PROPOSED INTEGRATED MODEL ACCURACY ONDIFFERENT CATEGORY

Object Name	No. of Experiment Image	No. of Object Detected Accurately
Tiger	214	167
Lion	234	177
Elephant	154	114
Zebra	183	138
Cat	191	147
Chicken	183	133
Cow	162	119
Deer	113	85
Dog	185	143
Bottle	227	177
Car	241	188
Mango	138	103
Plane	217	169
Ship	207	157

TABLE IX. ACCURACY OF THE PROPOSED INTEGRATED MODEL DIFFERENT COLOR SYSTEM

No. of Object	No. of Experiment Image	HIS Color Histogram	Gray Scale Color Histogram	RGB Color Histogram
5	500	79.91%	65.75%	86.54%
10	1000	77.67%	61.19%	82.56%
15	1500	74.12%	58.84%	79.19%
20	2000	72.58%	56.27%	76.28%

V. Conclusion

The paper has proposed an efficient content based image retrieval model using semantic concept for both the shaped and unshaped objects. The proposed integrated semantic model reduces the semantic image translation gaps between the human and computer. The proposed five steps based unshaped model shows how effectively local semantic content can be utilized for the image categorization. The neighborhood probability process further increases the mentioned effectiveness. The simulation results show that the proposed unshaped model achieves89.86%classification accuracy, which is much higher than the most efficient existing unshaped state-of-the-art method [15].The proposed unshaped model is then extended for the shaped objectretrievably adding two more steps and modifying the existing stages. In the proposed shaped model, the objectsaredetectedusing both itscolors and shapes. Theinitial step of the proposed extended model uses normalized color finding factor, which minimizes the search space by a margin. Then the remaining steps of the proposed integrated model compute in these reduce search space to improve the retrieval accuracy. The simulation results show that the integrated shaped model achieves 76.28% classification accuracy, which is much higher than the existing shaped approaches [4, 42]. The proposed integrated model compares an image object with the entire learned categories. This is time consuming specifically for large database. Minimizing these time complexities can be an interesting future work. The indexing mechanism can be used for this purpose. In such case, the index categories that have the maximum probability to exist in the object will become the final output. The proposed integrated model can also be used alike the proposed unshaped model*e.g.*, surveillance systems [9], driver assistance systems [10], human-robot interaction [11], interactive entertainments [12], smart assistance [13], military appliances [14] etc.

References

[1]. S. Agarwal, A. Awan andD. Roth, "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), 2005.

[3]. S. Feng, R. Manmatha and V. Lavrenko, "Multiple bernoulli relevance models for image and video annotation", in CIVR, Dublin, Ireland, 2004.

[4]. F. S. Khan, R. M. Anwer, J. V. D. Weijer, A. D. Bagdanov, M. Vanrell and A. M. Lopez, "Color Attributes for Object Detection", Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp: 3306-3313, 2012.

[5]. R. Zhang, Z. Zhang and Z. Qin, "Semantic repository modeling in image database", IEEE International Conference on Multimedia and Expo (ICME), vol.3, pp:2079-2082, 2004.

[6]. G.Passino,I. Patras and E. Izquierdo, "Pyramidal Model for Image Semantic Segmentation", 20thInternational Conference on Pattern Recognition (ICPR), pp.1554-1557, 2010.

[7]. Z. Xu and Zhang S., "Artificial Visual Cortical Responding Model in Image Semantic Processing", IEEE/ACS International Conference on Computer Systems and Applications, pp:722-725, 2007.

[8]. V. Lavrenko, R. Manmatha and J. Jeon, "A model for learning the semantics of pictures", in Advances in Neural Information Processing Systems (NIPS), MIT Press, pp: 553-560, 2003.

[9]. Z. Wu, andR.J. Radke, "Real-time airport security checkpoint surveillance using a camera network", in Computer Vision and Pattern Recognition Workshops, pp.25-32, 2011.

[10]. S.Sivaraman and M. M. Trivedi, "Towards CooperativePredictive Driver Assistance", inProceedings of the 16thInternational IEEE Annual Conference on Intelligent Transportation Systems (ITSC), pp: 1719-1724, 2013.

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

[12]. J. Chung, J. R. Kim and K. Shim, "Vision Based Motion Tracking System for Interactive Entertainment Applications", IEEE TENCON, IEEE Region 10 , pp: 1-6, 2005.

[13]. A. C. M. Fong, B. Fong, and C. K. Li, "Assistive independent living for senior citizens with special cognitive needs", IEEE 1stGlobal Conference on Consumer Electronics (GCCE), pp: 448-449, 2012.

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

[15]. J. Vogel and B.Schiele, "Semantic Modeling of Natural Scenes for Content-Based Image Retrieval", in Int. J. Comput. Vision vol.72, no.2 pp-133-157, 2007.

[16]. A. Oliva, A. Torralba, A. Guerin-Dugue and J. Herault, "Global semantic classification of scenes using power spectrum templates", in Proceedings of the international conference on Challenge of Image Retrieval, pp: 9-9, 1999.

[17]. A. Mojsilovic, J. Gomes and B. Rogowitz, "Semantic-friendly indexing and querying of images based on the extraction of the objective semantic cues", in International Journal of Computer Vision, vol. 56, no. 1/2, pp: 79–107, 2004.

[18]. N. Serrano, A. Savakis and J. Luo, "Improved scene classification using efficient low-level features and semantic cues", in Pattern Recognition, vol. 37, no. 9, pp: 1773–1784, 2004.

[19]. M. Boutell, J. Luo, X. Shen and C. Brown, "Learning multi-label scene classification",inPattern Recognition, vol. 37, no. 9, pp. 1757–1771, 2004.

[20]. Y. Chen, H. Sampathkumar, B.Luo and X. Chen, "Bridging the Semantic Gap in Vertical Image Search by Integrating Text and Visual Features", IEEE Transactions on in Knowledge and Data Engineering, vol.25, no.10, pp:2257-2270, 2013.

[21]. E.Agichtein, E. Brill and Susan Dumais, "Improving web search ranking by incorporating user behavior information," in Proceedings of the 29thannual international ACM SIGIR conference on Research and development in information retrieval, pp 19-26, 2006.

[22]. J. Eakins and M. Graham, "Content-based image retrieval", A report to the JISC Technology Applications Programme, Tech. rep., Institute for Image Data Research, University of Northumbria at Newcastle, 1999.

[23]. A. Smeulders, M. Worring, S. Santini, A. Gupta and R. Jain, "Content-based image retrieval at the end of the early years", in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, no. 12, pp. 1349–1380, December 2000.

[24]. R. Veltkamp and M. Tanase, "Content-based image retrieval systems: A survey", Tech. rep., Department of Computer Science, URL http://www.aa-lab.cs.uu.nl/cbirsurvey/cbir-survey/, Utrecht University, 2001.

[25]. Ji-quan Ma, "Content-Based Image Retrieval with HSV Color Space and Texture Features", in Proc. of the International Conference on Web Information Systems and Mining, pp: 61-63, 2009.

[26]. A. Essam, El-Kwae and M. R. Kabuka, "Efficient content-based indexing of large image databases", in ACM Trans. Inf. Syst. vol.18, no.2, pp: 171-210, 2000.

[27]. N. Zhang and Y. Song, "An Image Indexing and Searching System Based Both on Keyword and Content", in Proceedings of the 4thinternational conference on Intelligent Computing: Advanced Intelligent Computing Theories and Applications-with Aspects of Theoretical and Methodological Issues, pp: 1032-1039, 2008.

[28]. D. D.Burdescu, C. G.Mihai, L.Stanescu and M.Brezovan, "Automatic image annotation and semantic based image retrieval for medical domain", inNeurocomput.,vol.109, pp: 33-48, 2013.

[29]. A. K.Choupo, L.Berti and A. Morin, "Optimizing progressive query-by-example over pre-clustered large image databases", In Proceedings of the 2ndinternational workshop on Computer vision meets databases, pp: 13-20, 2005.

[30]. O.Linde and T.Lindeberg, "Composed complex-cue histograms: An investigation of the information content in receptive field based image descriptors for object recognition", in Com. Vis. Image Underst. Vol.116, no.4, pp:538-560, 2012.

[31]. T.Deselaers, D.Keysers and H. Ney, "Features for image retrieval: an experimental comparison", in Inf. Retr. vol.11, no.2, pp: 77-107, 2008.

[32]. Alta Vista, URL: www.altavista.com, Last accessed Feb 9, 2015.

[33]. C. Carson, S.Belongie, H. Greenspan and J. Malik, "Blobworld: Image Segmentation Using Expectation-Maximization and Its Application to Image Querying",inIEEE Trans. Pattern Anal. Mach. Intell. 24, vol.8 pp: 1026-1038, 2002.

[34]. Z. Li, O. R. Zaïane and B. Yan, "C-BIRD: Content-Based Image Retrieval from Digital Libraries Using Illumination Invariance and Recognition Kernel", in Proceedings of the 9thInternational Workshop on Database and Expert Systems Applications, pp:361-367, 1998.

[35]. T. D.Mascio, D.Frigioni and L. Tarantino, "VISTO: A new CBIR system for vector images", in Inf. Syst. vol.35, no.7, pp:709-734, 2010.

[36]. Image Retro, URL: http://carol.wins.uva.nl/vendrig/imageretro/, Last accessed Feb 15, 2015.
[37]. P. Chandrika and C. V. Jawahar, "Multi modal semantic indexing for image retrieval", in Proceedings of the ACM Int. Conference on Image and Video Retrieval, pp:342-349, 2010.
[38]. R. Hu, D. Larlus and Gabriela Csurka, "On the use of regions for semantic image segmentation", in Proceedings of the Eighth Indian Conference on Computer Vision, Graphics and Image Processing, vol. 51, pp: 1-8 , 2012.
[39]. L. D. Nguyen, G. E. Yap, Y. Liu, A. H. Tan, L. T. Chia, and J. H. Lim, "A Bayesian approach integrating regional and global features for image semantic learning", in Proceedings of the IEEE international conference on Multimedia and Expo, pp: 546-549, 2009.
[40]. R. S. Choras, "Feature extraction for CBIR and biometrics applications", In Proc. of the 7thConference on Applied Computer Science, vol. 7, pp: 1-9, 2007.
[41]. W.Gao, X. Zhang, L. Yang and Huizhong Liu, "An improved Sobel edge detection", in 3rd IEEE Int. Conference on Computer Science and Information Technology (ICCSIT), vol.5, pp.67-71, 2010.
[42]. S. Agarwal, A. Awan and D. Roth, "Learning to detect objects in images via a sparse, part-based representation", inIEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 26, no.11, pp: 1475-1490, 2004.