An Efficient Semantic Image Retrieval based on Color and Texture Features and Data Mining Techniques

3 authors, including:

Hala Zayed
Egypt University of Informatics (EUI)
94 PUBLICATIONS   **1,632** CITATIONS

An Efficient Semantic Image Retrieval based on Color and Texture Features and Data Mining Techniques

Doaa M. Alebiary
Department of computer
Science, Faculty of computers
and informatics
Benha University Benha,
Egypt

Noura A. Semary
Department of Information
Technology,Faculty of
computers and information
Menufia University Shebeen
El-Kom, Egypt

Hala H. Zayed
Department of computer
Science, Faculty of computers
and informatics
Benha University, Benha,
Egypt

ABSTRACT

Content-based Image Retrieval (CBIR) is retrieving the desired images from huge collections. The user queries are becoming very specific and traditional text-based methods cannot efficiently handle them. CBIR system retrieves the image via low-level features such as color, texture and shape. In this work, we propose CBIR system that retrieves images from a database based on the semantic features of them.

Our methodology divide the query image into 100 regions. And then, extracts Features Vector from each region and label each one with the suitable concept like (Sky, Sand, Water, trunks, foliage, rocks,..., and Grass). The labeling process in performed semi-automatically using k-means clustering and KNN classification algorithms. The system has been evaluated by recall and precision measures and compared to other recent works. The results of the paper reflects the efficiency of the system for retrieving images with up to 98% recognition ratio.

General Terms

Image processing, Image representations, Visual content-based indexing and retrieval.

Keywords

Big Data; Content-Based Image Retrieval; High-Level Semantics; Semantic Gap.

1. INTRODUCTION

Image search engines attempt to give access to the wide range of images available on the Internet. There are two approaches for image retrieval: text-based and content-based. The text-based approach can be tracked back to 1970s [9].

In such systems, the images are manually annotated by text descriptors, which are then used by a database management system to perform image retrieval. There are two disadvantages with this approach, the first is that a considerable level of human effort is required for manual annotation. The second is the annotation inaccuracy due to the subjectivity of human perception. To overcome the above disadvantages in text-based retrieval system, content-based image retrieval (CBIR) was introduced in the early 1980s.

Content-Based Image Retrieval (CBIR) systems aim to recognize and retrieve information based on content of images instead of looking at metadata provided with the images. The problem with searching in these features in images is the

semantic gap. The semantic gap is the difference between the high-level user semantics and the low-level features which need to be connected. We describe a CBIR methodology for the retrieval of images of natural scenes, whereas for humans

the content of an image refers to what is seen on the image, e.g. " a forest, a house, a lake ". One of the research issues in content-based image retrieval is to reduce this semantic gap between the image understanding of humans and the image understanding of the computer, Humans tend to use high-level features (concepts), such as keywords, text descriptors, to interpret images and measure their similarity. While the features are automatically extracted using computer vision techniques are mostly low-level features (color, texture, shape, spatial layout, etc.). In general, there is no direct link between the high-level concepts and the low-level features. Moreover, when database becoming larger the semantic annotations process becoming harder. In this paper, we propose a semi-automatic semantic based CBIR that using data mining techniques for automatic semantic database building. The main used features to describe the semantic features are a combination of color and texture. The proposed methodology is applied on natural images. More details will be presented in section 3. The organization of this paper comes as follows. In Section 2 we present a state of the art review of the most important works of image retrieval, mainly CBIR systems.

In Section 3, each of the stages of the proposal (training and testing) is detailed. In Section 4 several experiments to test the performance of our methodology are presented. Finally the conclusions and further research are detailed in Section 5.

2. RELATED WORKS

Large storage spaces and a huge number of images can be found over the Internet. With this huge image database, people want to find accurate ways to search in these images. Bimbo et. al [2], worked in CBIR by using automatic feature extraction and interactivity with visual content such as color, texture and object shape. Vogel et. al [13], proposed a CBIR system that extracts the semantic content of images. The basic idea of the semantic modeling is to classify local image regions into semantic concept classes. Li et. al [8] used advancing statistical modeling and optimization techniques to train computers about hundreds of semantic concepts using example pictures from each concept and data representation. Hiremath et.al [6] proposed a framework for combining color, texture and shape information, and achieved higher retrieval efficiency. The image is partitioned into non-overlapping tiles of equal size. Vogel et. al [12], proposed a retrieval system that splits the retrieval process into two stages. In first only small patches of the image are analyzed and second stage the patch information is processed and the relevant images are retrieved. While Vogel et. al in [14], took a closer look at the influence and interaction of local vs. global information in scene categorization. Serrano et. al [11], described a new statistically-based methodology to organize and retrieve

images of natural scenes by combining feature extraction, automatic clustering, automatic indexing and classification techniques. Helala [5], propose a new approach that extracts interest salient regions that work as local descriptors, a greedy graph matching algorithm with a proposed modified scoring function is applied to determine the final image rank. Helala [4], proposes a new region-based image retrieval technique called Principal Regions Image Retrieval (PRIR). This technique starts by segmenting an image to the most general principal regions and applies a fuzzy feature histogram to describe the color and texture properties of each segmented region. In this work, the retrieval is performed by two steps. First, the images is divided into equal sized block. The semantic description of each image block is defined semi-automatically using data mining techniques. Second, the image is analyzed to semantic descriptors which is used in the final retrieval stage.

3. PROPOSED METHOD
This section describes each step in the proposed methodology for semantic image retrieval of natural scenes from a database. The methodology has two basic stages Semantic database building (SDB) stage and semantic image retrieval (SIR) stage.

3.1 Semantic database building (SDB)
The SDB block diagram is shown in Figure 1 First, the images of the database are loaded and divided in equal blocks. Color and Texture features are extracted from each block and saved in Low-Level features database (DB1). For automatic annotation, the records of DB1 are clustered by K-Means to K classes. Each class is annotated to human perception keywords. Each image is described by a semantic feature vector which is saved in an indexed database DB2. Each step in this stage is detailed in the following subsections.

3.1.1 Preprocessing
before feature extraction, all the images in the database (assume N images) are scaled to 250×250 pixels. Then all the images are converted from RGB format to HSV format.

3.1.2 Features Extraction
All images in the database are divided into a regular grid of 10×10 image regions with size 25×25 pixels. For each region, three features are extracted for each block in three channels hue (H), saturation (S) and value (V). Two color features; the standard deviation and the mean, and one texture feature; the homogeneity (obtained from the co-occurrence matrix) [7].

3.1.3 Standard deviation
In image processing shows the variety or "dispersion" exists from the average (mean, or expected value). A low standard deviation refers to the data points tend to be very close to the mean, High standard deviation refers to the data points are spread out over a large range of values [7]. Standard deviation for each block is given by (1)

$$std(x,y) = \sqrt{\frac{1}{mn} \sum_{r,c\epsilon w} (f(r,c) - \frac{1}{mn} \sum_{r,c\epsilon w} f(r,c))^2} \qquad (1)$$

where x and y are the row/column indexes of the block, m and n are the height and width of each block ,respectively. r and c are the row and column of the pixel value within the block w and f is the image channel.
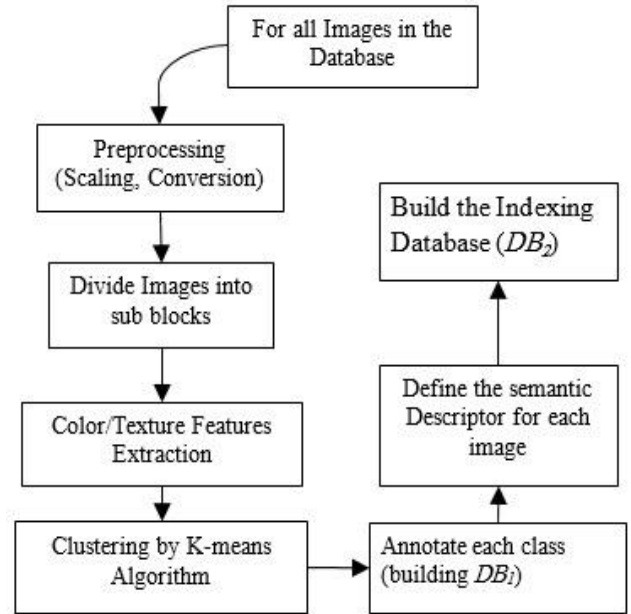

Fig. 1 Flow Chart For Building the Semantic Indexed Database

The Mean is calculated by averaging all the pixels values within the block. Mathematical formulation of the mean is given in (2)

$$mean(x,y) = \frac{1}{mn} \sum_{r,c\epsilon w} f(r,c) \qquad (2)$$

Homogeneity measures the uniformity of the non-zero entries in the gray level Co-occurrence Matrix (GLCM). It weights values by the inverse of contrast weight [3]

$$hem(x,y) = \sum_{r,c\epsilon w} \frac{1}{1-c^2} glcm(r,c) \qquad (3)$$

The GLCM homogeneity of any texture is high if GLCM concentrates along the diagonal, meaning that there are a lot of pixels with the same or very similar grey level value. The larger the changes in grey values, the lower the GLCM homogeneity making higher the GLCM contrast. The range of homogeneity is [0, 1]. If the image has little variation then homogeneity is high and if there is no variation then homogeneity is equal to 1 [10]. Therefore, high homogeneity refers to textures that contain ideal repetitive structures, while low homogeneity refers to big variation in both, texture elements and their spatial arrangements. An "inhomogeneous texture" refers to an image that has almost no repetition of texture elements and spatial similarity in it is absent [1].

3.1.4 Clustering (K-means)
To cluster the result of N×100 feature vectors (100 for each of the n images), we use k-means algorithm to classify these N×100 features into four classes, and then manually labeled these four classes to Sky, sand, water and grass. By the end of this phase, N 100 natural scene image regions are annotated with classes like sky, sand, water and grass. Finally, we created a database "DB1" which contains the centers of each class and an indexed database DB2 that contains N record of 4 semantic features for each image. The recorded semantic features contains the percent of each class in the image. Fig 2 shows an example of final annotated image.

3.2 Semantic Image Retrieval (SIR)

In the retrieval phase (Fig 3), a query image will go through the same steps to be automatically annotated before retrieving similar perceptual images from the indexed database.

First, the query image is divided into a regular grid of 10×10 image regions with size 25×25 pixels, then the three features (the standard deviation, the mean and the homogeneity) are extracted for each regions in three channels hue (H), saturation (S) and value (V). After that, each region is classified according to the nearest distance record in DB1. Finally, the ratio of its classes forms the semantic feature vector which is compared to the indexed database (DB2). For matching strategy, 1-NN classifier is used.

Fig. 2 Semantic Concepts

3.3 Retrieval Evaluation

For evaluating the retrieval accuracy Precision (P) and Recall (R) are calculated as (4) and (5) respectively :

$$P = \frac{Number\ of\ relevant\ images\ retrieved}{Total\ number\ of\ images\ retrieved} \times 100\% \quad (4)$$

$$R = \frac{Number\ of\ relevant\ images\ retrieved}{Total\ number\ of\ images\ in\ the\ database} \times 100\% \quad (5)$$
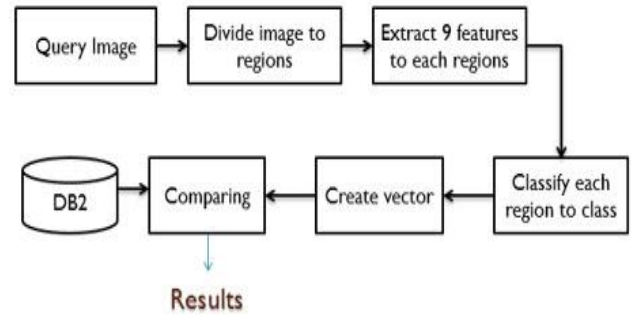
Fig. 3 Retrieved Phase

4. EXPERIMENTAL RESULTS

The System was developed by MATLAB R2012a running on an Intel coreTM I3 CPU (2.27GHz) and 2GB of RAM machine under Windows 7 32bit operating system. In our experiment, the database was built of 100 images of natural scenes collected from the internet. Accordingly, 10,000 (100×100) blocks are analyzed and their features were extracted (90,000 features). Table 1 shows how these 10,000 vectors were classified into class 1 (sand), class 2 (sky), class 3 (water), and class 4 (grass). After that, each region in image is labeled to its nearest class. Finally, each image semantic feature vector which contains the percentage of each class in image is created and saved in DB2. The indexed database DB2 is shown in Table 2 where each record represents an image that voted for its corresponding four classes. (notice: and the sum of each line is equal to 100). For example, the image 1.jpg contains 3% Sand, 26% Sky, 36% Water, 35% Grass while image 10.jpg contains 18% Sand, 0% Sky, 7% Water, 75% Grass, this refers to that this image have high percentage of grass so the semantic meaning of this image is Grass.

Table 1 Number of Regions In Each Concept Class

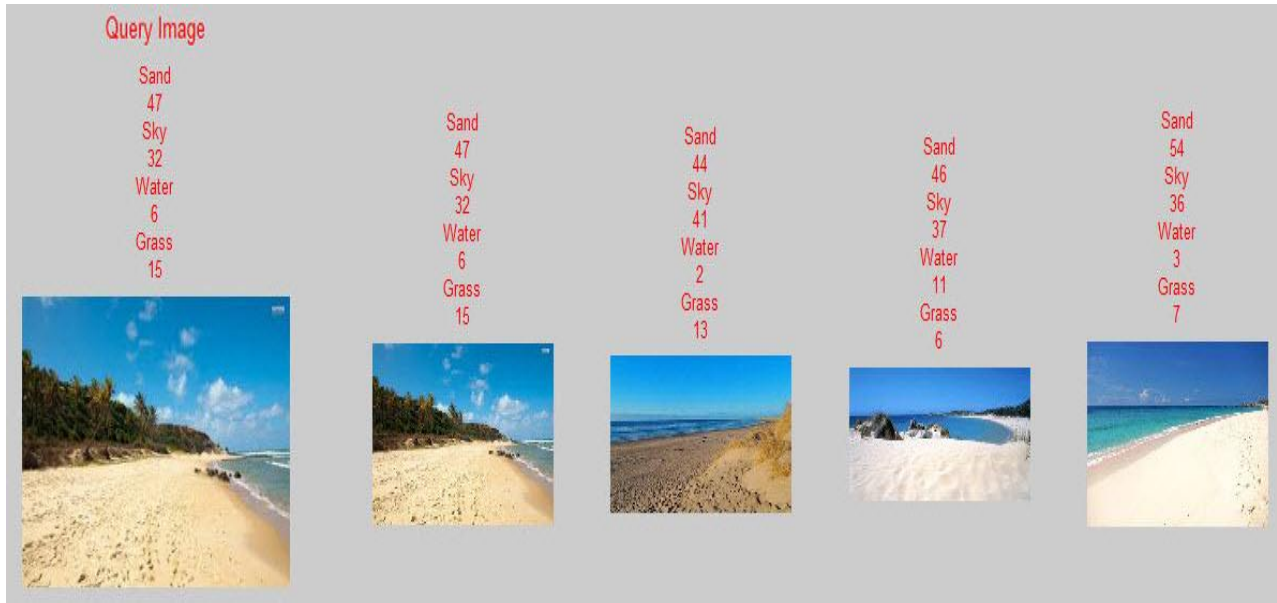| Concept Class	Number of the describing features per image classes	
1 (Sand)	2574	25.7%
2 (Sky)	2423	24.2%
3 (Water)	2582	25.8%
4 (Grass)	2421	24.2%
Total	10,000	100%

Fig. 4 An example of semantic image retrieved

Table 2 Indexed Database Structure "DB2"

Image Name	Sand	Sky	Water	Grass
1.jpg	3	26	36	35
10.jpg	18	0	7	7
14.jpg	7	59	23	11
…	…	…	…	…
100.jpg	30	0	47	23

Table 3 Example of Retrieve Image

Image Name	Sand	Sky	Water	Grass
1.jpg	*47%*	*32%*	*6%*	*15%*
2.jpg	*44%*	*41%*	*2%*	*13%*
3.jpg	*46%*	*37%*	*11%*	*6%*
4.jpg	*54%*	*36%*	*3%*	*7%*

For example: a query image has 47% Sand, 32% Sky, 6% Water, and 15% Grass is tested by the system. The major four similar votes from DB2 are presented in Table 3 . The system result is shown in Figure 4.

We took a random selection of 20 images from the database to measure the efficiency of our proposal we calculated the precision and recall. The recall of our proposed system is 94% while the precision is 98%.

4.1 Comparison with other works
The system has been compared to the system of [11], where the authors extracted random 300 points in the image and get the same features. The results of retrieving the image in Figure 4 is presented in Figure 5. More comparison results are presented in Figure 6.
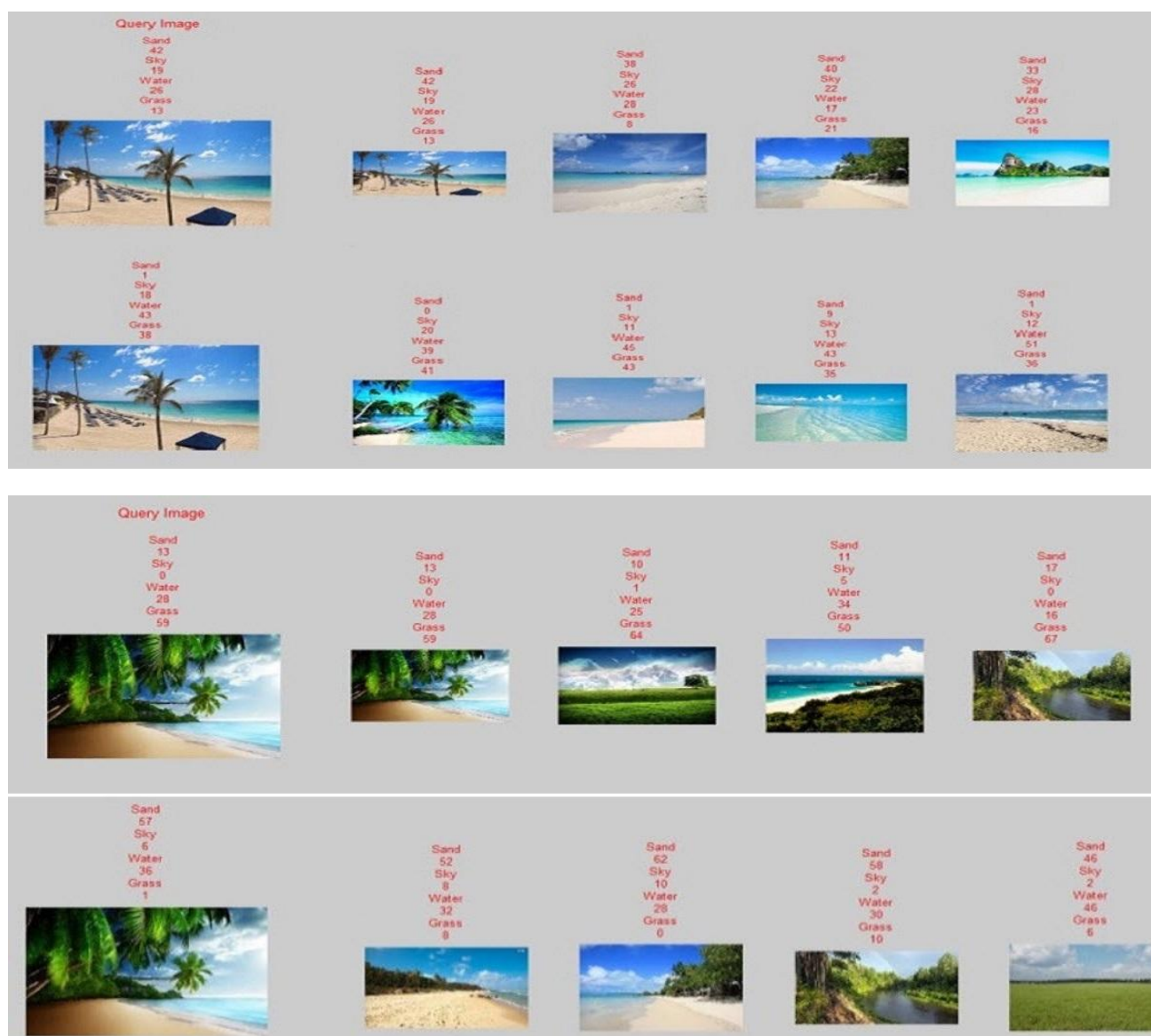
Fig. 5 The Retrieval System of [11]

Fig. 6 the Retrieval System of [11]

5. CONCLUSION

In this paper, we proposed a new methodology that allows the retrieval of natural scene images automatically. This process has been done in two steps. First, semantic database building (SDB) and semantic image retrieval (SIR). During SDB phase, we extract color (mean and standard deviation) and texture (homogeneity) features from the images in the database and transfer these low-level features to semantic descriptors using a clustering technique. K-means algorithm is used to form four different semantic clusters. K-NN classier was used to build one indexed database from the combination of all the describing vectors. During the retrieval phase the retrieved images are the most similar images given by the query image. The final results show that the system accuracy is approximately 98%. The advantages of the proposed system are the retrieval using semantic concepts and image labeling automation. Our future work is to add texture features, and trying to extract features from each region in image parallel to reduce running time.

6. REFERENCES

[1] Al Sebaey, N.A.E.M., 2007. Gray image coloring using texture similarity measures.

[2] Del Bimbo, A., 1998, June. A perspective view on visual information retrieval systems. In Content-Based Access of Image and Video Libraries, 1998. Proceedings. IEEE Workshop on (pp. 108-109). IEEE.

[3] Gebejes, A. and Huertas, R., 2013, March. Texture characterization based on grey-level co-occurrence matrix. In Proceedings in Conference of Informatics and Management Sciences (No. 1).

[4] Helala, M.A.E., Selim, M.M. and Zayed, H.H., 2009, December. An image retrieval approach based on composite features and graph matching. In Computer and Electrical Engineering, 2009. ICCEE'09. Second International Conference on (Vol. 1, pp. 466-473). IEEE.

[5] Helala, M.A., Selim, M.M. and Zayed, H.H., 2012. A content based image retrieval approach based on principal regions detection. International Journal of Computer Science Issues, 9(4), pp.204-213.

[6] Hiremath, P.S. and Pujari, J., 2007, December. Content based image retrieval using color, texture and shape features. In Advanced Computing and Communications, 2007. ADCOM 2007. International Conference on (pp. 780-784). IEEE.

[7] Kumar, V. and Gupta, P., 2012. Importance of statistical measures in digital image processing. International Journal of Emerging Technology and Advanced Engineering, ISSN, pp.2250-2459.

[8] Li, J. and Wang, J.Z., 2008. Real-time computerized annotation of pictures. IEEE transactions on pattern analysis and machine intelligence, 30(6), pp.985-1002.

[9] Liu, Y., Zhang, D., Lu, G. and Ma, W.Y., 2007. A survey of content-based image retrieval with high-level semantics. Pattern recognition, 40(1), pp.262-282.

[10] Semary, N.A., Hadhoud, M.M., El Kilani, W.S. and Ismail, N.A., 2007, March. Texture recognition based natural gray images coloring technique. In 2007 National Radio Science Conference (pp. 1-12). IEEE.

[11] Serrano-Talamantes, J.F., Avilés-Cruz, C., Villegas-Cortez, J. and Sossa-Azuela, J.H., 2013. Self organizing natural scene image retrieval. Expert Systems with Applications, 40(7), pp.2398-2409.

[12] Vogel, J. and Schiele, B., 2006. Performance evaluation and optimization for content-based image retrieval. Pattern Recognition, 39(5), pp.897-909.

[13] Vogel, J. and Schiele, B., 2007. Semantic modeling of natural scenes for content-based image retrieval. International Journal of Computer Vision, 72(2), pp.133-157.

[14] Vogel, J., Schwaninger, A., Wallraven, C. and Bülthoff, H.H., 2006, July. Categorization of natural scenes: local vs. global information. In Proceedings of the 3rd symposium on Applied perception in graphics and visualization (pp. 33-40). ACM.