LOW-LEVEL ANALYSIS AS A BASIC PROCESS FOR CONTENT-BASED IMAGE RETRIEVAL (CBIR) SYSTEM

Tatiana Jaworska

Systems Research Institute, Polish Academy of Sciences
Newelska 6 St, 01-447 Warsaw, Poland
e-mail: Tatiana.Jaworska@ibspan.waw.pl

Abstract: This article describes the way in which image is prepared for content-based image retrieval system. Automated image extraction is crucial for CBIR systems. Our system is dedicated to support estate agents. All efforts have been put into extracting elements of houses and bungalows from an image and finding their characteristic features in the unsupervised way. Hence, the paper presents segmentation algorithm based on a pixel colour and a novel method of texture identification which is based on wavelet transformation. After a set of low-level features for all objects is computed, the database is filled with these features. *Copyright © 2007 IFAC*

Keywords: content-based image retrieval, low-level analysis, image segmentation, object extraction, texture extraction.

1. INTRODUCTION

Determining how to store images in big databases, and later, how to retrieve information from them, is an active area of research for many computer science fields, including graphics, image processing, information retrieval and databases.

Although, attempts have been made to perform CBIR in an efficient way based on shape, colour, texture and spatial relations, it has yet to attain maturity. A major problem in this area is computer perception. In other words, there remains a big gap between low-level features like shape, colour, texture and spatial relations, and high-level features like windows, roofs, flowers, etc.

The purpose of this paper is to investigate image processing with special attention given to segmentation and selection of separate objects from the whole image. After computing the object features, the image and its objects are included into the database.

In the last 15 years, CBIR techniques have drawn much interest, and image retrieval techniques have been proposed in context of searching information from image databases. In the 90's the Chabot project at UC Berkeley (Ogle, 1995) was initialized to study retrieval and storage of a vast collection of digitized images. Also, at IBM Almaden Research Centre CBIR was prepared with a very interesting query system which gave the possibility for the users to define their own graphical query (Flickner *et al*., 1995), (Niblack *et al*., 1993). This approach let the users find images on the basis of extraction of a set of features. These features were defined for the whole images. This approach was improved in later research works which made possible finding the foreground object on the image basing on colour or shape (Tan *et al*., 2001), (Hsu *et al*., 2000). There also exists a database in which shape is used as a main distinguishing feature, containing about 1100 images of marine creatures, created by Mokhtarian, F. S. Abbasi and J. Kittler (Mokhtarian *et al*., 1996) at Department of Electronic and Electrical Engineering UK. Each image shows one distinct species on a uniform background. Every image is processed to

recover the boundary contour, which is then represented by three global shape parameters and the maxima of the curvature zero-crossing contours in its Curvature Scale Space image.

Bamidele, Stentiford and Morphett (Bamidele *et al*., 2004) have made some attempts to use high-level metadata. In their approach a visual attention map (VAM) is created to define regions of interest (ROI). VAM is calculated basing on dissimilarity of regions depending on colour or shape. ROIs are used to find strong similarity between objects in spite of being placed on different image backgrounds. Images in a collection are processed off-line to produce metadata stored in the database. VAM is applied to image query, whereas ROI is used to determine the similarity of images in database.

Though much effort has been invested in CBIR, many techniques have been proposed and many prototype systems have been developed, the problems with retrieving images according to image content are far from being solved. In the field of CBIR, a new research trend is to use high-level descriptions, instead of low-level features, in the matching process. Hence, the crucial point is how to extract high-level descriptions from images and to fill the gap between the low-level features and human perception of image contents.

The newest approaches combine many different methods to fill this gap, from conversion of high level semantic features using fuzzy production rules based on extraction of low level features (Stanchev, 2003) to using tree-structured features and multi-layer self-organizing map (Chow *et al*., 2006). Some promising researches have been led by Castelli (Castelli *et al*., 1998), Jaimes (Jaimes, 1999), Hong (Hong *et al*., 1999), who have suggested the use of a hierarchical structure in order to reach progressive image analysis.

In this paper we would like to focus on the low-level stage, namely segmentation level because without it building the higher level representation is simply impossible. As a result of segmentation process we obtain objects which are already a part of a higher level representation. This approach is little similar to the Wang and Rui concept (Wang *et al*., 2004). They, however, went on to semantic-level model and spatial relationship based on constraint region found in images, whereas we go further and show how to exactly extract particular objects and what spatial relationships exist between them. In order to achieve this purpose we present two new methods: one is a very fast algorithm for colour image segmentation, and the second one is a new approach to description of textured objects, using discrete wavelet transformation. For objects extracted by these two methods we intend to create spatial relations, as it has been proposed, for instance, by Song and Zhang (Song, 2003) who have analyzed scenery image by monotonic tree. In the next stage a full visual perception level will be built.
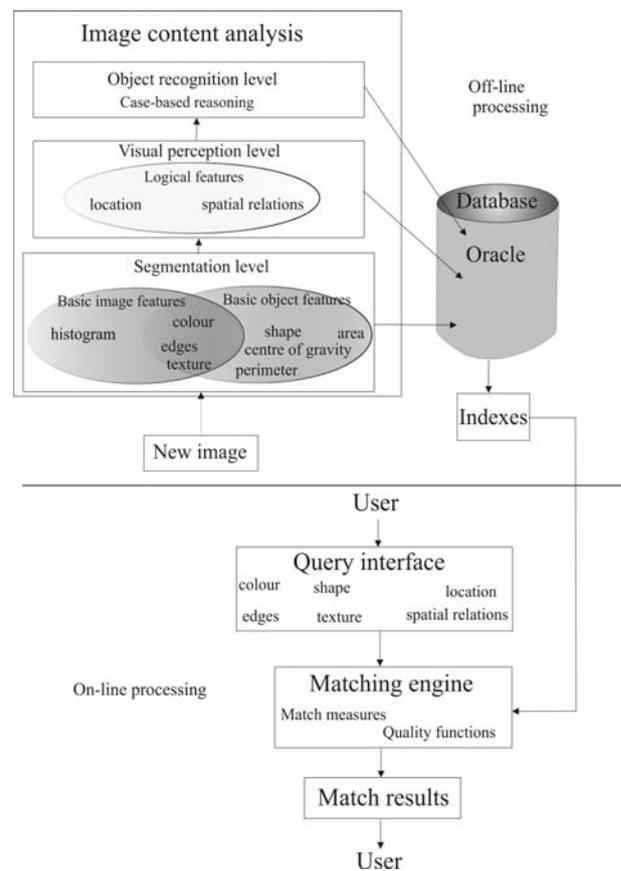
2. CBIR CONCEPTION OVERVIEW

Fig. 1. Block diagram of our content-based image retrieval system.

Our CBIR system is dedicated to support estate agents. In the estate database there are images of houses, bungalows, and other buildings. To be effective in terms of presentation and choice of houses, the system has to be able to find the image of a house with defined architectural elements, for example: windows, roofs, doors, etc (Jaworska, 2005). Figure 1 shows the block diagram of our CBIR system. As it is seen, the biggest and the most crucial part of this system is *image content analysis* block processed off-line. The hierarchical structure from the segmentation level to the object recognition level covers the present approach to a multi-layer image description model in order to reach progressive image analysis and understanding (Gao *et al*., 2000). In this structure, image content is analyzed and represented on three levels. There is a context between adjacent levels i.e. the representation for the upper level is directly extracted from the lower level.

The first stage of our analysis is to split the original image into several meaningful clusters; each of them provides certain semantics in terms of human understanding of image content. Then, proper features are extracted from these clusters to represent the image content on the visual perception level. In the interest of the following processes, such as object

recognition, the image features should be selected carefully. Nevertheless, our efforts have been put into extracting elements from an image in the unsupervised way. Another important part of the *image content analysis* block is relationship determination, which provides more semantic information about the different objects of the image. On the object recognition level, a semi-interactive procedure is used to achieve the correct recognition result. The case-based reasoning method is used, after the basic objects have been defined and classified to the particular classes.

New images (in JPEG format) as well as all results achieved on each level of *image content analysis* are stored in the related database. For our project Oracle DB system has been chosen on account of its flexibility and completeness. In order to enforce the process of image retrieval many multi-dimensional indexes have to be implemented for this database. Some of them are proposed by Amsaleg and Gros (Amsaleg *et al*., 2004) and also by Głomb (Głomb, 2006).

Fig. 2. Example of an original image.

The bottom half of the diagram is dedicated to users, hence, it is processed on-line. It comprises of the user's *query interface* and the *matching engine*. In our system the user's interface is offered on WWW page. Users can select separate features such as colour, or texture, they can also select an architectonical element, which automatically defines many features as a query. Indeed, if the users have a vague target image in mind, they have a possibility to compose their imaginary house and the system presents them some optional houses based on these chosen elements (Fauqueur, 2006).

The next element of the system is the *matching engine* which uses the distance function between the features on the low-level (Rubner *et al*., 2000), the dissimilarity function on the middle-level and the case-based reasoning on the high-level to search for images which are best compatible with the user's query. Furthermore, at this stage a careful image content analysis is very crucial, otherwise, the proper result will not be reached.

3. A NEW FAST ALGORITHM FOR OBJECT EXTRACTION FROM COLOUR IMAGES

From many existing methods of image segmentation we prefer unsupervised ones. Although we began with the C-means algorithm the results were unsatisfying because we did not obtain image elements separated according to colours.

These results forced us to work out a new algorithm which uses colour information about a single point to greater extent than the C-means algorithm does. With the aim of labelling a pixel we chose the biggest value from the triple (R,G,B) and we defined it as a cluster colour. In this way we obtained three segments – red, green and blue. Additionally, points with equal values of RGB were labelled as grey. Tentatively it was accepted, but for better result each colour was divided into three shades, according to the darkness of colour. These shades are shown as three regions (I, II, III) which determine point brightness. The idea of the segmentation is illustrated in figure 3. The radius *r* of the dividing sphere was counted in Euclidean measure, namely
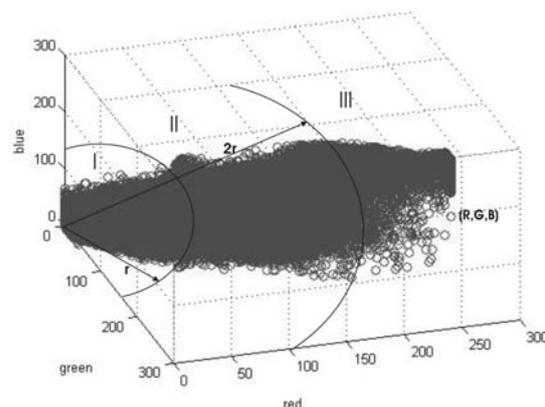

Fig. 3. The way of labelling the set of pixels. Regions I, II, III show pixel brightness and the biggest value of triple (R,G,B) determines its colour.

$$r = \frac{\sqrt{R_{max}^2 + G_{max}^2 + B_{max}^2}}{3} \qquad (1)$$

Generally, $R_{max} = G_{max} = B_{max} \# 255$ because full saturation of colours is rare. Moreover, we added three segments: black, grey and white for pixels where R=G=B or was not exactly equal according to their region (I, II, III). We assumed that 'not exactly equal' meant that $|R - G| < \Phi$ and $|R - B| < \Phi$, where $10 < \Phi < 15$. Having done this, we obtained images segmented into 12 clusters. We called this algorithm 'colour one'. Figure 4 presents the image shown in fig. 2 divided into 12 clusters using the above-described algorithm.

This approach is very fast because it uses only operations of comparison for each pixel, and not multiplication nor square as it is necessary for the C-means algorithm. Moreover, the "colour" algorithm is tenfold faster than the C-means algorithm. It should be pointed out that the bigger the image, the

greater the difference between the algorithms in terms of counting time.
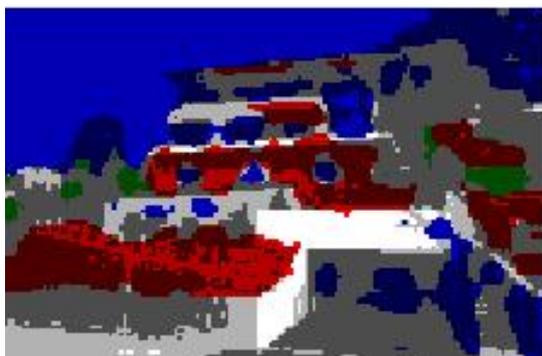

Fig. 4. 12 cluster segmentation of fig. 2 obtained by using the 'colour' algorithm.

4. OBJECT EXTRACTION ON THE BASE OF THE NEW ALGORITHM

Basing on this segmentation separate objects are obtained. We additionally made morphological operations such as close and open for each of 12 segments in order to receive objects from each one. It is worth to mention that the size of the structuring elements for these morphological operations is fitted automatically according to the size of image.

Fig. 5. Objects from fig. 2 presented in their average colours.

After extracting separate elements, the following features for all objects were counted: average colour, texture parameters, region-based shape descriptors. contour based shape descriptors, location in the image as a region-based representation. Figure 5 is an example of one of the above mentioned features; all elements are presented in their average colour.

5 THE DETERMINATION OF TEXTURE PARAMETERS

The texture information presented in images is one of the most powerful additional tools available. There are many methods which can be used for texture characterization. Unfortunately, they are mostly useless for our purpose.

Fig. 6. Example of an original image where the roof is a textured surface

One of them is the two-dimensional frequency transformation. For our purposes we could apply as well the classical Fourier transformation as several spatial-domain texture-sensitive operators, for instance, the Laplacian 3x3 or 5x5, the Gaussian 5x5, Hurst, Haralick, or Frei and Chen (Russ, 1995). Regrettably, all of them are useful for relatively small neighbourhoods.
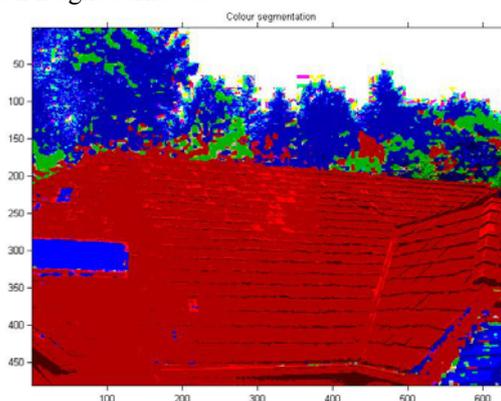

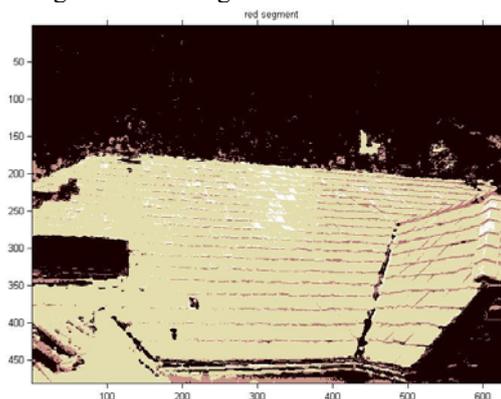Fig. 7. 12 clusters segmentation of fig. 7 obtained by using the 'colour' algorithm.

Fig. 8. The red segment (in three levels of brightness) extracted from the whole segmentation from fig. 7.

The next methods are based on the transformation domain algorithms. In 2001 Balmelli and Mojsilović (Balmelli, 2001) proposed the wavelet domain for texture and pattern. They have found features such as directionality, symmetry and regularity for regular textures, geometrical patterns and floral ornaments. Unfortunately, they have not proposed any

application of their method for real images. So far only Lewis and Fauzi manage to perform an automatic texture segmentation for CBIR based on discrete WT (DWT). They have applied their method for image retrieval in museum collections (Fauzi, 2006).

In our work we decided to use the Fast Wavelet Transform (FWT). It is efficient and productive enough for frequent use for our purpose.

One of the most important features of wavelet details is their directionality. It means that the convolution of the horizontal wavelet and the horizontal elements of image results in big values of horizontal details d_1^1 (called by Matlab as cH1 coefficients), whereas the convolution of the vertical wavelet and the vertical elements of image results in big values of vertical details d_1^2 (called cV1 coefficients). If we use this feature and compute the convolution of an image consisting of regular tiles or bricks and relevant wavelet, we obtain a 2D transform whose maximum values are placed in the connection spots among these tiles or bricks. We have applied the Haar wavelet for all our counts because it is the simplest wavelet and it seems to be the most suitable kind of wavelet for analysis of geometrical elements.
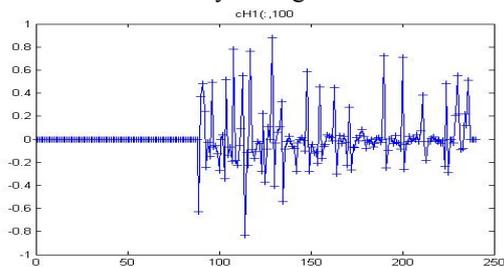

Fig. 9. Horizontal wavelet coefficients presented along the 100th column of the image transform (for the Haar wavelet, where $j = 1$). Numbers of the Haar wavelets for the first level of multiresolution analysis are on the horizontal axis and values of coefficients cH1 are on vertical axis.

Therefore, we have applied the Haar wavelet to the roof region shown in fig. 8. Then, we obtained three matrices of details d_1^1, d_1^2 and d_1^3. The cross-section through the 100th column of the horizontal details matrix d_1^1 (cH1) is presented in figure 10. Maxima and minima in this figure are equivalent to connections between tiles in fig. 7. Having computed horizontal details, we have measured distances between maxima for each column of this matrix and we have measured distances between minima for each column of this matrix (shown in fig. 10). We have located one threshold on the level of 1% of the maximum value of the whole matrix and we have measured distances between positive coefficients on that level. Analogically, we have set the next threshold on the level of 1% of the minimum value of the whole matrix and we have measured distances between negative coefficients on that level. It has turned out that these distances which are equivalent

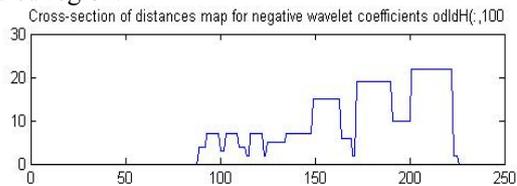to the size of tiles are good distinctive parameters for textured region.

Fig. 10. Cross-section through the 100th column of the distances map for negative horizontal wavelet coefficients. Numbers of the Haar wavelets for the first level of multiresolution analysis are on the horizontal axis and distances between the maximal negative wavelet coefficients are on vertical axis.

After counting the distances we have created two distance maps for all positive and negative horizontal coefficients. Figure 11 presents one of these distance maps. Analogical procedure has been carried out for vertical wavelet coefficients cV1. Distance map for negative vertical wavelet coefficients is presented in figure 12. We have to remember that wavelet transforms are counted by subsampling, therefore, the distance maps are half smaller than the original image.
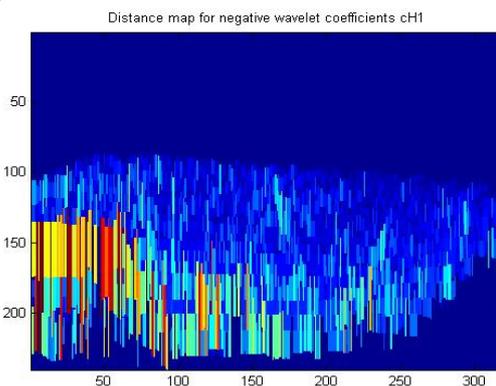

Fig. 11. Distance map for negative horizontal wavelet coefficients cH1.

Basing on the above distance maps we can estimate that the vertical size of tiles is equal to 15-20 wavelet lengths which means 30-40 pixels and horizontal size of tiles is equal to 22-40 wavelet lengths which means 44-80 pixels.
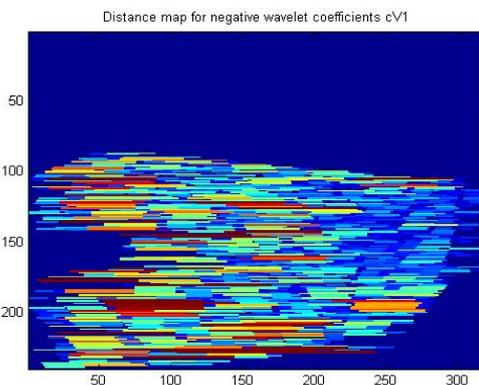

Fig. 12. Distance map for negative vertical wavelet coefficients cV1.

6. CONCLUSIONS AND FURTHER WORKS

To sum up, this paper shows how to extract elements from images in the unsupervised way and analyze objects parameters. We have focused on the description of texture parameters because it was the most difficult task.

The achieved results indicate that it is possible to separate objects in the image with acceptable accuracy for further interpretation in the unsupervised way. In computer terms, objects are recognized by finding the above-mentioned features of each object and a new object is classified to one of the previous created classes. So far, we have no interpretation which of these objects are doors, windows, etc. At present, the database structure is being prepared. This structure will cover all elements necessary for image content analysis; namely basic object features as well as logical and spatial relations.

In the future works we plan to use the case-based reasoning method for object interpretation in human terms. Logical features and spatial relations will then be used for this purpose.

REFERENCES

Amsaleg, L., Gros, P., Berrani, S-A. (2004). Robust Object Recognition in Images and the Related Database Problems, *Multimedia Tools and Applications*, Vol. 23, No. 1, Springer Netherlands, pp. 221-235.

Balmelli, L., Mojsilović, A. (2001). Wavelet domain features for texture/pattern description, classification and replicability analysis. In: A. Petrosian and F. Meyer, ed., *Wavelets in Signal and Image Analysis*. Kluwer Academic Publishers, London, Chap. 7, pp. 193-212.

Bamidele, A., Stentiford, F. W., Morphett, J. (2004). An attention-based approach to content-based image retrieval, *BT Technology Journal*, Vol. 22, No. 3, pp. 151-160.

Castelli, V., Bergman, L. D., Kontoyiannis I., et al. (1998). Progressive search and retrieval in large image archives, *IBMJ Research Development*, Vol. 42, No. 2, pp. 253-268.

Chow, T., Rahman, M., Wu, S. (2006). Content-based image retrieval by using tree-structured features and multi-layer self-organizing map, *Pattern Analysis and Applications,* Vol. 9, No 1, Springer-Verlag, London, pp. 1-20.

Fauqueur, J., Boujemaa, N. (2006). *Mental image search by Boolean composition of region categories, Multimedia Tools and Applications,* Vol. 31, No. 1 Springer Netherlands, pp. 95-117.

Fauzi, M., Lewis, P. (2006). Automatic texture segmenta-tion for content-based image retrieval application*, Pattern Analysis and Applications*, Springer-Verlag, London, (in printing).

Flickner, M., Sawhney, H., et al. (1995). Query by Image and Video Content: The QBIC System, *IEEE Computer*, Vol. 28, No. 9, pp. 23-32.

Gao Y. Y., Zhang Y. J., Yu F. (2000) *Semantic-based image description model and its implementation for image retrieval*. Proceeding of the First International Conference on Image and Graphics, pp. 657-660.

Głomb, P. (2006). Metody indeksowanie ruchomych obrazów wideo, *Ph.D. Dissertation*, Institute of Theoretical and Applied Informatics PAS, Gliwice.

Hong, D. Z., Wu, J. K., Singh, S. S. (1999). Refining image retrieval based on context-driven method, *SPIE Proceedings on Storage and Retrieval for Image and Video Database VII,* Vol. 3656, pp. 581-593.

Hsu, W., Chua, T. S., Pung, H. K., 2000. Approximation Content-based Object-Level Image Retrieval*, Multimedia Tools and Applications*, Vol. 12, Springer Netherlands, pp. 59-79.

Jaimes, A., Chang, S. F. (1999). Model-based classification of visual information for content-based retrieval, *SPIE Proceedings on Storage and Retrieval for Image and Video Database VII,* Vol. 3656, pp. 402-414.

Jaworska, T., Partyka, A. (2005). Research: Content-based image retrieval system [in Polish], *Report RB/37/2005*, Systems Research Institute, PAS.

Mokhtarian, F., Abbasi, S., Kittler J. (1996). Robust and Efficient Shape Indexing through Curvature Scale Space, *Proc. British Machine Vision Conference*, pp. 53-62.

Niblack, W., Flickner, M., et al. (1993). The QBIC Project: Querying Images by Content Using Colour, Texture and Shape, *SPIE*, Vol. 1908, pp. 173-187.

Ogle, V., Stonebraker, M. (1995). CHABOT: Retrieval from a Relational Database of Images, *IEEE Computer*, Vol. 28, No 9, pp. 40-48.

Russ, J. C., 1995 The image processing. Handbook, *CRC*, London, pp. 361-385.

Rubner, Y., Tomasi, C., Guibas, L. (2000). The Earth Mover's distance as a Metric for Imager Retrieval, *International Journal of Computer Vision,* Vol. 40, No 2, Springer Netherlands, pp. 99-121.

Song, Y., Zhang, A. (2003). Analyzing scenery images by monotonic tree, *Multimedia Systems*, Vol. 8, Springer-Verlag, London, pp. 495-511.

Stanchev, P., 2003. Using Image Mining for Image Retrieval, In *Proceeding of IASTED Conference, Cancun, Mexico*, pp. 214-218.

Tan, K-L., Ooi, B. Ch., Yee, Ch. Y. (2001). An Evaluation of Color-Spatial Retrieval Techniques for Large Image Databases, *Multimedia Tools and Applications*, Vol. 14, Springer Netherlands, pp. 55-78.

Wang, T., Rui, Y., Sun, J.-G. (2004). Constraint Based Region Matching for Image Retrieval, *International Journal of Computer Vision*, Vol. 56, No 1/2, pp. 37-45.