Adaptive Hierarchical Density Histogram for Complex Binary Image Retrieval

Panagiotis Sidiropoulos, Stefanos Vrochidis, Ioannis Kompatsiaris
Informatics and Telematics Institute
6th Km Charilaou-Thermi Road, Thessaloniki, Greece
psid@iti.gr stefanos@iti.gr ikom@iti.gr

Abstract

This paper proposes a novel binary image descriptor, namely the Adaptive Hierarchical Density Histogram, that can be utilized for complex binary image retrieval. This novel descriptor exploits the distribution of the image points on a two-dimensional area. To reflect effectively this distribution, we propose an adaptive pyramidal decomposition of the image into non-overlapping rectangular regions and the extraction of the density histogram of each region. This hierarchical decomposition algorithm is based on the recursive calculation of geometric centroids. The presented technique is experimentally shown to combine efficient performance, low computational cost and scalability. Comparison with other prevailing approaches demonstrates its high potential.

1 Introduction

Despite the significant advances in colour image processing, still many image databases consist of binary (i.e. black-and-white) images, including trademarks, patent images, technical drawings, or other specific applications like road signs, botanical collections or medical images. Binary images contain minimum colour and texture information, so they cannot be effectively described by general purpose content based algorithms, which highly depend on the aforementioned characteristics. So, while the employment of descriptors such as the known Scale Invariant Feature Transform (SIFT) [13] in binary image databases could be possible, their usage is rather limited. Instead, content-based binary image analysis techniques are expected to include a shape-based feature vector extraction method [2]–[4], [7]–[8], i.e. one that aim to describe the image geometric information accurately. A generic case content-based binary image retrieval scheme has to deal with the unique characteristics of binary images. More specifically, such images (e.g. technical drawings, patent images, etc) originate from a noisy analog sketch of various drawing styles that was exposed to arbitrary time degradation, as well as to the content degrading procedure of digitising.

In the image processing domain, several shape-based binary image processing techniques have been proposed. In [7]–[8], a known approach for binary images content-based retrieval is presented, that recognises line-patterns, in order to produce attributed graphs or the histogram of local attributes. In [10]-[11], the binary image is considered as a set of line segments. The lengths and orientations of triples of adjacent line segments are quantised into pre-defined codewords and the resulting histogram is used as binary image descriptor. Variations of the method has been proposed for object detection [11], shape retrieval and classification [10]. In a recent work [12], the combination of local and global features is proposed in order to capture both the essence and details of a binary shape. In another approach, a two-step algorithm is proposed to retrieve relevant images from a binary trademark database [2]–[3], which included the computation and quantisation of image edge angles, thus constructing the Edge Direction Histogram [3] (EDH). This descriptor was combined with Hu's invariant moments [1] and template matching in a sophisticated and computationally expensive scheme that cannot be easily generalised. Recently, a large number of built-upon EDH features have been proposed. The most prominent is the edge orientation autocorrelogram, (EOAC) [5], which has been introduced for general purpose CBIR. Edge orientation autocorrelogram is a two dimensional histogram, in which the (j, k) element indicates the number of similar edges with the $j - th$ orientation that lie k pixel distance apart. EOAC and edge-based techniques in general, inherently fail to achieve independence from drawing style. This is because the edge directions can not be accurately computed in cases of very thin or very thick line drawings, small isolated blobs drawings and filled-in figures [3]. Nevertheless, the edge orientation autocorrelogram (EOAC) has been compared with many shape-based retrieval features and revealed the scheme's superiority [5].

This paper introduces a novel retrieval technique for binary image retrieval, based on a new feature called the

978-1-4244-8027-2/10/$26.00 ©2010 IEEE

Adaptive Hierarchical Density Histogram (AHDH). The approach is inspired by the adaptive pyramidal decomposition of the image into regions based on the recursive calculation of geometric centroids proposed in [5] in order to produce an efficient binary image descriptor by generating the density histogram of each region. As the technique is based on the geometric characteristics of the image by exploiting the distribution of the image points on a two-dimensional area, it shows robustness against drawing style variation and moderate additional noise. In addition, taking into account the small feature vector size and the fact that no segmentation process is required, the algorithm is proven to be computationally inexpensive and consequently it is capable of dealing with large binary image databases.

The rest of the paper is structured as follows: In section 2 the algorithm for the adaptive pyramidical decomposition of an image [6] is summarised. The algorithm for the generation of the Adaptive Hierarchical Density Histogram is presented and analysed in section 3. In section 4 a set of experiments in a patent image database is conducted, which include comparisons that demonstrate the improved performance of the proposed method over the well-established methods of EOAC [5] and [6]. Finally, section 5 concludes this work.

2 Adaptive Hierarchical Geometrical Centroids Partitioning

In [6] Yang et al. have presented an adaptive pyramidal image partitioning scheme in order to generate an image descriptor that was composed of the calculated geometric centroids. In an iterative process, the initial image is decomposed into non-overlapping regions in a recursive way. In this algorithm region partitioning is performed based on the extraction of the geometric centroids. In each iteration (or level of decomposition) l, a number of 4^{l-1} regions R_i^l have to be processed. Considering a separate cartesian coordinates system for each region R_i^l, the coordinates of the respective geometrical centroid $c_R = (x_c, x_c)$ is given by [6]:

$$x_c = \frac{\sum\limits_{(x,y)\in B_i^l} x}{N_i^l}, \quad y_c = \frac{\sum\limits_{(x,y)\in B_i^l} y}{N_i^l} \tag{1}$$

where N_i^l denotes the amount of black-pixels set B_i^l in the processed region R_i^l, and the illuminance of a binary image pixel with coordinates (x, y), can take only values 0 (for pixel that belong to the background, or 'white' pixels) or 1 (for pixel that belong to the foreground, or 'black' pixels). These centroids partition the image plane in an adaptive hierarchical biased orthogonal grid (figure 1).

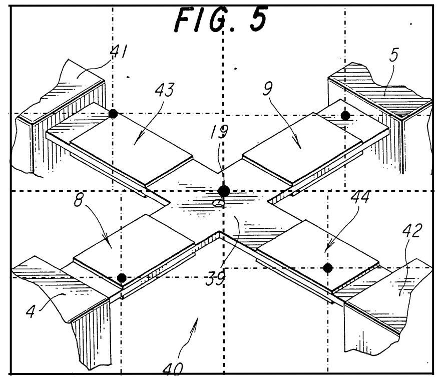

Figure 1. A binary patent image and its first and second level centroids (marked with dots) and partitions.

3 Adaptive Hierarchical Density Histogram (AHDH)

In this section a novel feature, namely the Adaptive Hierarchical Density Histogram, is introduced. The algorithm for the generation of AHDH consists of two main parts: a) the Region Partitioning based on the generation of the adaptive geometric centroids as discussed in previous section and b) the Adaptive Hierarchical Density Histogram generation, which is the main contribution of this work. A schematic view of the algorithm is presented in figure 2.

In contrary to the most shape-based techniques, in which a binary image is considered as a complex geometrical shape or a set of simple geometrical primitives, the orientations and relative positions of which would produce a fair description of its topological structure, in our approach the binary image is considered as a two dimensional plane. We define as B the set of black pixels that comprise the schematic diagram on the plane, while N is the cardinality of the elements of B. The coordinates of these pixels are normalised in order to be translation invariant and in addition a simple noise reduction procedure is performed during a pre-processing stage. Regions are produced iteratively following the technique that was introduced in [6] and summarised in section 2, leading to the construction of an adaptive asymmetric orthogonal grid, which covers the entire black and white image (figure 1).

For each iteration or level l, Region Partitioning is performed by estimating the geometric centroid of all regions R_i^l formulated so far and then splitting each region into 4 sub-regions $SR_{i,j}^l$, $j = \{1, 2, 3, 4\}$ using as center the geometric centroid. The initial region is the whole image and if

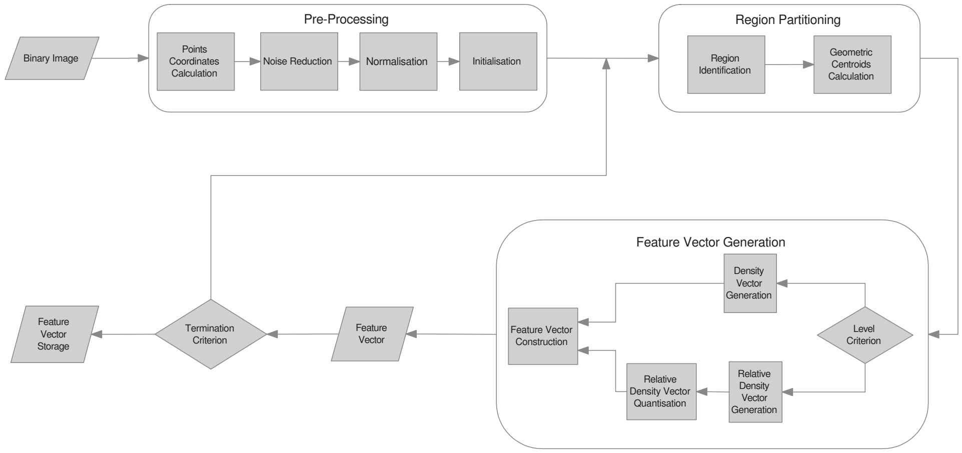

Figure 2. Overview of the presented algorithm. Initially translation invariance and noise reduction is achieved through a simple pre-processing step. Subsequently, the binary image is segmented into smaller regions with the iterative procedure of Region Partioning. Then, the novel feature is extracted in the Feature Vector Generation part, which is the main contribution of this work. This iterative procedure is terminated when a manually selected criterion is satisfied.

$l > 1$ then for all regions R_i^l there is a sub-region SR_{i_1,j_1}^{l-1} such that $R_i^l \equiv SR_{i_1,j_1}^{l-1}$.This iterative procedure is continued until a termination criterion is satisfied.

The procedure of feature extraction depends on the level value; in 'lower' levels, the feature is simply a vector of the distribution of the N_i^l black pixels into the 4 sub-regions, while in 'higher' levels a two-classes classification of each sub-region is employed. The two classes are labeled as 'Full' and 'Empty' and are defined by the percentage of N_i^l black pixels that lie into the sub-region in comparison to the percentage of region's area E_i^l that belong to the sub-region. The combination of the classes of all 4 sub-regions of a region produces a 'distribution word'. Finally, the 'distribution word' histogram is used as the levels feature. As the region partitioning step generates an adaptive hierarchical orthogonal partition and the feature extraction involves the estimation of each region's density the proposed generated feature is named 'Adaptive Hierarchical Density Histogram'.

3.1 Adaptive Hierarchical Density Histogram Generation

The most important part of the algorithm is the generation of the novel feature AHDH. We consider two different approaches for the construction of the feature vector: a) density features and b) quantised relative density features. For all levels l less than a threshold l_d density features are estimated, while if $l \geq l_d$ quantised relative density features are computed. Threshold l_d is experimentally defined. The features are estimated accordingly, for all regions at each level and the overall feature vector is updated at each iteration of the algorithm.

Yang et al. proposed a feature vector wich consists of the centroids of the three first iterations [6]. The number of the centroids is an exponentially increasing function of the amount of levels, and in order to keep the dimensionality of the feature vector low, the centroids can be estimated only for the first levels. This undermines the potential of the approach, since most of the image local geometrical information lies in the deepest levels. Furthermore, it seems that the distribution of region's black pixels conveys more information than the geometric centroid. For instance, in figure 1 the fact that in the second level partition the 'first line, second column' sub-region is almost empty of black pixels is usually much more useful than the exact knowledge of the centroid of the region that this sub-region belongs. Consequently, we propose the employment of a density vector, instead of Yang's centroid vector.

In $l-th$ level (i.e. after l iterations), we have identified 4^{l-1} regions R_i^l with area E_i^l and N_i^l black pixels. For each of these regions, the centroid estimation results to a partition into 4 new sub-regions $SR_{i,j}^l$, $j = \{1,2,3,4\}$ with area $E_{i,j}^l$ and $N_{i,j}^l$ black pixels. It is obvious that $\sum_{j=1}^{4} N_{i,j}^l = N_i^l$ and $\sum_{j=1}^{4} E_{i,j}^l = E_i^l$. Two variables are defined based on these: the density $d_{i,j}^l$ of a sub-region $SR_{i,j}^l$ of a region R_i^l and the relative density $\hat{d}_{i,j}^l$. The density $d_{i,j}^l$ is defined as the amount of the black pixels of a sub-region $SR^l_{i,j}$ divided by N^l_i, while the relative density $\hat{d}^l_{i,j}$ is the sub-region density compared to the ratio of the sub-region area $E^l_{i,j}$ over the region area E^l_i:

$$d^l_{i,j} = \frac{N^l_{i,j}}{N^l_i}, \qquad \hat{d}^l_{i,j} = \frac{E^l_i N^l_{i,j}}{N^l_i E^l_{i,j}} \tag{2}$$

where $1 \leq i \leq 4^{l-1}$ and $1 \leq j \leq 4$. Subsequently an $4^{l-1} \times 4$ feature array FA_l is constructed either by the densities or the relative densities of each new sub-region.

$$FA_l = \begin{vmatrix} d^l_{1,1} & d^l_{1,2} & d^l_{1,3} & d^l_{1,4} \\ d^l_{2,1} & d^l_{2,2} & d^l_{2,3} & d^l_{2,4} \\ .... & .... & .... & .... \\ d^l_{4^{l-1},1} & d^l_{4^{l-1},2} & d^l_{4^{l-1},3} & d^l_{4^{l-1},4} \end{vmatrix} \tag{3}$$

Then, the FA_l is serialised to form the l-level feature FV_l:

$$FV_l = \left\{ d^l_{1,1}, d^l_{1,2}, d^l_{1,3}, d^l_{1,4}, .., d^l_{4^{l-1},3}, d^l_{4^{l-1},4} \right\} \tag{4}$$

In the same way, we construct the feature $\hat{FV}_l$, which involves $\hat{d}^l_{i,j}$ instead of $d^l_{i,j}$:

$$\hat{FV}_l = \left\{ \hat{d}^l_{1,1}, \hat{d}^l_{1,2}, \hat{d}^l_{1,3}, \hat{d}^l_{1,4}, .., \hat{d}^l_{4^{l-1},3}, \hat{d}^l_{4^{l-1},4} \right\} \tag{5}$$

The employment of density or relative density features depends on the partition level. For lower levels, the density vector is used. The feature of these levels is simply the serialised density feature of Equation (4). The dimension of the density feature increases exponentially with the number of levels. On the other hand, as the higher level triggers an image decomposition in exponentially smaller parts, the information that each feature element contains is rapidly decreasing. To overcome those inherent deficiencies, the relative density feature is employed in the higher levels (Equation (5)). Hence, when the number of iteration, exceeds an experimentally evaluated threshold, then relative density instead of density is computed in the algorithm.

3.2 Relative Density Vector Quantisation

Unlike to centroid vectors, density and relative density vectors can easily be quantised. In this work we have chosen to quantise only the relative density, and not the density, vectors due to the intuitionally straightforward extraction of the between-classes boundary that they provide. A sub-region $SR^l_{i,j}$, for which $\hat{d}^l_{i,j} \geq 1$, is labeled 'Full', else it is labelled 'Empty'. The rationale behind the use of the Empty and Full labels can be statistically explained. If the black pixels of a region R^l_i would fall in the four sub-regions following a uniform random distribution, the expected value of relative density would be equal to 1 for every sub-region. Consequently, a sub-region is labeled Full or Empty depending on the amount of pixels that lie in its interior compared to the statistically expected value. On the contrary, density vectors quantisation depend on the selection of an arbitrate between-classes threshold.

At this stage, we introduce a Lexicon L of 'distribution words' w, which represents the 16 combinations of the 4 Full or Empty sub-regions of a processed region. Defining that E corresponds to Empty and F to Full, the lexicon L has the following format:

$$L = [EEEE, EEEF, ..., FFFF] = [w_0, w_1, ..., w_{15}] \tag{6}$$

However, the valid words that are used are actually 15. The 16-th, non valid word is the 'EEEE' or w_0, since, by the definition of the relative density the amount of black pixels of a sub-region can not be less than the expected value for all four sub-regions of any region. Based on this lexicon, the quantised feature for each level is:

$$\hat{FV}_{q,l} = [h(w_1)^l, h(w_2)^l, ..., h(w_{15})^l] \tag{7}$$

where $h(w_i)^l$ is estimated by counting the number of appearances of the respective word, normalised over the total number of the level l sub-regions (i.e $h(w_1)^5$ is the histogram value of the word EEEF in the $5-th$ level).

Finally, the new constructed feature is superimposed to the feature vector that was generated during the previous iterations:

$$FV = [FV_1 \, FV_2 \, ...FV_{l_d - 1} \, \hat{FV}_{q,l_d} \, \hat{FV}_{q,l_d+1} \, ...FV_{q,l}]$$

where l_d is the first level for which quantised relative density features are extracted. After the algorithm is terminated, the final format of FV represents the Adaptive Hierarchical Density Histogram of the image and can be used for retrieval purposes.

It must be mentioned that relative density vectors are employed and quantised only to allow us to reach deepest levels, where local information about the black pixels topological structure exists. In the first levels, this technique, is not only unnecessary but also inefficient, since it is obvious that the global distribution of black pixels is much better represented by the density vector instead of the knowledge that its relative density vector belongs to a certain class.

4 Experimental Results

In this section the experimental results for the evaluation of the proposed algorithm are presented. The potential of the proposed technique is demonstrated through comparison with: (i) the edge orientation autocorrelogram

(EOAC), which is considered one of the most prevalent approach in this domain, outperforming most of the non-segmentation based retrieval methods [5], and (ii) Yang's technique, which was the basis for our approach, utilising a 3-level centroid vector [6]. The experiments were conducted in a database of complex binary images, extracted from patent documents from European Patent Office.

More specifically, in the experimental framework we employed a database[1] that includes 2000 binary patent images. The query base was created by 120 randomly selected binary patent images with 2 to 73 relevant images in the database. The annotation of the database was performed with the cooperation of patent searchers in the context of the EU project PATExpert. The AHDH method employed a constant termination level, which was experimentally tuned to $l = 10$. The first level for which quantised relative density features are extracted were manually set to $l_d = 3$. EOAC and centroid vector algorithms were executed with the parameters that [5] and [9] have applied. In all methods, L1 distance was utilised as a similarity measure.

The Precision-Recall curves are illustrated in figure 3. It can be observed that for identical Recall rates, AHDH Precision rate is at least 20-40% higher than the other implemented techniques, while for identical Precision rates, Recall rate is 5-25% better. In the case that Recall and Precision values are equal, AHDH would lead to a mutual enhancement of at least 14.7% both for Recall and Precision values. The above comparison indicates the improved performance of AHDH over these prevalent binary image descriptors. In order to test the aforementioned techniques in terms of time response and scalability, we employed a database of 10000 images, also extracted from patent documents from European Patent Office. The mean time-responses of the system for a queries by visual example using AHDH, Yang's and EOAC technique were 9.1, 8.5 and 90 seconds respectively.

Additionally, a query-based experiment, in which each image example of the first database (i.e. 2000 patent images) was associated with the 25 most similar images retrieved for each query, was conducted. This is the case in many retrieval systems, where a user is assumed to decide about the number of the results to be retrieved. Images with more than 25 similar images where excluded from the query database, leading to a query base of 96 images at total. For this experiment, 86.9% of the manually annotated near-replicas or similar images were successfully retrieved. Two different query images and the first retrieved images, which involve visual search for technical drawings of circular shape and flowcharts respectively are shown in figure 4. On the other hand, figure 5 depicts a retrieval example, in which the algorithm fails to produce quality results. In this case, the example patent image, which illustrates an item

[1] Publicly available at http://mklab.iti.gr/content/patent-database

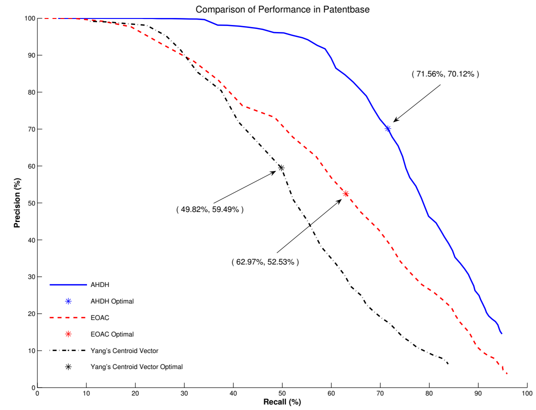

Figure 3. Precision - Recall diagram of AHDH, EOAC and centroid vector for patent binary images.

rotated by different angles and the results contain only one similar depiction and cyclic objects. This is due to the fact that the proposed method is rotation variant and fails to retrieve similar images under different orientations.

From the aforementioned experimental results, it can be seen that Adaptive Hierarchical Density Histograms outperforms EOAC in terms of precision, recall and time response. Furthermore, the proposed descriptor is equivalent with Geometrical Adaptive Hierarchical Centroids in terms of computational complexity, however it demonstrates significantly better recall and precision rates.

It should be mentioned that, similarly to most shape descriptors including Edge Orientation Autocorrelogram, Adaptive Hierarchical Density Histogram is experimentally found to be rotation variant for high rotation angles. However, the small feature vector size and the relatively low computational cost that is required to extract the descriptor can allow a simple countermeasure by creating multiple feature vectors for variant image orientations.

5 Conclusions and Future Work

In this paper the novel feature namely Adaptive Hierarchical Density Histograms for binary image retrieval was introduced. Evaluation of the performance of the algorithm in terms of recall and precision as well as of time response and comparison with other important techniques in the field, showed the potential of the proposed retrieval method. On the other hand, the major drawback of the AHDH is the absence of inherent geometrical invariance, which can be considered as the main restraining factor of its performance. Future work will deal with further research that would pro-

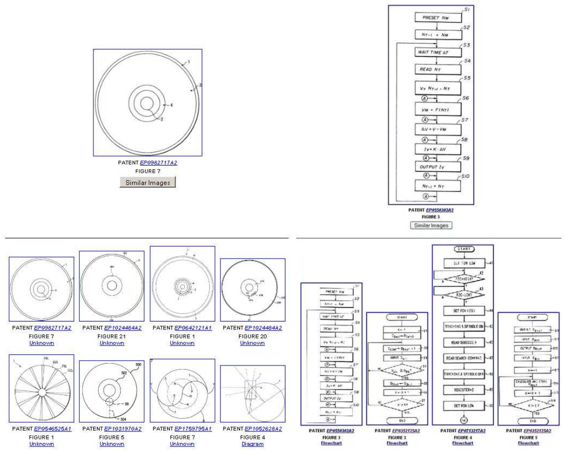

Figure 4. Two queries by visual example of patent images and the first retrieved results.

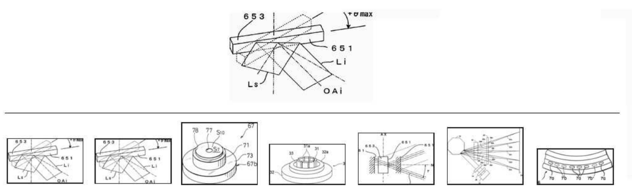

Figure 5. A patent image example that fails to produce quality results.

vide a geometrical invariant variation of the point-density histogram.

Acknowledgements

This work was supported by the European project PAT-Expert (FP6-028116), as well as by the project CHORUS+ (FP7-249008), both funded by the European Commission.

References

[1] M. Hu, Visual pattern recognition by moment invariants, IRE Transactions of Information Theory 8 (1962) 179-187.

[2] A.K. Jain and A. Vailaya, Image retrieval using color and shape, Pattern Recognition, 29 (1996) 1233-1244.

[3] A.K. Jain and A. Vailaya, Shape-based retrieval: a case study with trademark image database, Pattern Recognition 31 (1998) 1369-1390.

[4] J. Poel, C. de Almeida, L. Batista, A new multi-scale, curvature-based shape representation technique for image retrieval based on dsp techniques, In: Proceedings of the Fifth International Conference on Hybrid Intelligent Systems, Rio de Janeiro, Brazil, 2005, pp. 373-378.

[5] F. Mahmoudi, J. Shanbehzadeh, A.-M. Eftekhari-Moghadam, H. Soltanian-Zadeh, Image retrieval based on shape similarity by edge orientation autocorrelogram, Pattern Recognition 36 (2003) 1725-1736.

[6] M. Yang, G. Qiu, Y. Huang, D. Elliman, Near-Duplicate Image Recognition and Content-based Image Retrieval using Adaptive Hierarchical Geometric Centroids, In: Proceedings of the 18th International Conference on Pattern Recognition (ICPR 2006), Hong Kong, China, 2006, pp. 958-961.

[7] B. Huet, N. J. Kern, G. Guarascio and B. Merialdo, Relational Skeletons For Retrieval In Patent Drawings, In: Proceedings of International Conference in Image Processing (ICIP 2001), Thessaloniki, Greece, 2001, pp. 737-740.

[8] B. Huet and E. R. Hancock, Fuzzy relational distance for large-scale object recognition, In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR 1998), Santa Barbara, California, USA, 1998, pp. 138-143.

[9] A. Tiwari, V. Bansal, PATSEEK: Content Based Image Retrieval System for Patent Database, In: Proceedings of International Conference on Electronic Business, Beijing, China, 2004, pp. 1167-1171.

[10] G. Zhu, X. Yu, Y. Li, D. Doermann, Learning visual shape lexicon for document image content recognition, In: Proceedings of the 10th European Conference on Computer Vision, Marseille, France, 2008, pp. 745-758.

[11] X. Yu, Y. Li, C. Fermuller, D. Doermann, Object detection using a shape codebook, In: British Machine Vision Conference, 2007.

[12] C.-H. Wei, , Y. Li, , W.-Y. Chau, and C.-T. Li, Trademark image retrieval using synthetic features for describing global shape and interior structure, Pattern Recognition 42 (2009) 386-394.

[13] D. G. Lowe, Distinctive Image Features from Scale-Invariant Keypoints, International Journal of Computer Vision 60 (2004) 91-110.