International Journal of Knowledge-based and Intelligent Engineering Systems 19 (2015) 27–33
DOI 10.3233/KES-150305
IOS Press

Automatic logo based document image retrieval

M.S. Shirdhonkar[a,*] and Manesh B. Kokare[b]
[a]*BLDEA's Dr. P.G.H. CET, Bijapur, Karnataka, India*
[b]*SGGS Institute of Technology and Engineering, Nanded, Maharashtra, India*

Abstract. Recently, there is need for the automatic logo detection and document image retrieval because of increasing requirements of intelligent document images. In this work, we have developed an automatic logo based document image retrieval system in which we first proposed an automatic logo detection and extraction from the document images using discrete wavelet transform (DWT). Then, we have proposed an approach to retrieve document images based on the logo using combined rotated complex wavelet filters (RCWF) and dual tree complex wavelet transform (DT-CWT). Since these combined features give information in twelve different directions. Our approach is segmentation free. Finally, we evaluate the effectiveness of our approach using large database collections of real-world complex documents.

Keywords: Document image retrieval, document similarity measurements, logo detection, wavelet transforms

1. Introduction

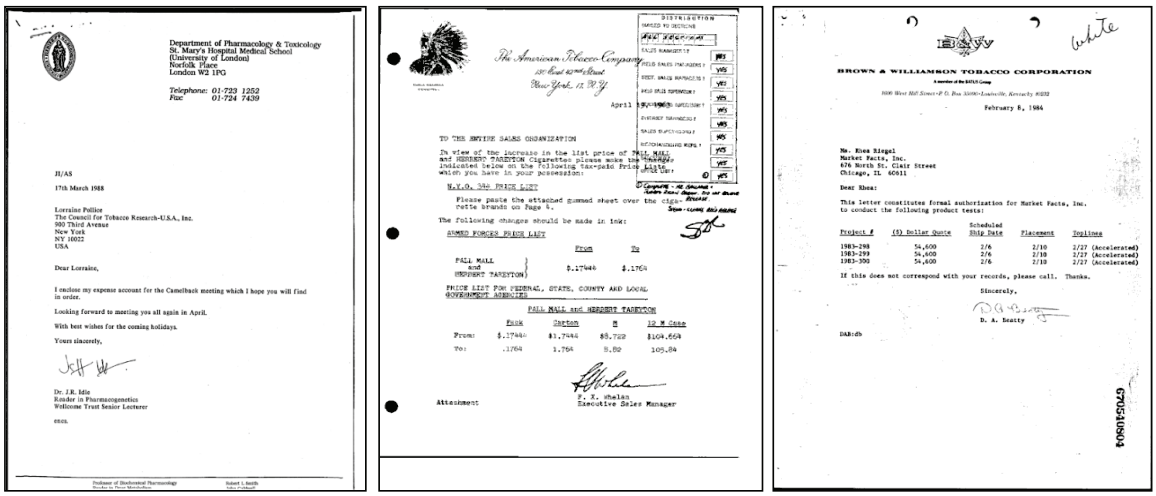
Graphics detection and recognition are fundamental research problems in document image analysis and retrieval. As one of the most pervasive graphical elements in business and government documents, logos may enable immediate identification of organizational entities and serve extensively as a declaration of a document's source and ownership. Complex documents present a great challenge to the field of document recognition and retrieval. The combined presence of noise, handwriting, signature, logos, machine-print with different fonts, and rule lines impose a lot of restrictions to algorithms that work relatively well on simple documents. The primary task of processing these complex documents is that of isolating different contents present in the document. Once the contents such as handwriting, machine-print, signature and logos are separated out, they can now be called as indexed documents which are ready to be used by a context-based image retrieval system. The problem of logo detection and recognition is of great interest to the document image analysis and retrieval communities because it enables immediate identification of the source of document based on the originating organization [1]. In the context of document image retrieval, logo provides an important form of indexing that enables effective exploration of data. The document image retrieval based on logo is the process of retrieving the closest matching document to the questioned documents from a database of known documents. Figure 1 shows examples of three different documents with different logos. The retrieval task would be to retrieve all the other documents with the same logo document. This would involve detecting logo from the documents and then performing a match on these documents.

The main contribution of this paper is that, firstly, we have proposed a novel logo detection and extraction from the document image using discrete wavelet transform. Secondly, document image retrieval based on logo using combined rotated complex wavelet filter and dual tree complex wavelet transform, in which document matching was performed using the Canberra distance. The experimental results of proposed method were satisfactory and give better results.

The rest of the paper is organized as follows. Section 2 discusses the literature survey. Section 3 dis-

*Corresponding author: M.S. Shirdhonkar, BLDEA's Dr. P.G.H. CET, Bijapur, Karnataka, India. E-mail: ms_shirdhonkar@rediffmail.com.

ISSN 1327-2314/15/$35.00



(a) Document 1 (b) Document 2 (c) Document 3

Fig. 1. Three different documents with logo from Tobacco-800 dataset.

cusses the feature extraction methods. In Section 4, the proposed system is discussed. In Section 5, the experimental results are presented and finally, Section 6 concludes the work.

2. Related work

The literature has focused almost on logo detection. In 1997, Seiden et al. [1], considered the binary classification problem of whether an image segment contained a logo. In their approach, the document page is first segmented using the top-down X-Y cut algorithm. A total of sixteen features of the connected components in each segment are extracted and used by a rule based classification scheme. Most of the research is focused only on logo recognition [2,3], where it is assumed that the segmentation of the logos has been done by a different module. In 2003, Pham [4] presented a simple logo detection method based on the assumption that the spatial density of the foreground pixels in a logo region is greater than that in non logo regions. A document image is first binaries into foreground and background pixels by global thresholding. Then, the spatial density within each fixed size window is computed and the region with the highest density is hypothesized as a logo region. In 2007, Zhu and Doermann [5], presented a multi-scale approach to logo detection and extraction in document images. A trained Fisher classifier performs initial classification at a coarse image scale. Each logo candidate region is further classified at successively finer scales by a cascade of simple classifiers. In 2009, Zhu and Doermann [6], developed an automatic logo based document retrieval system, that handles logo detection and segmentation by boosting a cascade of classifier across multiple image scales and logo matching using translation, scale and rotation invariant shape descriptors and matching algorithms. In 2012, Tivoli [7] proposed a framework for classify non-textual document image retrieval approaches and they evaluated the based on important measures such as appearance features, structural etc. In 2013 Keyvanpour and Tivoli [8] proposed a framework for classify document image retrieval approaches and they evaluated approaches based on important measures such as application type, appearance features, structural, language independent and cost.

3. Feature extraction methods

The major task of feature extraction is to reduce image data to much smaller in size which represents the important characteristics of the image. We propose the use of DT-CWT and DT-RCWF jointly, which captures the information in twelve different directions.

3.1. Discrete wavelet transform

The multi resolution wavelet transform decomposes a signal into low pass and high pass information. The low pass information represents a smoothed version and the main body of the original data. The high pass information represents data of sharper variations and

Fig. 2. Impulse response of 0°, 90° and ± 45° of DWT.

details. Discrete Wavelet Transform decomposes the image into four sub-images when one level of decomposing is used. One of these sub-images is a smoothed version of the original image corresponding to the low pass information and the other three ones are high pass information that represents the horizontal, vertical and diagonal edges of the image respectively. When two images are similar, their difference exists in high-frequency information. A DWT with N decomposition levels has $3N+1$ frequency bands with $3N$ high-frequency bands. The impulse response associated with 2-D discrete wavelet transform (see Fig. 2).

3.2. Dual tree complex wavelet transforms

The Drawbacks of the DWT are overcome by the complex wavelet transform (CWT), by introducing limited redundancy into the transform. But still it suffers from problem like no perfect reconstruction is possible using CWT decomposition beyond level 1, when input to each level becomes complex. To overcome this, Kingsbury [9] proposed a new transform, which provides perfect reconstruction along with providing the other advantages of complex wavelet, which is DT-CWT. The DT-CWT uses a dual tree of real part of wavelet transform instead of using complex coefficients. This introduces a limited amount of redundancy and provides perfect reconstruction along with providing the other advantages of complex wavelets. The DT-CWT is implemented using separable transforms and by combining sub-band signals appropriately. Even though it is non-separable, it inherits the computational efficiency of separable transforms. Specifically, the 1-D DT-CWT is implemented using two filter banks in parallel, operating on the same data. For d-dimensional input, a L scale DT-CWT outputs an array of real scaling coefficients corresponding to the low pass subbands in each dimension. The total redundancy of the transform is 2^d and independent of L. The mechanism of the DT-CWT is not covered here. See [9,10], for a comprehensive explanation of the transform and details of filter design for the trees. A complex valued $\psi(t)$ can be obtained as

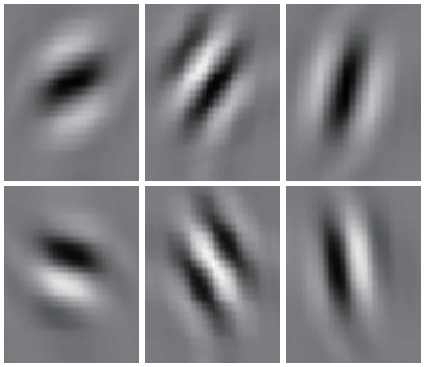


Fig. 3. Impulse response of six wavelet filters +15°, +45°, +75°, −15°, −45° and −75° of complex wavelet.

$$\psi(x) = \psi_h(x) + j\psi_g(x) \tag{1}$$

where $\psi_h(x)$ and $\psi_g(x)$ are both real-valued wavelets. The impulse responses of six wavelets associated with 2-D dual tree complex wavelet transform (see Fig. 3).

3.3. Dual tree rotated complex wavelet filters

Directional 2D RCWF are obtained by rotating the directional 2D DT-CWT filters by 45°, so that decomposition is performed along new direction, which are apart from decomposition 45° directions of CWT. The size of a filter is $(2N-1)\times(2N-1)$, where N is the length of 1-D filter. The decomposition of input image with 2-D RCWF followed by 2-D down sampling operation is performed up to the desired level. The computational complexity associated with RCWF decomposition is the same as that of standard 2-D CWT, if both are implemented in the 2-D frequency domain. The set of RCWFs retains the orthogonality property. The six subbands of 2D DT-RCWF gives information strongly oriented at (30°, 0°, −30°, 60°, 90°, 120°). The mechanism of the DT-RCWF is not covered here. See [11], for a comprehensive explanation of the transform and details of filter design for the trees. Thus, the 2D DT-CWT and RCWF provide us with more directional selectivity in the direction

$$\left\{\begin{array}{l}(+15^\circ,+45^\circ,+75^\circ,-15^\circ,-45^\circ,-75^\circ),\\(0^\circ,+30^\circ,+60^\circ,+90^\circ,120^\circ,-30^\circ)\end{array}\right\}$$

than the DWT whose directional sensitivity is in only three directions {0°, ± 45°, 90°}. The six wavelets associated with rotated complex wavelet filters (see Fig. 4).

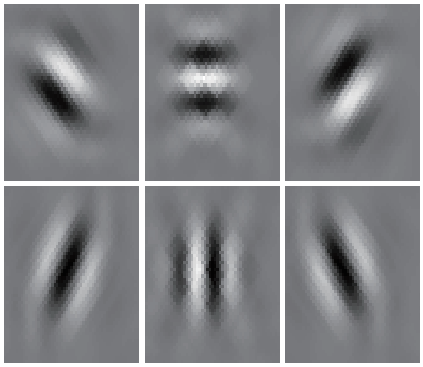


Fig. 4. Impulse response of $-30°$, $0°$, $+30°$, $+60°$, $90°$ and $120°$ of rotated complex wavelet filters.

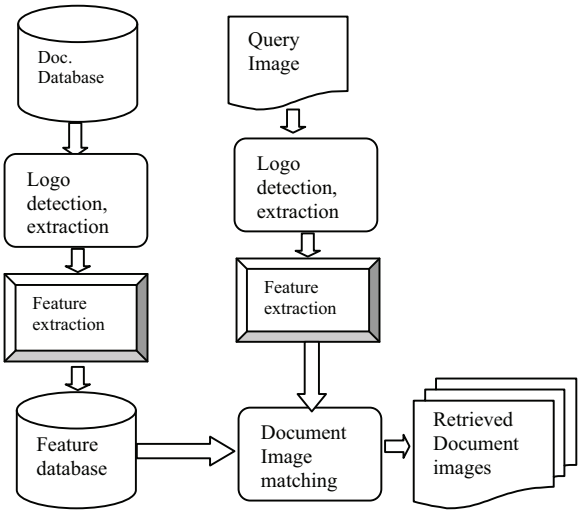


Fig. 5. System architecture for proposed system.

4. Proposed system

The objective of the proposed work is to detect and extract the logo from the document image and then retrieve the document image based on logo. The basic architecture of the system is as shown in Fig. 5. The performance of the proposed system can be tested by retrieving the desired number of the document images from the document image database. The average logo detection rate and document image retrieval rate is the main performance measures in the proposed system.

4.1. Logo detection and segmentation

Logos often appear as mixed text and graphics regions. In this section, we present a logo detection and extraction approach. The detection is formulated on the principle that spatial density of the foreground pixels within a given windowed image that contains a logo is greater than those of non-logo regions. We used discrete wavelet transform (DWT) to find the spatial density of the windowed image. Computing its spatial density as follows:

Let I be a document image of size $P \times Q$ and $w \in I$ a window of size mxn, which is chosen to be an approximation of a logo area. For each of the window w, we applied discrete wavelet transform to calculate the DWT coefficients. The number of windows w is calculated using $L = \mathit{round}(Q/n)$. The document image I is segmented into L windows w of size mxn. for each window w, we have computed the DWT coefficients, then calculated energy and standard deviation of these coefficients.

Let A_K and B_K are the energy and standard deviation vectors of the window w_k. Let C_k be the combined vector of these two vectors.

$$C_k = [A_K, B_K]\,. \tag{2}$$

Let $\delta(w_k)$ be density function of window w_k.

$$\delta(w_k) = \sum_k C_k \tag{3}$$

Finally, window W^* is detected as the region that contains a logo.

$$w_i^* = \arg\max_k \delta(w_k) \tag{4}$$

We also detected the second max value of $\delta(w)$. Let us denote it as w_j^*. Finally, we merge these two windows based on i and j value. If $i < j$, we have concatenated w_i^* with w_j^*, otherwise w_j^* with w_i^* window. The merged window we call it W^* as the region that contains logo. Figure 6 shows different categories of detected logos from the Tobacco-800 document image database [12]

4.2. Document image retrieval

In Section 4.1, we have detected and extracted the logo from document image by segmenting and extracting features using DWT. For document image retrieval based on logo we have used combined rotated complex wavelet filters and dual tree complex wavelet transform, which gives information in twelve different orientations. In the conduction the experiments, first, we detected and extracted the logo from each document image from the document database. Then, the extracted logo W^* is decomposed using DT-CWT

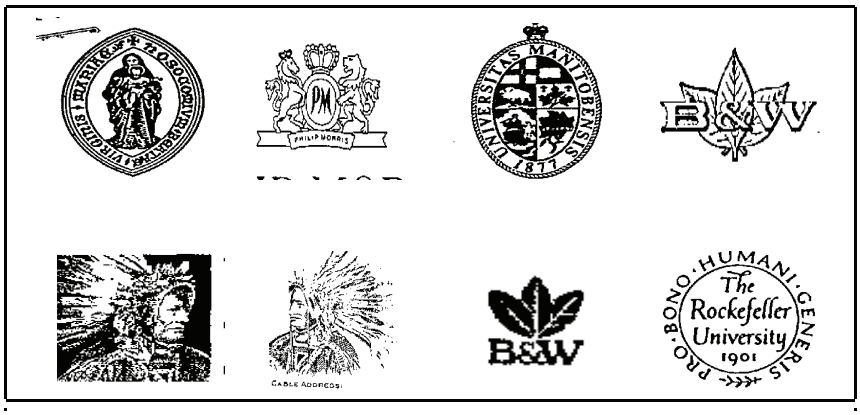


Fig. 6. Different categories of detected logos from the Tobacco-800 document image database.

and DT-RCWF up to third level. To construct the feature vectors of each logo of each document image in database, we decomposed each extracted logo from document image using DT-CWT and DT-RCWF up to third level. The energy and standard deviation were computed separately on each subband and the feature vector was formed using these two parameter values. The retrieval performance with combination of these two feature parameters always outperformed that using these features individually. The energy (E_k) and (δ_k) standard deviation of k^{th} sub-band is computed as follows

$$E_k = \frac{1}{M \times N} \sum_{i=1}^{M} \sum_{j=1}^{N} |W_k(i,j)| \tag{5}$$

$$\sigma_k = \left[\frac{1}{M \times N} \sum_{i=1}^{N} \sum_{j=1}^{M} (W_k(i,j) - \mu_k)^2 \right]^{\frac{1}{2}} \tag{6}$$

Where $W_k(i,j)$ is the k^{th} wavelet-decomposed subband, $M \times N$ is the size of wavelet decomposed subband, and μ_k is the mean of the k^{th} subband. The resulting feature vector using energy and standard deviation are $\bar{f}_E = [E_1 \;\; E_2 \ldots E_n]$ and $\bar{f}_\sigma = [\sigma_1 \;\; \sigma_2 \ldots \sigma_n]$ respectively. So combined feature vector of extracted logo is

$$\bar{f}_{\sigma\mu} = [\sigma_1 \;\; \sigma_2 \ldots \sigma_n \;\; E_1 \;\; E_2 \ldots E_n] \tag{7}$$

4.3. Document image matching

There are several ways to work out the distance between two points in multidimensional space. Here query is the document image, which is further processed to compute the feature vector as given in Section 4.1. The process involves two steps: first detection and extraction of the logo from query document image, then, computation of the features of extracted logo from query document image. The most commonly used is the Canberra distance measure. It can be considered the shortest distance between two points. If x and y are the feature vectors of the database and query image respectively, and have dimension d, then the Canberra distance is given by Eq. (8).

$$\mathit{Canb}(x,y) = \sum_{i=1}^{d} \frac{|x_i - y_i|}{|x_i| + |y_i|} \tag{8}$$

The average retrieval rate for the query document image is measured by counting the number of document images from the same category which are found in the top 'N' matches.

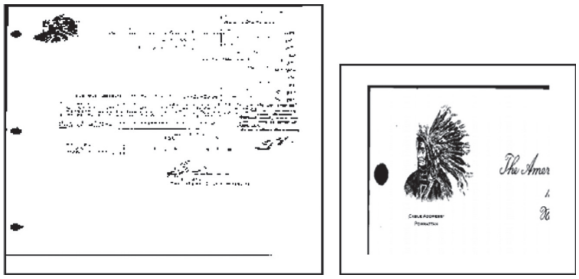
5. Experimental results

5.1. Image database

To evaluate system performance in automatic logo based document image retrieval, we used a large document collection, Tobacco-800 dataset. Tobacco-800 is a public subset of the IIT CDIP Test collection based on 42 million pages of documents obtained from UCSF [13] and released by Tobacco companies under the Master settlement Agreement. Tobacco-800 is a realistic dataset for document analysis and retrieval, as these documents were collected and scanned using a wide variety of equipment over time. We have tested in our system a total of 72 documents with logo from Tobacco-800 for detection and retrieval performance, in which there are eight different logo document images, among which the number of logos per class documents varies from 2 to 19.

Table 1
Detection rates of eight test models

Test category number	Detection rate using mountain function (%)	Detection rate using proposed approach (%)
1	75	75
2	78.9	89
3	50	75
4	08	69
5	50	100
6	64.7	100
7	75	100
8	100	100
Average detection rate (%)	62.7	88.5



(a) (b)

Fig. 7. Logo detection example (a) Original document image (b) After detection and extraction of logo.

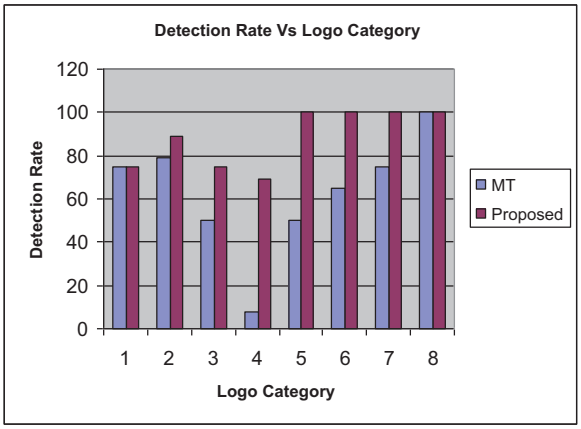


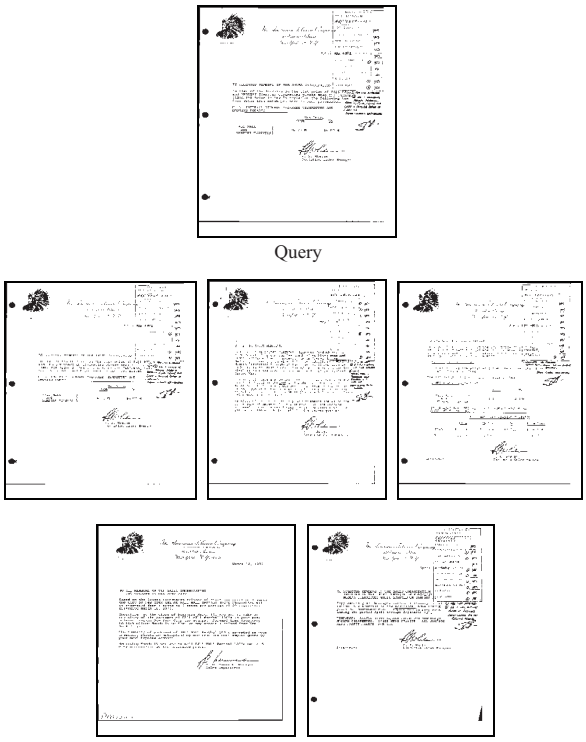
Fig. 8. Logo detection rate. (Colours are visible in the online version of the article; http://dx.doi.org/10.3233/KES-150305)

5.2. Retrieval performance

For each experiment, one image was selected at random as the query document image from each logo type document and thus retrieved images were obtained. In the following we compare two competing methods using a realistic and public data set. Calculating spatial density using mountain function (MT) described in [4].

Table 2
Average retrieval performance

Number of top matches	Precision (%)	Recall (%)
Top 1	100	22.22
Top 2	99.44	41.77
Top 3	77.77	44.88
Top 4	69.44	51.33
Top 5	64.44	56.11



Query

Fig. 9. List of the five most similar retrieved document images from the database.

It is based on the number of foreground and background pixels, which is time consuming task, and proposed method algorithm described in this paper outperforms in the detection rate, and it is faster. The average logo detection rate is 88.5%. The results of these tests are summarized in Table 1 [13].

Figure 7 shows Logo detection example. Logo detection rate of the proposed system is shown in Fig. 8. For performance evaluation of the automatic logo based document image retrieval system, it is significant to define a suitable metric. Two metrics are employed in our experiments as follows.

$$Recall = \frac{Number\ of\ relevant\ documents\ retrieved}{Number\ of\ relevant\ documents} \quad (9)$$

$$Precision = \frac{Number\ of\ relevant\ documents\ retrieved}{Number\ of\ documents\ retrieved} \quad (10)$$

In the document image retrieval system (DIRS), we have given a document image as query image. Based on query, DIRS retrieves document images which are having similar logo as present in the query document image. Results correspond to precision and recall rate for a Top 1, Top 2, Top 3, Top 4 and Top 5 as shown in Table 2. We observed from Table 2 that, we found 64.44% of precision for the top 5 document image retrieval.

In Fig. 9, Document image retrieval example results are presented in a list of images having a query document image.

6. Conclusion

In this paper, we have presented an approach to automatically logo detect and extract from the document images using discrete wavelet transform (DWT). Secondly, we have proposed approach, which retrieves document images based on the logo using combined rotated complex wavelet filters(RCWF) and dual tree complex wavelet transform(DT-CWT). Since these combined features give information in twelve different directions. We quantitatively evaluated the effectiveness of our approach in challenging retrieval tests using public, real-world document image collection involving as large number of classes but relatively small number of logos instances per class. Document matching was performed using Canberra distance. Retrieval results with proposed method are very promising with precisions and recalls.

References

[1] S. Seiden, M. Dillencourt, S. Irani, R. Borrey and T. Murphy, Logo detection in document images, *Proceeding of the International Conference on Imaging Science, Systems, and Technology*, Las Vegas, Nevada, (1997), 446–449

[2] D.E. Rivilin and I. Weiss, Applying algebraic and differential invariants for logo region, *Mach Vision Appl* **9**(2) (1996), 73–86.

[3] J. Neumann, H. Samet and A. Soffer, Integration of local and global shape analysis for logo classifica tion, *Proc of the 4th Int Workshop on Visual Forms Ages* (2001), 769–778.

[4] T.D. Pham, Unconstrained logo detection in document images, *Journal of the Pattern Recognition* (2003), 3023–3025.

[5] G. Zhu and D. Doermann, Automatic document logo detection, *International Conference Document Analysis and Recognition* (2007), 864–8668.

[6] G. Zhu and D. Doermann, Logo matching for document image retrieval, *Proc of Int Conf on Document Analysis and Recognition* (2009), 606–610.

[7] R. Tavoli, Classification and evaluation of document image retrie val system images, *WSEAS Trans* **11**(11) (2012), 329–338.

[8] M. Keyvanpour and R. Tavoli, Document image retrieval: Algorithms, analysis and prom ising directions, *Journal of Software Engineering and Its Applications* **7**(1) (2013).

[9] N.G. Kingsbury, Complex wavelets for shift invariant analysis and filtering of signals, *J App Comput Harmon Anal* **10**(3) (2001), 234–253.

[10] I. Selesnick, R. Baraniuk and N. Kingsbury, The dual-tree complex wavelet transforms, *IEEE Signal Process Mag* **22**(6) (2005), 123–151.

[11] M.B. Kokare, P.K. Biswas and B.N. Chatterji, Texture image retrieval using new rotated complex wavelets filters, *IEEE Trans on Systems Man and Cybernetics-part B: Cybernetics* **35**(6) (2005), 1168–1178.

[12] G. Zhu and D. Doermann, Tobacco-800 complex document image database and groundtruth, online, http://lampsrv01.umiacs.umd.edu/projdb/edit/project.php?id=52, 2008.

[13] M.S. Shirdhonkar and M.B. Kokare, Automatic logo detection in document images, *Proceeding of 2010 IEEE International Conference on Computational Intelligence and Computer Research*, Coimbatore, India, during 28–29 Dec 2011, 905–907.