Преглед НЦД 8 (2006), 11–21

Emil Kelevedjiev,
Andrey Andreev
(Institute of Mathematics and Informatics,
Bulgarian Academy of Sciences)

WAVELETS AND WORD IMAGE MATCHING[1]

Abstract: An approach to word image matching (WIM) based on wavelet transform (WT) is examined. A detailed computer experiments was carried out with respect to the type of wavelet filters, types of the wavelet bases used, the best suited frequency band and quantization of wavelet coefficients for WIM. The experiments for detecting user's specified words in Bulgarian and other typed documents of different quality show good steady results (above 95% properly detected words in average).

Key words: document text image, bitmap file, word searching, wavelet transform, quantization.

1. Introduction

In spite of the fact that newly generated documents are in text format which allows convenient storage and easy remote access to them, the problem for digitizing a huge amount of typewritten (handwritten) scientific, historical and cultural documents of different origin still remains open. The combination of advanced facilities for scanning and conversion such documents to digital images and Optical Character Recognition (OCR) technology, which automatically produces from them a computer readable text, resolves to a great extent the problem. However, as it is said in [3]: "It is generally acknowledged that the OCR accuracy requirements for information retrieval are considerably lower than for many document processing applications". Motivated by this and some other results in the recent time, the goal of this paper is to improve searching and location of a user specified word after segmentation of the digitized text image. Such problem for word spotting in scanned images is a subject of intensive research in both cases of the handwritten and typewritten texts [1,2,3]. The wavelet coefficients are used as a basic tool for achieving this goal taking into account possibilities of:

1. Existence of huge amount of different wavelets (filters) that have different support, smoothness, etc.;
2. Creation of different bases using a fixed filter;
3. Quantization of wavelet coefficients in different metrics;
4. Using the one of the "main advantages" of a given wavelet basis to other bases, e.g. the faster diminishing of wavelet coefficients considered by absolute value;

[1] This work is prepared under the Marie Curie Action Contract No.: MTKD-CT-2004-509754 – Knowledge Transfer for Digitisation of Cultural and Scientific Heritage in Bulgaria

5. Lower complexity of wavelet transform with respect to other commonly used transformations, for example Fast Fourier Transform (FFT);
6. Smoothing properties of wavelets [4] that results in noise removal (noise almost always exists in older documents caused by different sources) using thresholding (quantization) of wavelet coefficients.

All the above listed properties of wavelets are considered more in details and they are tested numerically later in order to find out from practical point of view the best approach for using wavelet coefficients.

2. Preprocessing

Based on the system and the code described in [5] we assume that the binarization and enhancement steps in the sense shown below are done:

Source document ⇒	digitized ⇒ image	improved image
1. books	1. bitmap file	1. Wiener filter
2. manuscripts	2. text file	2. WT
3. newpapers	3. word file	3. smoothness
4. etc	4. etc	4. etc

Simply, we suppose that the code has done the following:

1. Noise removal and image enhancement (compare Fig. 1 and 2, where scanned hand-written text is presented before and after noise removal).

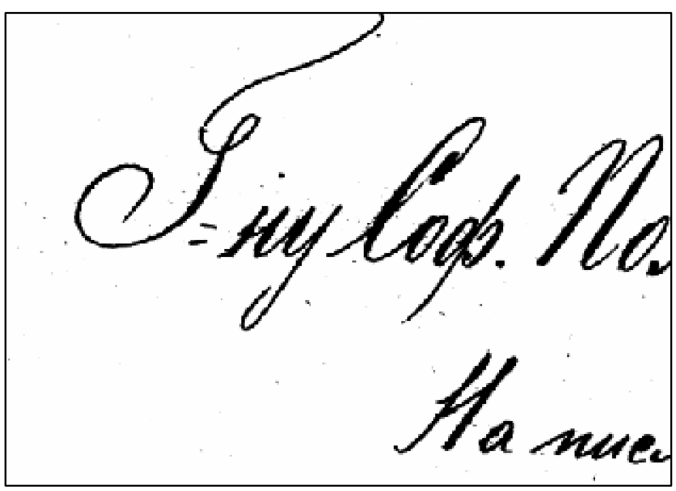


Fig. 1

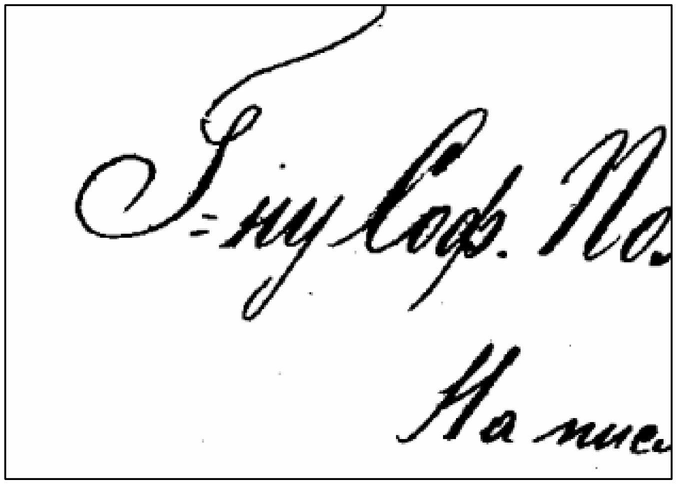
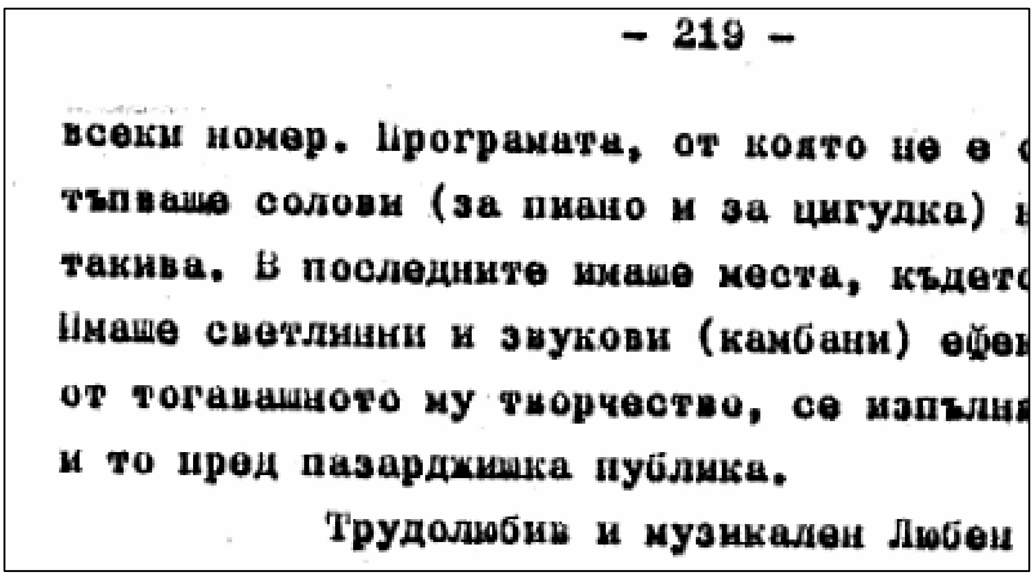


Fig. 2

2. Successful word segmentation of the digitized document and a bitmap file of each word (compare Fig. 3 and 4, where scanned typewritten text is presented);



\- 219 -

всеки номер. Програмата, от която не е

тъпваше солови (за пиано и за цигулка)

такива. В последните имаше места, където

Имаше светлинни и звукови (камбани) ефек

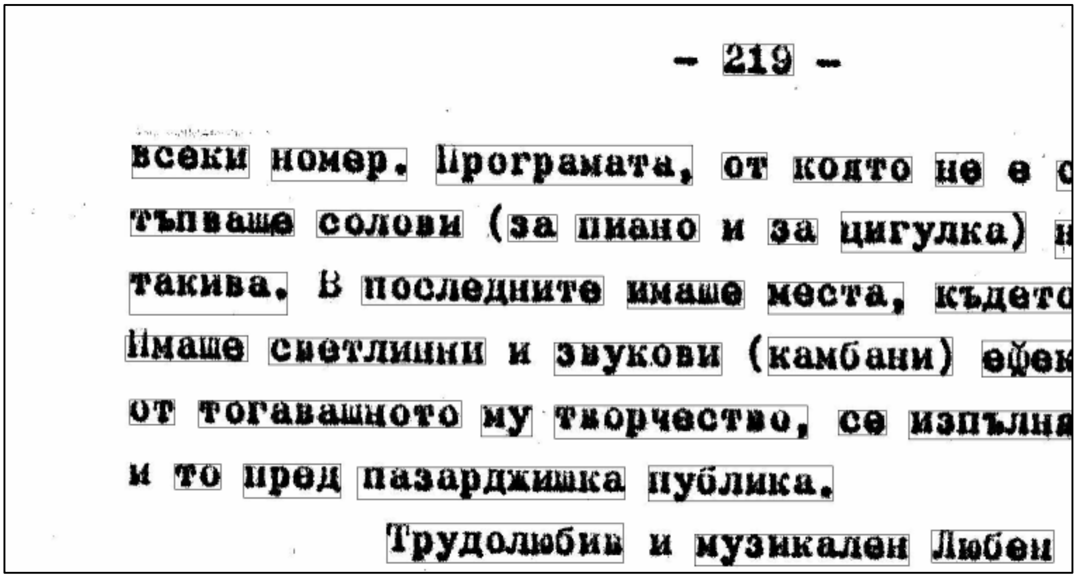
от тогавашното му творчество, се изпълни

и то пред пазарджишка публика.

Трудолюбив и музикален Любен

Fig. 3

3. Using interactive features of the code the alphabet was built selecting by the help of scanning technique all small and capital letters in the scanned document. This alphabet will be used later for creation of synthetic image word.


- 219 -

всеки номер. Програмата, от която не е с

тъпваше солови (за пиано и за цигулка) н

такива. В последните имаше места, където

Имаше светлинни и звукови (камбани) ефек

от тогавашното му творчество, се изпълня

и то пред пазарджишка публика.

Трудолюбив и музикален Любен

Fig. 4

4. Wavelets and Bitmap image file. Haar, Daubechies and Biorthogonal wavelets.

When representing an image via wavelets their coefficients behave differently with respect to the type of wavelet used and the properties of the image. In order to determine a suitable type of wavelet that reflects the features of the images that are typewritten Cyrillic words obtained by the code given in [5], we carried out series of experiments with 3 different wavelet functions, namely Haar, Daubechies-4 and 10-6 Biorthogonal wavelets. It is important to note that images like that obtained by scanning typewritten texts have a worse smoothness than usual pictures like “Lenna” used widely by researchers (compare Fig. 5 and 6).

Fig. 5

Fig. 6

If the function *f* is represented by orthogonal wavelet expansion

$$f(x) = \sum_i d_i \, \psi_i(x)$$

then the best way to pick N coefficients d_{ik} among $\{d_i\}$ making

$$\| f - f_N^{opt} \|_{L_2},$$

where

$$f_N^{opt} = \sum_{i_k=1}^{N} d_{i_k} \psi_{i_k}(x),$$

as small as possible is simply picking the N coefficients with largest absolute value. We determine the smoothness β of the Bitmap image file (as a 2-dimensional function) using the estimation [6, p. 111],

$$E_{L_2}^N(f) = \| f - f_N^{opt} \|_{L_2} = \frac{C}{N^{\beta}}, \qquad (1)$$

where the constant C in (1) is connected with a Besov norm of a function *f*. Using Haar wavelets the behavior of a relative error

$$\frac{E_{L_2}^N(f)}{\| f \|_{L_2}}$$

is given on Fig. 7 as a function of N for two images, Lenna from Fig. 6 and the middle image on Fig. 5.

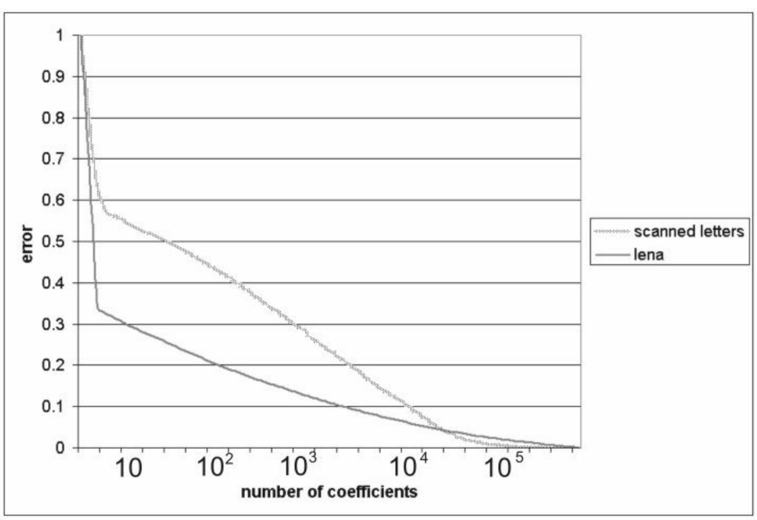


Fig. 7. Comparison of L_2 error for Lenna and word image

The similar experiments with other word images lead to the conclusion that the word images are of lower smoothness or what is more important is that such images can be characterized using less wavelet coefficients than must be used for portrait images like Lenna.

Although the “word images” have lower smoothness, the question which wavelets reflect in “best way” their essence remains open. As a first step in this direction, we carry out numerical experiments using: three relatively different wavelets – Haar, Daubechies-4 and 10-6 Biorthogonal wavelets [7]; a computational scheme for Biorthogonal wavelets with symmetric extension; two types of bases: Regular and Hyperbolic.

4. Biorthogonal wavelets

It is proved in [7] the existence of functions φ, ψ and φ', ψ' and coefficients

$$\{\alpha_j\}^{u}_{j=1-u},\quad \{\alpha'_j\}^{u'}_{j=1-u'}$$

for which under the notation $\varphi_{k,j}(x) = 2^{k/2}\varphi(2^k x - j)$ it is fulfilled that

$$\varphi = 2\sum_{j=1-u}^{u}\alpha_j\varphi_{1,j},\ \varphi' = 2\sum_{j=1-u'}^{u'}\alpha'_j\varphi'_{1,j},\ \psi = 2\sum_{j=1-u'}^{u'}\beta_j\varphi_{1,j},\ \psi' = 2\sum_{j=1-u}^{u}\beta'_j\varphi'_{1,j}$$

where $\beta_j = (-1)^j\alpha'_{1-j}$, $\beta'_j = (-1)^j\alpha_{1-j}$.

Table 1 gives the coefficients used in numerical experiments ($u = 5$, $u' = 3$).

j	α_j	α'_j
–4	0.01337437	
–3	0.00494231	
–2	–0.04754360	–0.09127176
–1	0.09432042	0.03372823
0	0.43490656	0.55754352
1	0.43490656	0.55754352
2	0.09432042	0.03372823
3	–0.04754360	–0.09127176
4	0.00494231	
5	0.01337437	

Table 1. 10-6 Biorthogonal wavelet filter

The computational scheme for Biorthogonal wavelets with symmetric extension is:

1. Let the function $f \in V_k$ is given by [7]:

$$f = \sum_{j=0}^{2^k-1} c_j^k \varphi_{k,j}$$

2. Decomposition step: for $0 \le j \le 2^{k-1} - 1$,

$$d_j^{k-1} = \sqrt{2}\sum_{l=1-u}^{u} \beta'_l \, c_{2j+l}^k, \quad c_j^{k-1} = \sqrt{2}\sum_{l=1-u'}^{u'} \alpha'_l \, c_{2j+l}^k,$$

using symmetric extensions

$$c_j^k = \begin{cases} c_{-1-j}^k, & j = -1, -2, -3, \ldots \\ c_{2^{k+1}-j-1}^k, & j = 2^k, 2^k+1, 2^k+2, \ldots \end{cases}$$

and f is represented in the form

$$f = \sum_{j=0}^{2^{k-1}-1} c_j^{k-1} \varphi_{k-1,j} + \sum_{j=0}^{2^{k-1}-1} d_j^{k-1} \psi_{k-1,j}$$

Reconstruction step: for $0 \le j \le 2^{k-1} - 1$,

$$c_{2j}^k = \sqrt{2}\left(\sum_{l=-[u/2]}^{[u/2]} \alpha_{2l} c_{j-l}^{k-l} + \sum_{l=-[(u'-1)/2]}^{[u'/2]} \beta_{2l} d_{j-l}^{k-l} \right),$$

$$c_{2j+1}^k = \sqrt{2}\left(\sum_{l=-[u/2]}^{[(u-1)/2]} \alpha_{2l-1} c_{j-l}^{k-l} + \sum_{l=-[u'/2]}^{[(u'-1)/2]} \beta_{2l+1} d_{j-l}^{k-l} \right)$$

using the symmetric extensions for c_j^k and

$$d_j^k = \begin{cases} -d_{-1-j}^k, & j = -1, -2, -2, \ldots \\ -d_{2^{k+1}-j-1}^k, & j = 2^k, 2^k+1, 2^k+2, \ldots \end{cases}$$

5. Regular and Hyperbolic wavelets in 2-D

Any square gray scale image can be considered as a matrix M of pixel values

$$M = \{p_{i,j}: \; i, j = 0, 1, \ldots, 2^n - 1, \; 0 \le p_{i,j} \le 255, \text{ where } p_{i,j} \text{ are integers}\}.$$

We associate with M the discrete function $f(x, y)$ of two variables

$$f(x, y) = \sum_{i,j=0}^{2^n-1} c_{i,j}^n \varphi_{n,i}(x) \varphi_{n,j}(y), \text{ for } c_{i,j}^n = p_{i,j}.$$

It is evident that f can be represented using any basis of 2^{2n} linearly independent functions from the following set of functions:

$$\{\varphi_{i,j}(x)\varphi_{k,l}(y), \; \varphi_{i,j}(x)\psi_{k,l}(y), \; \psi_{i,j}(x)\varphi_{k,l}(y), \; \psi_{i,j}(x)\psi_{k,l}(y)\}_{i,j,k,l=0}^{2^n-1}$$

For our goals we compare numerically the following two commonly used bases:

Regular basis: It is built by the following 2^{2n} functions:

$$\{\varphi_{0,0}(x)\varphi_{0,0}(y),\ \varphi_{k,l}(x)\psi_{k,j}(y),\ \psi_{k,l}(x)\varphi_{k,l}(y),\ \psi_{k,l}(x)\psi_{k,l}(y)\},$$
$$k = 0,1,\ldots,n-1,\ \ l, j = 0,1,\ldots,2^k - 1.$$

Hyperbolic basis: It is built by the following 2^{2n} functions:

$$\{\varphi_{0,0}(x)\varphi_{0,0}(y),\ \varphi_{0,0}(x)\psi_{k,j}(y),\ \psi_{k,l}(x)\varphi_{0,0}(y),\ \psi_{k,l}(x)\psi_{i,j}(y)\},$$
$$k = 0,1,\ldots,n-1,\ \ l = 0,1,\ldots,2^k - 1,\ \ j = 0,1,\ldots,2^i - 1.$$

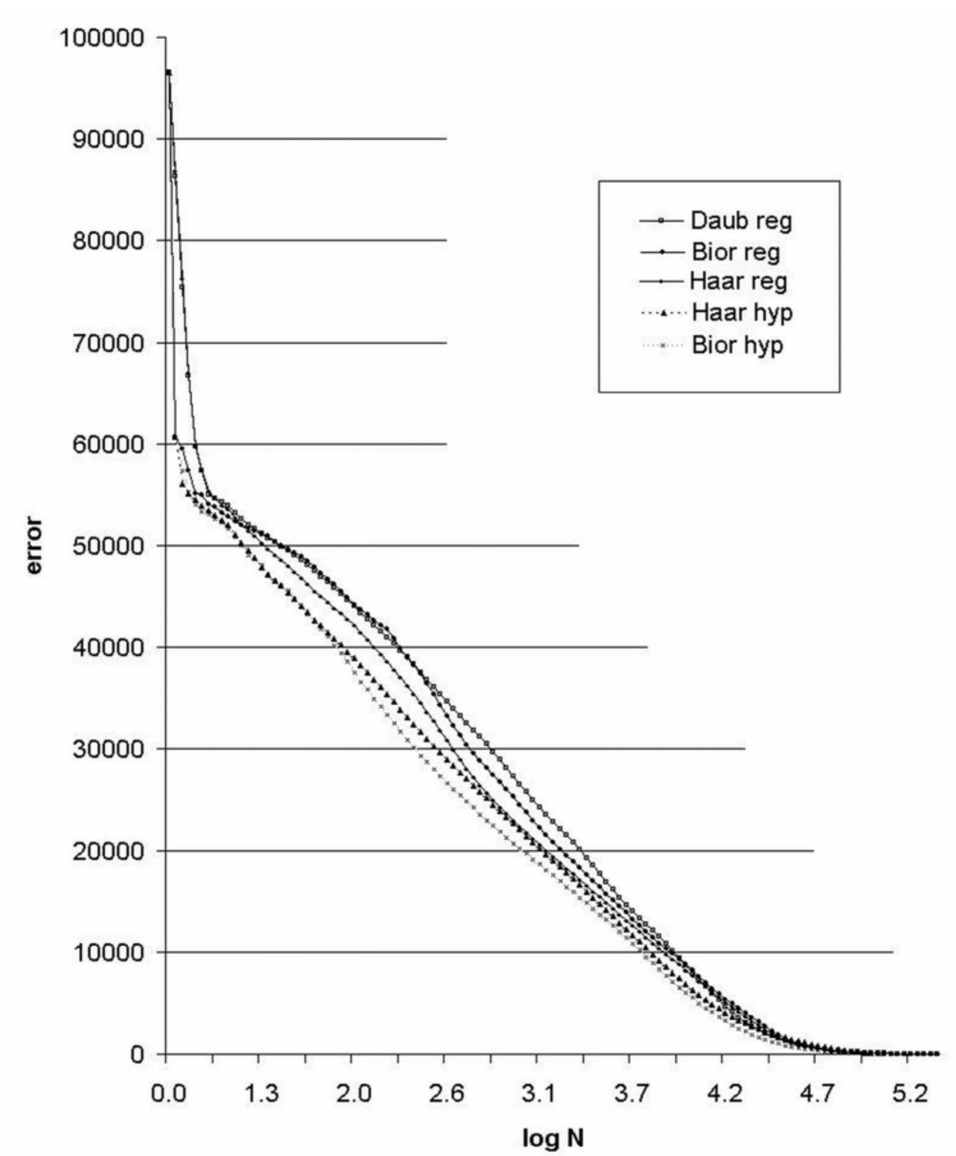


Fig. 8. L_2 approximation of a word image by different wavelets

L_2 error is shown on Fig. 8, when we approximate the middle word image given in Fig. 5. Here the x-axis shows in logarithmic scale the number *N* of wavelet coefficients used for reconstruction of the image.

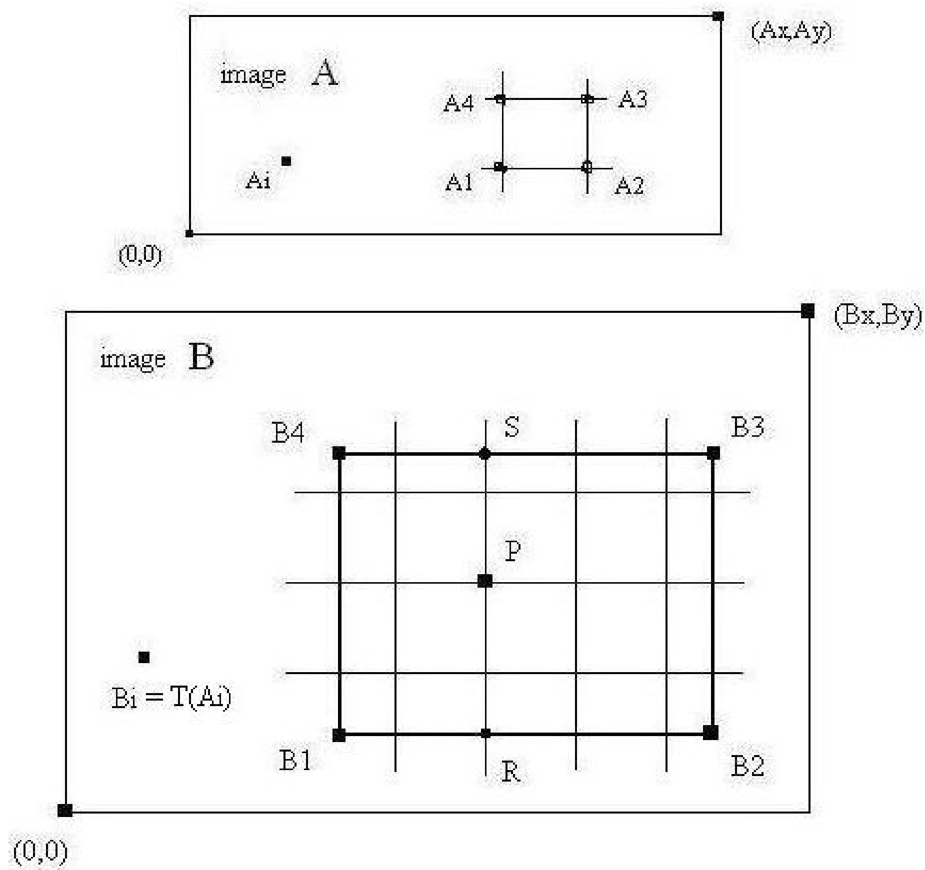


Fig. 9. Images *A* and *B* for which *B* = *T*(*A*)

6. Numerical experiments. Resizing algorithm.

In order to simplify the process of computations using wavelets we resize the word images like these shown on Fig. 5 to square image with the size which is $(2^n, 2^n)$, where n is 7, 8, or 9. For this aim we apply the following algorithm:

1. Let images *A* and *B* shown on Fig. 9, with sizes (Ax, Ay) and (Bx, By) satisfy $B = T(A)$, where the transformation *T* is defined by: if $Bi = T(Ai)$, where *Ai* is a pixel from *A* with coordinates $Ai = (Ai_x, Ai_y)$ and *Bi* is a point (generally not a pixel) from *B* with coordinates $Bi = (Bi_x, Bi_y)$ then

$$Bi_x = (Bx/Ax)\, Ai_x, \quad Bi_y = (By/Ay)\, Ai_y;$$

2. Let the pixels $A1 = (i, j)$, $A2 = (i + 1, j)$, $A3 = (i + 1, j + 1)$, $A4 = (i, j + 1)$ be mapped by T to the points $B1 = (B1_x, B1_y)$, $B2 = (B2_x, B2_y)$, $B3 = (B3_x, B3_y)$, $B4 = (B4_x, B4_y)$. The last 4 points build a rectangle.

If a pixel $P = (k, l)$ belongs to this rectangle then define its value by: If R is a pixel (or point) then by $|R|$ we denote its pixel value (or a value assigned to this point). We set $|A1| = |B1|$, $|A2| = |B2|$, $|A3| = |B3|$, $|A4| = |B4|$;

Let the points R and S, as given on Fig. 9, have coordinates $R = (k, R_y)$ and $S = (k, S_y)$. Then

$$|P| = ((1 - R_y)/(S_y - R_y))|R| + ((S_y - 1)/(S_y - R_y))|S|,$$

where

$$|R| = ((k - B1_x)/(B2_x - B1_x))|B1| + ((B2_x - k)/(B2_x - B1_x))|B2|,$$
$$|S| = ((k - B4_x)/(B3_x - B4_x))|B4| + ((B3_x - k)/(B4_x - B3_x)|B3|.$$

In the above notation let us point out to the fact that: T can be applied if $Bx \geq Ax$, $By \geq Ay$; if A and B are two images for which $B = T(A)$, then obviously the inverse transformation T^{-1} exists and $A = T^{-1}B$.

7. Application to image recognition

In order to compare two images, we could use the subband of wavelets coefficients. Usually about 10% of the biggest by absolute value coefficients contain enough information for the purposes of word matching. If two words, a and b, are represented by wavelet coefficients a_{ij} and b_{ij}, then we can check for the proximity of a and b by using some estimation. In our tests, we use the square metric:

$$\text{Distance}(a,b) = \sum_{i,j} (a_{i,j} - b_{i,j})^2$$

To find an image, which is "close" ("the same word") to the given one, some threshold parameter is used. The obtained results in this way show a good coincidence with the human perception. An advantage is a small amount of memory needed to maintain the sample.

Experiments are made with:
– several wavelet filters (Haar, Doubechies 4 and 8, etc);
– several cases of choosing which part of the coefficients has to be taken;
– several choices of the threshold parameter.

The following example was accomplished with: Doubechies 8, subbands part: 1/32, threshold = 10^6. The groups of the "similar" words on the rows in Fig. 10 are obtained from the scanned image of a typewritten page after applying a code for segmentation of words.

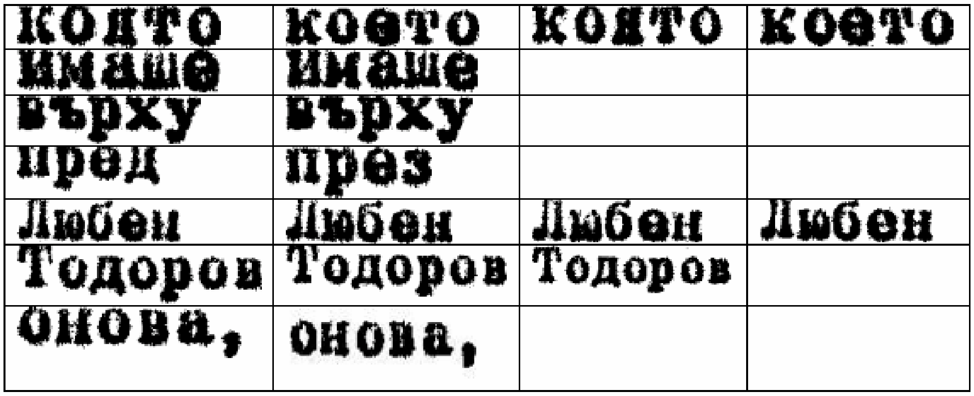


Fig. 10

Conclusions:

1. With fixed number of coefficients the hyperbolic type of wavelets extract more information about the image than regular type of wavelets when applied for word images;
2. Haar hyperbolic wavelet is of the same quality as 6-10 Biorthogonal wavelet if we keep 50-500 wavelet coefficients;
3. Taking into account the simplicity and speed of computations using Haar filter it makes sense to use Haar hyperbolic wavelets for word images;
4. The experiments with “word images” lead to the conclusion that the “word images” are of lower smoothness i.e. there is no need to use wavelets with high smoothness. It seems that Haar wavelet is a good choice as a tool for word matching.
5. The wavelet coefficients can be used as a tool for achieving progress in OCR;

References

1. Y. Lue, C. L. Tan, W. Huang, L. Fan, “An Approach to Word Image Matching Based on Weighted Hausdorff Distance”, *6th ICDAR*, 10–13 Sept. 2001, Seattle, USA.
2. Y. Lu, C. L. Tan, “Chinese word searching in imaged documents”, *International Journal of Pattern Recognition and Artificial Intelligence*, Vol. 18, no.2, pp. 229–246, 2004.
3. A. Kunde, “Handwritten word recognition using hidden Markov model”, in H. Bunke and P. S. P. Wang (Ed.) *Handbook of character recognition and documents image analysis*, World Scientific, 1977, pp. 157–182.
4. D. L. Donoho, I. M. Johnstone, “Ideal Denoising in an orthonormal basis chosen from a library of bases", *C. R. Acad. Sci. Paris*, Ser. *I*, 319(1994), 1317–1322.
5. Konidaris, B. Gatos, K. Nizios, I. Pratikakis, S. Theodoridis, S. J. Perantonis, “Keyword Guided Word Spotting in Historical Typewritten Documents Using Synthetic Data and User's feedback”, 2005, (submitted).
6. SIGGRAPH'95 Course Note, Wavelets and their Applications in Computer Graphics.
7. A. Cohen, I. Daubechies, J.-C. Feauveau, “Biorthogonal Bases of Compactly Supported Wavelets”, *Comm. Pure and Applied Math*. **45** (1992), 485–560.

keleved@math.bas.bg
aandreev@math.bas.bg