ZESZYTY NAUKOWE POLITECHNIKI ŚLĄSKIEJ 2012
Seria: TRANSPORT z.74 Nr kol. 1861

Wiesław PAMUŁA[1]

A WAVELET-BASED VEHICLES DETECTION ALGORITHM

Summary. The detection of vehicles, in video streams from road cameras, is generally performed by analyses of the occupancy of virtual detection fields defined in image frames. This principle of detection is sensitive to ambient light variations, vehicle shadows, and camera movement. The paper presents a method for detection of vehicles that uses transformed image frames. To facilitate detection each frame is converted into a vector of pixel values. Consecutive video vectors are transformed using one-dimensional DWT. Stopped vehicles are represented by stripes, whereas moving ones by checked patches. The width of a stripe indicates vehicle size, while the length shows how long the vehicle waited at the approach to the intersection.

ALGORYTM DETEKCJI POJAZDÓW OPARTY NA FALKACH

Streszczenie. Wykrywanie pojazdów w strumieniu wideo z kamer drogowych oparte jest zwykle na analizie zajętości wirtualnych pól detekcji. Ten sposób wykrywania jest czuły na zmiany oświetlenia, cienie pojazdów i ruchy kamery. Artykuł przedstawia metodę wykrywania, która wykorzystuje transformaty klatek obrazów. W celu umożliwienia sprawnej analizy zawartość klatki zamieniana jest najpierw na wektor wartości pikseli. Kolejne wektory wideo są transformowane z użyciem jednowymiarowego, dyskretnego przekształcenia falkowego. Zatrzymane pojazdy są reprezentowane przez paski, a ruchome przez kratkowane placki. Szerokość paska wskazuje na rozmiar pojazdu, a długość określa, jak długo pojazd oczekiwał na wlocie skrzyżowania.

1. INTRODUCTION

Intelligent transport systems (ITS) for controlling and managing road traffic require a large spectrum of traffic data. Depending on the level of control, the data is acquired locally at intersection links, along network routes and globally along travel routes. Data is gathered and stored for evaluating traffic control strategies. Much of the data is processed into synthetic measures of road traffic parameters.

ITS commonly utilise video based devices for collecting traffic data. Widely used systems such as Autoscope, Trafficam, Peek, Iteris provide data of limited reliability [1]. Systems

[1] *wieslaw.pamula@polsl.pl*

detect vehicles by analyses of the occupancy of virtual detection fields defined in image frames obtained from road traffic cameras.

Autoscope determines the occupancy by subtracting the current frame from a frame representing scene background [2]. It assumed that the differences account for vehicles. The background model is sensitive to ambient light changes, because it is a simple single mode statistical model. The resultant occupancy signals are affected by deficiencies of this model. Similarly, Trafficam uses segment analysis along traffic lanes with updated past contents [3].

Background modelling is the source of detection errors in the widely used video detection systems. More accurate modelling uses a mixture of Gaussians [4] or performs filtering in space-time domain, using Fourier transform [5]. The scene background is removed by identifying the high-energy frequency components in the Fourier spectrum, setting them to zero, and finally back transforming to the spatial domain image using the inverse Fourier transform. In the reconstructed image, the background structure texture in the original frame becomes a uniform region yielding better responses when calculating occupancy factors. The Fourier transforms are computed efficiently using DSP circuits.

Scale space analysis is commonly used in image processing for determining object features and tracking object movement [6]. Elaborating on this principle a method for detection of vehicles was devised. Discrete wavelet transform was chosen as the basis for the algorithm. The characteristics of the devised algorithm are presented and several results of its application are discussed.

2. DWT

Discrete wavelet transform gives a multi-resolution representation of signals [7]. Signal, scaling functions, and wavelets are discrete. The DWT of a sequence consists of two series expansions, one corresponding to the approximation and the other to the details of the sequence.

The DWT of an N-point sequence *p* is given by:

$$
\begin{aligned}
&DWT(p) = \frac{1}{\sqrt{N}}\sum_{i=0}^{N-1} p(i)\phi_{j_0,k}(i) + \frac{1}{\sqrt{N}}\sum_{i=0}^{N-1} p(i)\psi_{j,k}(i) \\
&\phi_{j_0,k}(i) = 2^{j/2}\phi(2^j i - k) \\
&\psi_{j,k}(i) = 2^{j/2}\psi(2^j i - k) \qquad j \ge j_0 ,
\end{aligned} \tag{1}
$$

where:
$\varphi_{j,k}$ – scaling functions,
$\psi_{j.k}$ – dyadic wavelets.

Wavelet transforms can be written using lifting steps; it follows that an integer version of these can be devised [8]. Transforms that map integers into integers are most desired for efficient implementation using hardware processing.

Images of traffic scenes contain objects, which may be approximated by cuboids or projections of these. Adequate transform representation may be attained with symmetric wavelets with a few vanishing moments [9]. Suitable are interpolating transforms of the form (N, Ñ) where N is the number of vanishing moments of the analyzing high pass filter, while Ñ is the number of vanishing moments of the synthesizing high pass filter. Examples are wavelet transforms built from the interpolating Deslauriers–Dubuc scaling functions such as (1, 1), (2, 2), (4, 4) (table 1).

The principle of construction of the lifting scheme, in the case of one dimensional data streams, can be shown in three steps. In the first step, the stream of values is split into two sub streams one consisting of values with odd indices the other consisting of values with even ones.

The split step is:

$$\begin{aligned} d_{0,i} &= p_{2i+1} \\ s_{0,i} &= p_{2i}\ , \end{aligned} \tag{2}$$

where:
p_i – i-th data value,
$d_{n,i}$ – high-pass coefficients of n-th sub band (also called difference coefficients),
$s_{n,i}$ – low-pass coefficients of n-th sub band (also called scaling coefficients).

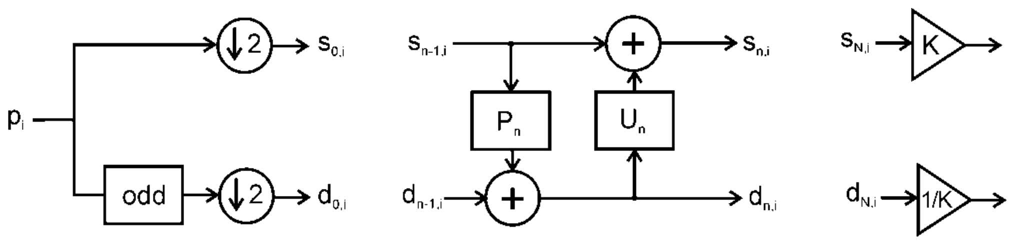

Fig. 1. Lifting scheme steps for computing wavelet coefficients, split, lifting, normalisation
Rys. 1. Kroki obliczania współczynników transformaty falkowej

Next, the odd and even sub streams are transformed by the prediction $P_n(k)$ and by the update, $U_n(k)$ filters.

The lifting step, illustrated also on fig.1, consists of computing:

$$\begin{aligned} d_{n,i} &= d_{n-1,i} + \sum_{k=1}^{N} P(k) s_{n-1,i} \\ s_{n,i} &= s_{n-1,i} + \sum_{k=1}^{N} U(k) d_{n,i}\ . \end{aligned} \tag{3}$$

N sub steps constitute this step.

The last step is the normalisation, which is necessary if the transform coefficients will be used for reconstructing the data. In the case of interpolating transforms K=1. Table 1 contains the prediction and update functions for wavelets used in representing image data.

The listed transforms are separable so their two-dimensional versions can be implemented in a row-column fashion. The versions are easily computed. A number of video streams from road cameras, registered at different times of day that is in different lightning conditions, were transformed. The performance of computing was evaluated for speed and for sensitivity of coefficient values to ambient light variations.

Table 1

Integer to integer wavelet transforms

(1,1)	$d_{1,i} = s_{0,2i+1} - s_{0,2i}$ $s_{1,i} = s_{0,2i} + \lfloor d_{1,i}/2 \rfloor$
(2,2)	$d_{1,i} = s_{0,2i+1} - \lfloor (s_{0,2i} + s_{0,2i+2})/2 + 1/2 \rfloor$ $s_{1,i} = s_{0,2i} + \lfloor (d_{1,i-1} + d_{1,i})/4 + 1/2 \rfloor$
(4,4)	$d_{1,i} = s_{0,2i+1} - \lfloor 9(s_{0,2i} + s_{0,2i+2})/16 - (s_{0,2i-2} + s_{0,2i+4})/16 + 1/2 \rfloor$ $s_{1,i} = s_{0,2i} + \lfloor 9(d_{1,i-1} + d_{1,i})/32 - (d_{1,i-2} + d_{1,i+1})/32 + 1/2 \rfloor$

No significant differences were noted so the (1, 1) interpolating transform was chosen for further tests as it is the least computationally demanding.

Transform coefficients are localised in space so may be regarded as a sort of descriptions of space features of the image frame content. One cannot map these, in most cases, to physical features of objects on the image. Some coefficients resemble object edges (fig. 2). This characteristic will be exploited in the proposed method of detecting vehicles.

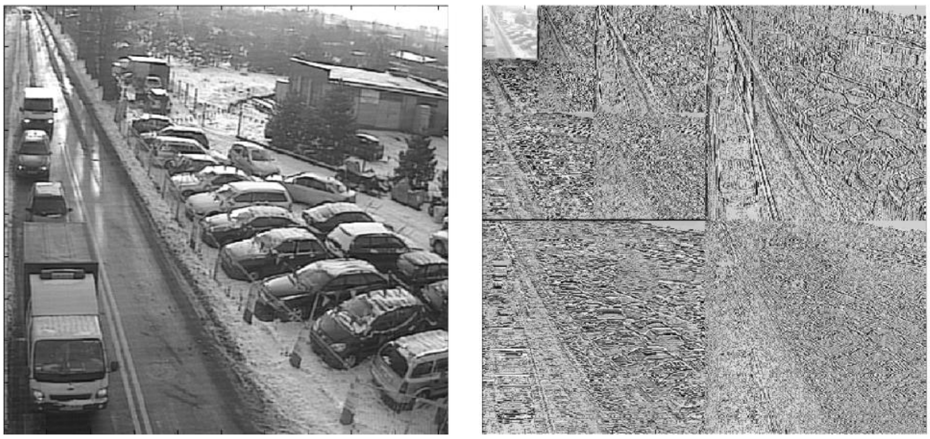

Fig. 2. A video frame of a traffic scene and its 2D DWT coefficients at different scales
Rys. 2. Klatka wideo obrazu sytuacji drogowej i współczynniki 2D DWT w różnych skalach

The shrinking squares on the diagonal of the DWT image represent high-pass coefficients of consecutive sub bands. The higher the sub band the less legible is the contents of the square. Coefficients describe contents details of a growing patch of the original image.

Patch area grows 4 times with each sub band. Starting from single pixel values the fourth sub band contains coefficients describing 16x16 pixel patches. For images acquired by a camera mounted on a commonly used observation post this accounts for an area of about 4 to 9 m^2 in the middle of the image, which is just about the size of a passenger car.

Coefficients of the fourth sub band may describe cars in a traffic scene. Wavelet transform coefficients carry information on all of the image content so this assumption is only an estimate of the object features.

Moving objects are signalised by changing values of transform coefficients and translation of positions of these. As vehicles move through the field of view, their movement is registered as variable changes of coefficient values in contrary to gradual value changes of coefficients describing the background. Ambient light variations influence both coefficient groups, but as the movement rate is usually higher than light change, the predominant change would be accounted for vehicle changing positions on the image.

Stopped vehicles produce distinct groups of coefficients. These coefficients differ very much from background ones, because it is highly unlikely that a vehicle has exactly the same features as the patch of background that is covered by the vehicle. The coefficient values change little during the stop.

3. VIDEO FRAME REPRESENTATION

Road traffic cameras provide video streams at the rate of 25 frames per second, this allows for a 40 ms resolution in time analysis of traffic scenes. Such a high update rate is useful for detailed traffic tracking, but very demanding for processing such data. Reducing the flow without losing vital information is the aim of processing.

Video frames coming at consecutive time intervals constitute a 3D data object. The three-dimensional data object is difficult for analysis; especially using hardware based computing devices. It would be more desirable to have a two dimensional representation.

A way to construct a two dimensional representation is to convert the image into a pixel vector *W* and assemble, consecutive in time, vectors into a 2D surface. The image to vector conversion function determines surface correspondence to the real world scene. A space-filling curve suggests itself as a function of first choice because of its uniform coverage of the frames surface. Space-filling curves are used for mesh generation, multi-dimensional indexing in computer database management systems.

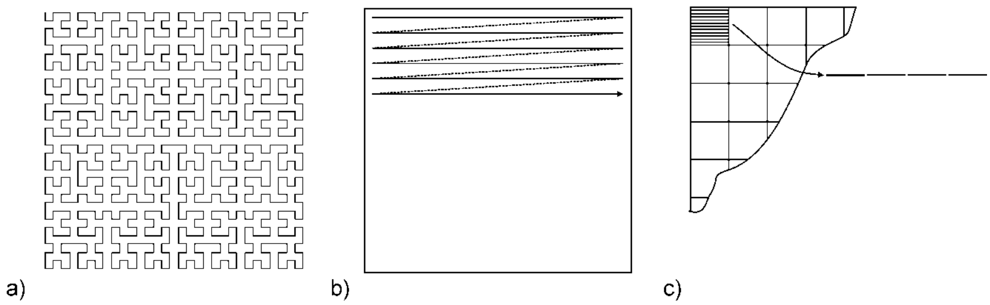

Fig. 3. Vector mapping of frame pixels: a) Hilbert curve, b) row by row, c) square wise ordering
Rys. 3. Wektorowe odwzorowanie pikseli klatki: a) po krzywej Hilberta, b) wiersz po wierszu, c) kwadratami

Peano and Hilbert proposed the first constructions of space filling curves in the end of the XIX century [8]. It is a continuous mapping of a closed interval into a closed two-dimensional area. The curve passes through every point of the area. The construction is usually defined as an iterative procedure. In practical applications for fixed sized matrices of values, such as images *I(x, y)*, it becomes a mapping of consecutive addresses of values:

$$W(i) = I(H(i \operatorname{div} n, i \operatorname{mod} m)) \qquad i = 1..n \cdot m, \tag{4}$$

where:
H(x, y) – Hilbert curve addressing matrix.

Another solution is to take consecutive rows and join their ends to form the pixel vector. This complies with the serial character of a video stream from a road camera, but destroys important two-dimensional relations between pixels (fig. 3b).

To conserve these a square wise ordering was proposed. The frame was divided into squares, which were individually scanned row by row, and then adjacent rows were added to the pixel vector (fig. 3c):

$$\begin{aligned} &W(i) = I(y,x) \\ &y = (i \operatorname{div} ma)ma + (i \bmod a)m \\ &x = ((i \bmod m) \operatorname{div} a)a + (i \bmod a)\,, \end{aligned} \tag{5}$$

where:
a – length of square side,
m – number of image frame columns.

The size of the squares corresponds to the envisaged scale of analysis. As noted in section 2 a size of 16x16 pixels is comparable to vehicle sizes and this size was used for converting frames square wise. To implement the square wise ordering a video small buffer was required to save a row of image squares.

The size of the vector, in all cases, is equal to the number of pixels in the image frame. Standardised PAL frames commonly used by road cameras have 720x576 pixels. Although the standard defines such a frame size, the actual resolution is smaller, highly dependent on the quality of the image-sensing device.

Wavelet calculation based on lifting schemes is optimally done when the number of operands is a power of two in the case of one-dimension transforms or power of four in two dimensions. The largest number of pixels being a power of two, which may be cut out of a PAL frame, is 2^{18} that is a 512x512 pixel square.

The following prerequisites were defined for processing frames in the analysis of video data for detecting vehicles:

- use DWT based on (1,1) interpolating wavelets,
- analyse coefficients corresponding to image patches of the size 16x16 pixels,
- use Hilbert curve mapping or square wise ordering for pixel vector representation of frame content.

4. DETECTION OF VEHICLES

Incoming image frames were converted into pixel vectors and transformed using (1.1) interpolation DWT. To achieve the desired coefficient description of image patches the transform was computed up to the sub band s:

$$\begin{aligned} &s = \log_2 n_s \\ &s = \log_2 (16 \cdot 16) = 8\,, \end{aligned} \tag{6}$$

where:
n_s – number of pixels in an image patch (square).

Variations of ambient illumination or camera internal parameters have more influence on the low-pass information in images, than on the high-pass content; therefore, high-pass coefficients for this band were taken for analysis.

There were $N_c = 2^{18}/(16\cdot16) = 1024$ coefficients. This small number of coefficients compared to the number of frame pixels was a great advantage for carrying out the analysis process. These were integers from a range of values significantly smaller than the initial pixel range of values. For each sub band, the range shrunk about 1/4 (table 1). The pixels have had values in the range 0–255, at the 8th sub band the range shrunk to $(3/4)^8$ of the original. This means that a coefficient could be coded with 6 bits instead of the initial byte. A shorter code requires fewer resources, which is advantageous when designing hardware for processing.

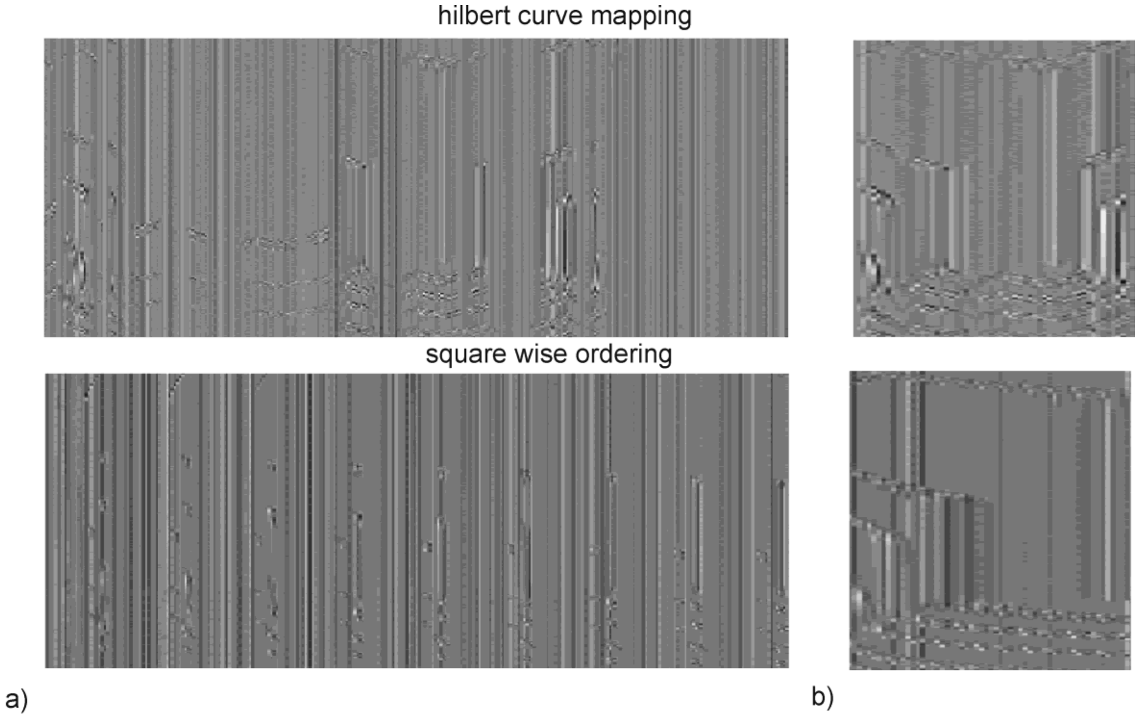

Fig. 4. DWT transforms: a) first 300 coefficients, b) coefficients corresponding to a traffic lane
Rys. 4. Transformata falkowa: a) pierwsze 300 współczynników, b) współczynniki odpowiadające pasowi ruchu

Figure 4 presents samples of a frame sequence of transformed frames. Frames were registered every 80 ms. The left images contain about 200 segments with the first 300 high-pass coefficients of the eighth sub band. This corresponds to the upper left part of the traffic lane on fig. 2. Background features remain as stable vertical stripes. In both cases of pixel mappings, object features, described by coefficients, are spread along the horizontal axis (fig. 4a). This gives a vague view of objects and their behavior.

Hilbert curve based analysis surface, rendered approximately vehicle queuing. This was because the curve meanders on the image frame clustering pixels. Square wise ordering was hard to interpret, as there was no clue to the position of the traffic lane and to other points of interest in the traffic scene.

In order to facilitate a more efficient analysis, coefficients describing areas outside traffic lanes were deleted. Fig. 4b presents a sample of the remaining coefficients that maps vehicle behavior in time. Stopped vehicles are represented by stripes whereas moving ones by checked patches. Square wise ordering of pixels produced an ordered series of stripes. The coefficients mapped square areas of the traffic lane, uniformly row by row, beginning from the furthest to the nearest part of it. In the case of Hilbert curve, stripes of different lengths were mixed, as the curve did not scan the area of the lane row by row but meandered.

Using Hilbert curve mapping, analysis may be done without extracting traffic lane coefficients. In the case of applying square wise ordering, it is necessary to delete non-lane coefficients to achieve a meaningful view of vehicle behavior for analysis.

A vehicle occupies a number of patches (squares) of the image; this is marked by coefficient values differing from background coefficients. The width of a group, of equal in length stripes, reflects the vehicle's size. The width changes when the vehicle stops in different parts of the image. This effect is due to the projection of the vehicles size registered by the camera. Depending on the position of the camera relative to observed traffic and related projection parameters, this change may prevent accurate classification of vehicles.

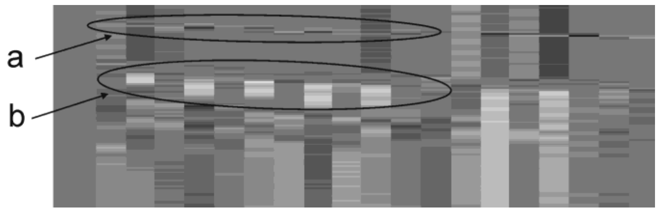

Fig. 5. Moving objects a) small like cars, b) large like buses
Rys. 5. Przemieszczające się obiekty a) małe obiekty - samochody, b) duże obiekty – autobusy

Distinguishing buses from passenger cars can be done in most cases as projection parameters seldom distort the view so much as to equalize car and bus projection views.

Signalised link accumulates vehicles when the traffic lights go red. For a set period of time, a queue builds up (fig. 4b). A characteristic horizontal checked line leads to each group of stripes. The width of this line signifies the size of the vehicle. Large vehicles have wide lines leading to their stop.

Fig. 5 illustrates the differences between moving small objects such as cars and large ones like buses.

No calibration of the camera field of view was done so only approximate classification was possible. The distinction between buses and cars was adequate and may be utilised in traffic control for instance for prioritising public transport.

5. CONCLUSIONS AND FUTURE WORK

Wavelet analysis of image sequences proves useful for detecting moving vehicles in road traffic. It is possible to distinguish vehicle classes such as cars and buses. The estimation of queue lengths and occupancy factors may also be done. Such data is frequently used for predicting traffic flow characteristics for traffic controllers working together along a traffic lane.

No effort yet was undertaken to devise automatic procedures for determining individual vehicles and their tracks. This task may be accomplished by analysis of the coefficient surface utilising image-processing algorithms. Methods for extracting features will be especially useful. Features such as stripes, checked lines have clearly defined characteristics, which may facilitate efficient processing.

On line, analysis methods of the coefficient surface will be the goal of future research.

Bibliography

1. Chitturi M.V., Medina J.C., Benekohal R. F.: Effect of shadows and time of day on performance of video detection systems at signalized intersections. „Transportation Research Part C” vol. 18, 2010, p. 176-186.
2. Michalopoulos P.G., Fundakowski R.A., Geokezas M.,Fitch R.C.: Vehicle detection through image processing for traffic surveillance and control, US Patent 4,847,772 Jul.11.1989.
3. Bunnen B., Bogaert M., Versaver J.: Traffic monitoring device and method, US Patent 5.912,634 Jun 15. 1999.
4. Zhang W., Fang Z.X., Yang X.: Moving vehicles segmentation based on Bayesian framework for Gaussian motion model, „Pattern Recognition Letters”, vol. 27, 2006, p. 956-967.
5. Tsai D-M., Chiu W-Y.: Motion detection using Fourier image reconstruction, „Pattern Recognition Letters”, vol. 29, 2008, p. 2145-2155.
6. Zhou H., Yuan Y., Shi C.: Object tracking using SIFT features and mean shift, „Computer Vision and Image Understanding”, vol. 113, 2009, p. 345-352.
7. Strang, G.: Wavelets and Dilation Equations: A Brief Introduction, „SIAM Review”, vol. 31, 1989, p. 614-627.
8. Calderbank A. R., Daubechies I., Sweldens W.: Wavelet Transforms that Map Integers to Integers, „Applied and computational harmonic analysis”, vol. 5, 1998, p. 332-369.
9. Andreopoulos Y., Munteanu A., Van der Auwera G., Cornelis J.P.H., Schelkens P.: Complete-to-Overcomplete Discrete Wavelet Transforms: Theory and Applications, „IEEE Transactions On Signal Processing”, vol. 53, 2005, p. 1398-1412.
10. Moon B., Jagadish H.V., Faloutsos C.: Analysis of clustering properties of Hilbert space filling curve, „IEEE Transactions on Knowledge and Data Engineering”, vol. 13, 2001, p.124-141.