2D DT-CWT CBIR With Adaptive Selection of the Decomposition Level

Ivo Draganov[1]✉ [0000-0001-5076-3020] and Stella Vetova[1]

[1] Technical University of Sofia, 8 Kliment Ohridski Blvd., 1756 Sofia, Bulgaria
idraganov@tu-sofia.bg, vetova.bas@gmail.com

Abstract. In this paper, a novel algorithm for content-based image retrieval is presented based on the two-dimensional dual tree complex wavelet transform. It includes adaptive selection of the decomposition level consistent with preliminary set retrieval time. Depending on the size of the image queries further adaptation could be applied based on the volume of information contained in the sub-band of approximation coefficients from the wavelet spectrum so the overall retrieval accuracy remains as high as possible without violating the time requirements. Specific to this approach algorithm for image database indexing is also developed. Test results are obtained with the Wang database. The average precision proved to be higher than that of other popular approaches from the practice for 5 of the image categories, reaching 100% for 2 of them. It remains almost equal for other 2 of the testing categories. It is considered to be especially applicable into CBIR systems with critical time requirements variable in different use case scenarios, including preserving cultural heritage, assuring scalability while preserving retrieval accuracy.

Keywords: 2D DT-CWT • CBIR • Decomposition level • Retrieval time • Retrieval accuracy.

1 Introduction

Content-based image retrieval (CBIR) using the two-dimensional Dual Tree Complex Wavelet Transform (2D DT-CWT) has been used for a number of years now [1-5]. CBIR systems are now widely used in preserving cultural heritage on a world-wide scale processing images of artifacts, artificially synthesized 3D models and entire sites. In [1] Patil and Talbar tested both the DT-DWT and DT-CWT for image retrieval. They selected fourth level for decomposition with major consideration of the feature size. Further, this approach is developed by Vhanmane and Sangve [2] taking the same features with Grey Level Co-occurrence Matrix (GLCM) over them. Slightly higher retrieval accuracy is achieved. In order to increase the retrieval precision, Patil and Kokare [3] include relevance feedback, provided by the user, and used rotated complex wavelet filters. They found that three iterations are enough to gain saturation of the average retrieval precision of just over 90% from most of the image queries. The performance of both R-DT-DWT and C-DT-DWT is compared to that of 4-

level Curvelet in [4] for 4th level of decomposition leading to slightly better results for the latter for a portion of tested images. Ciu et al. [5] propose modified DT-CWT with inclusion of hashes for feature generation. Testing database has smaller size in their study and only normalized Hamming distance presents the retrieval efficiency of the algorithm. In each of these studies, no measured or predicted retrieval times are reported, nor how the actual indexing of the database is performed.

In this paper, detailed analysis of the time consumption by the 2D DT-CWT to form feature vectors is presented in the next section, followed by description of two novel algorithms for both database indexing and image retrieval with adaptive selection of the decomposition level given required retrieval time and considering the information stored in the low-frequency sub-bands. In Section 4 experimental results are presented supporting the applicability of the proposed algorithms followed by a conclusion in Section 5.

2 2D DT-CWT Properties

2.1 Maximum Level of Decomposition and Stored Information Estimation

The coefficients of the two dimensional Dual-Tree Complex Wavelet Transform are obtained according to [6]:

$$c(m,n) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} I(p,q)\phi(p-m, q-n),, \quad (1)$$

$$d(j,m,n) = 2^j \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} I(p,q)\psi(2^j p - m, 2^j\, q - n), \quad (2)$$

where *I(p,q)* is the image intensity with *p* and *q* – the coordinates of the current pixel, *p* □ *{0, P-1}*, *q* □ *{0,Q-1}*; ψ and ϕ – the wavelet and scale functions; *m* and *n* – the positions of the wavelet coefficients within the spectrum *c* (wavelet) and *d* (scale); *j* – scale factor.

The decomposition structure in l=2 levels is given in Fig. 1 [7]. According to it and as well as to the invariance to translation [8], which determines that the first level of the scheme includes filters of a different order (let denote them by (α, β)) from those in the following levels (let them be (γ, δ)), one pixel for l levels of full decomposition takes the time:

$$t_f = 10[(\alpha+\beta)t_* + 2(\alpha+\beta-2)t_+] + 12(l-1)[(\gamma+\delta)t_* + 2(\gamma+\delta-2)t_+], \quad (3)$$

where t^* is the time required to perform a multiplication of fractions, and t_+ - the time to perform a single addition for the resulting products.

The time needed for estimating only the approximation coefficients at level l is:

$$t_\alpha = 3[(\alpha+\beta)t_* + 2(\alpha+\beta-2)t_+] + 4(l-1)[(\gamma+\beta)t_* + 2(\gamma+\beta-2)t_+]. \quad (4)$$

Kingsbury [8] proposes sets of digital filters satisfying the condition for decomposition (1) and (2). The simplest set includes (5, 3) LeGall filters for l=1 and 6-tap Q-shift filters for the next levels. Thus, selecting these filters for testing α=5, β=3,

γ=δ=6, denoting with T_l the average allowable time for retrieval per pixel and using only the approximation coefficients for feature vectors formation the maximum level of decomposition is:

$$l_{max} = \left\lfloor \frac{T_l - 24t_* - 9t_+}{48t_* + 20t_+} + 1 \right\rfloor. \tag{5}$$

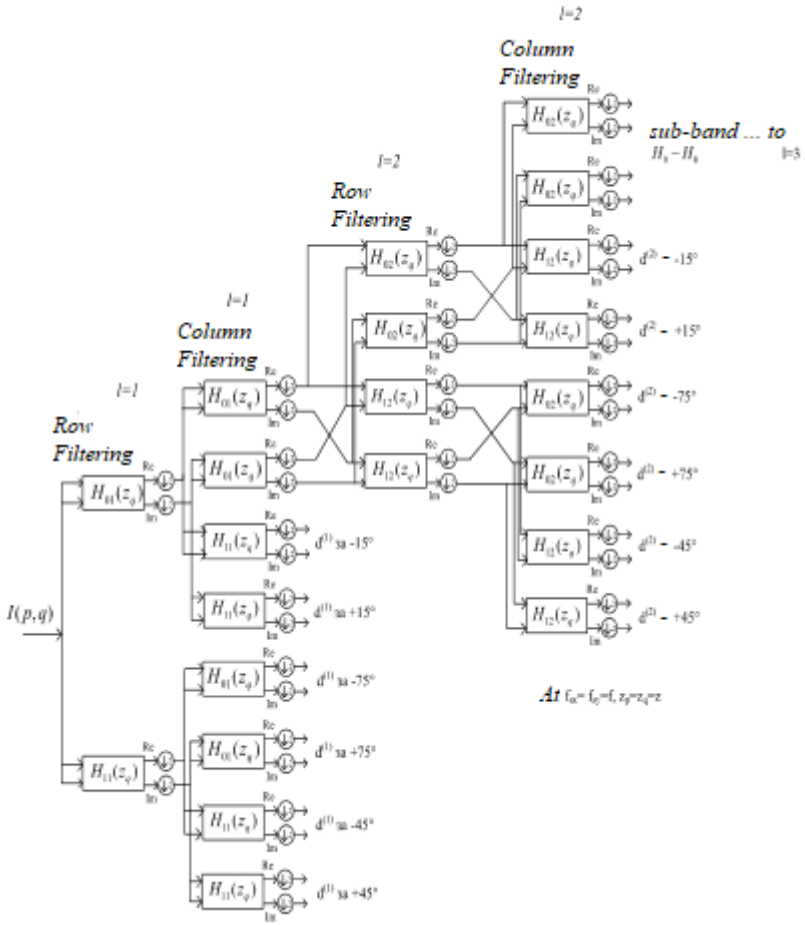

Fig. 1. 2D DT-CWT decomposition in two levels

The spatial position of each sub-band within the wavelet spectrum is given in Fig. 2. Transition from two-dimensional spatial coordinate system (p, q) and the related one from spectral domain (m, n) to one-dimensional representation could be done via:

$$\begin{aligned} v &= px + q, \\ w &= my + n. \end{aligned} \tag{6}$$

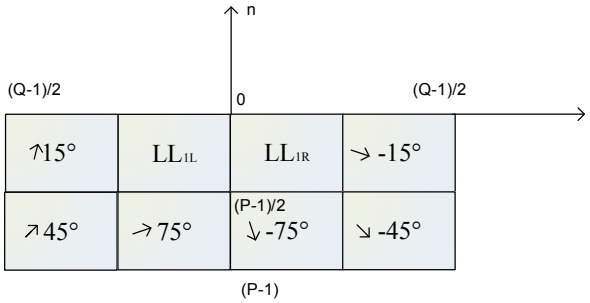

Fig. 2. Spectrum sub-bands obtained by 2D DT-CWT

Thus, the approximation coefficients from the low-frequency sub-bands processed only by rows for l=1 (Fig. 1) are:

$$\begin{aligned} d_r'^{(l)}[w] &= \sum_{v=0}^{x.y} h_0^{(l)}[2v-w]I[v], \\ d_c'^{(l)}[w] &= \sum_{v=0}^{x.y} h_0'^{(l)}[2v-w]I[v]. \end{aligned} \tag{7}$$

where $d_r'^{(l)}[w]$ and $d_c'^{(l)}[w]$ are their real and imaginary part. The full decomposition after processing by columns leads to:

$$\begin{aligned} d_{r1}''^{(l)}[w'] &= \sum_{w=0}^{\frac{x}{2}.y} d_r'^{(l)}[w]h_0^{(l)}[2w-w'], \\ d_{c1}''^{(l)}[w'] &= \sum_{w=0}^{\frac{x}{2}.y} d_r'^{(l)}[w]h_0'^{(l)}[2w-w'], \\ d_{r2}''^{(l)}[w'] &= \sum_{w=0}^{\frac{x}{2}.y} d_c'^{(l)}[w]h_0^{(l)}[2w-w'], \\ d_{c2}''^{(l)}[w'] &= \sum_{w=0}^{\frac{x}{2}.y} d_c'^{(l)}[w]h_0'^{(l)}[2w-w']. \end{aligned} \tag{8}$$

Both low-frequency sub-bands contain complex coefficients represented as:

$$\begin{aligned} \dot{d}_1''^{(l)} &= d_{r1}''^{(l)} + jd_{c1}''^{(l)}, \\ \dot{d}_2''^{(l)} &= d_{r2}''^{(l)} + jd_{c2}''^{(l)}. \end{aligned} \tag{9}$$

For $l \geq 2$ analogous to (7) and (8) derivations could be done, so the energy contained in the approximation coefficients from level l is:

$$E_a^{(l)} = \left(d_{r1}''^{(l)}\right)^2 + \left(d_{c1}''^{(l)}\right)^2 + \left(d_{r2}''^{(l)}\right)^2 + \left(d_{c2}''^{(l)}\right)^2, \tag{10}$$

while the total energy of the image is:

$$E = \sum_{p=0}^{x-1}\sum_{q=0}^{y-1} I^2(p,q). \tag{11}$$

The portion of information carried by the approximation coefficients then is:

$$S_l = -log_2 \frac{E_a^{(l)}}{E}. \tag{12}$$

2.2 Similarity Estimation Time

Let $I_1(p,q)$ and $I_2(p,q)$ are two images of equal size $p \in \{0,P-1\}$ and $q \in \{0,Q-1\}$. After l levels of decomposition, the area occupied by the approximation coefficients is:

$$\left[m \in \left\{\frac{2^l-1}{2^{l+1}}(P-1), \frac{2^l+1}{2^{l+1}}(P-1)\right\}, n \in \left\{\frac{2^l-1}{2^l}(Q-1), (Q-1)\right\}\right]. \quad (13)$$

The feature vectors have the following components:

$$\vec{\xi}^{(l)} = [d''^{(l)}\left(\frac{2^l-1}{2^{l+1}}(P-1), \frac{2^l-1}{2^l}(Q-1)\right), d''^{(l)}\left(\frac{2^l-1}{2^{l+1}}(P-1), \frac{2^l-1}{2^l}(Q-1)+1\right), \ldots,$$
$$d''^{(l)}\left(\frac{2^l-1}{2^{l+1}}(P-1), (Q-1)\right), d''^{(l)}\left(\frac{2^l-1}{2^{l+1}}(P-1)+1, \frac{2^l-1}{2^l}(Q-1)\right), \ldots, \quad (14)$$
$$d''^{(l)}\left(\frac{2^l-1}{2^{l+1}}(P-1), (Q-1)\right)].$$

The later are totally in number:

$$\eta^{(l)} = \left\{\left[\frac{2^l+1}{2^{l+1}} - \frac{2^l-1}{2^{l+1}}\right]P\right\}\left[Q - \frac{2^l-1}{2^l}Q\right] = \frac{PQ}{2^{2l}}. \quad (15)$$

In this study, the Hausdorff distance is selected as a measure between the vectors based on its advantages assuring higher retrieval accuracy [9]. It is found by:

$$d_H^{(l)}(I_1, I_2) = \max\left(h_{I_1,I_2}^{(l)}(\cup_{m,n}\vec{\xi}_{I_1}^{(l)}, \cup_{m,n}\vec{\xi}_{I_2}^{(l)}), h_{I_2,I_1}^{(l)}(\cup_{m,n}\vec{\xi}_{I_2}^{(l)}, \cup_{m,n}\vec{\xi}_{I_1}^{(l)})\right), \quad (16)$$

where $h_{I_1,I_2}^{(l)} = \max_{\vec{\xi}_{I_1}^{(l)}} \min_{\vec{\xi}_{I_2}^{(l)}} \left\|\vec{\xi}_{I_1}^{(l)} - \vec{\xi}_{I_2}^{(l)}\right\|$ is the directional Hausdorff distance from the set of $\vec{\xi}_{I_1,m,n}^{(l)}$ to the set of $\vec{\xi}_{I_2,m,n}^{(l)}$ for all m and n within the range $\left\{m \in \left(\frac{x}{4}, \frac{3x}{4}\right), n \in \left(\frac{y}{2}, y\right)\right\}$. The norm of the directional distance is found by using the Euclidean distance. For each point in the feature space all distances to all the other points belonging to I_2 are calculated and then the minimal is selected. The process leads to $\eta^{(l)}$ calculated distances each of which takes 2 multiplications, 1 addition and 1 square root. The minimal value is found by binary search in this case which needs $log_2(\eta^{(l)})$ comparisons.

For all the other points from I_1 all steps are repeated and it consumes time equal to:

$$t_{h_{I_1,I_2}^{(l)}} = \eta^{(l)}\left[\eta^{(l)}(2t_* + t_+ + t_r) + t_>\right] + t_>, \quad (17)$$

where $t_>$ is the time needed to find the maximum among all minimums according to the expression for directional Hausdorff distance. The related distance $h_{I_1,I_2}^{(l)}$ is calculated in analogous fashion where $t_{h_{I_1,I_2}^{(l)}} = t_{h_{I_2,I_1}^{(l)}}$. To all these time components a time needed to find the maximum from (16) is added, which is $t_{>DH} = 1$ comparisons or the global time for similarity estimation based on content between the two images is:

$$t_H^{(l)} = 2\{\eta^{(l)}[\eta^{(l)}(2t_* + t_+ + t_r) + t_>] + t_>\} + t_{>DH}. \quad (18)$$

2.3 Database Indexing Time

Let's have database comprising of B images, all of PxQ pixels in size. A comparison based on content similarity takes place between a query and all of them. At selected level of decomposition l the time needed for calculating all feature vectors according to (4) is:

$$t_{DB} = BPQt_a. \quad (19)$$

Just for the query it is $t_q = PQt_a$. On the other hand the time needed to find similarity with all images from the database and returning them as results in order of relevance at rank R is:

$$t_{DB} = Bt_H + log_2(B). \quad (20)$$

The logarithm component relates to the sorting time of the results. Only R of them output as indexes.

3 Proposed Algorithms

3.1 Algorithm for Indexing Image Database

The steps for feature vectors estimation of the images from the database of the CBIR system using 2D DT-CWT are given in Fig. 3. The aim is to find the level of decomposition l_{end} at which the preserved information (energy) is at least as high as preliminary set one. For all levels, $l \leq l_{end}$ the feature vectors are found and stored for all B images within the same database.

The algorithm takes as an input the whole content of the database. If necessary, normalization of the size of particular images could be done using the bi-cubic interpolation for better preserving the original shape of objects. Then, the minimum allowed energy of the approximation coefficients is being set and starting from level 1, iteratively, the 2D DT-CWT is applied from the spectrum of which as a result the actual approximation coefficients are extracted (contained in the LL (L - Low-frequency) sub-bands). They cumulative energy is found and when all images of the database have been passed over, their average energy is also found. For each level of decomposition l the feature vectors of all images are being preserved in a separate record within the database. Then, comes the comparison with the currently registered amount of energy and if it is still higher than the preliminary set threshold, further decomposition takes place one level forward. Once the condition has been met, then the reached level is registered indicating the ending point of indexing the database.

The use of the average energy at each iteration of the indexing process for the whole database assures that any significant deviation caused by particular content leading to more evenly distributed spectrum will not affect the average retrieval accuracy for the entire collection of scenes when ever growing number of queries are passed to the system.

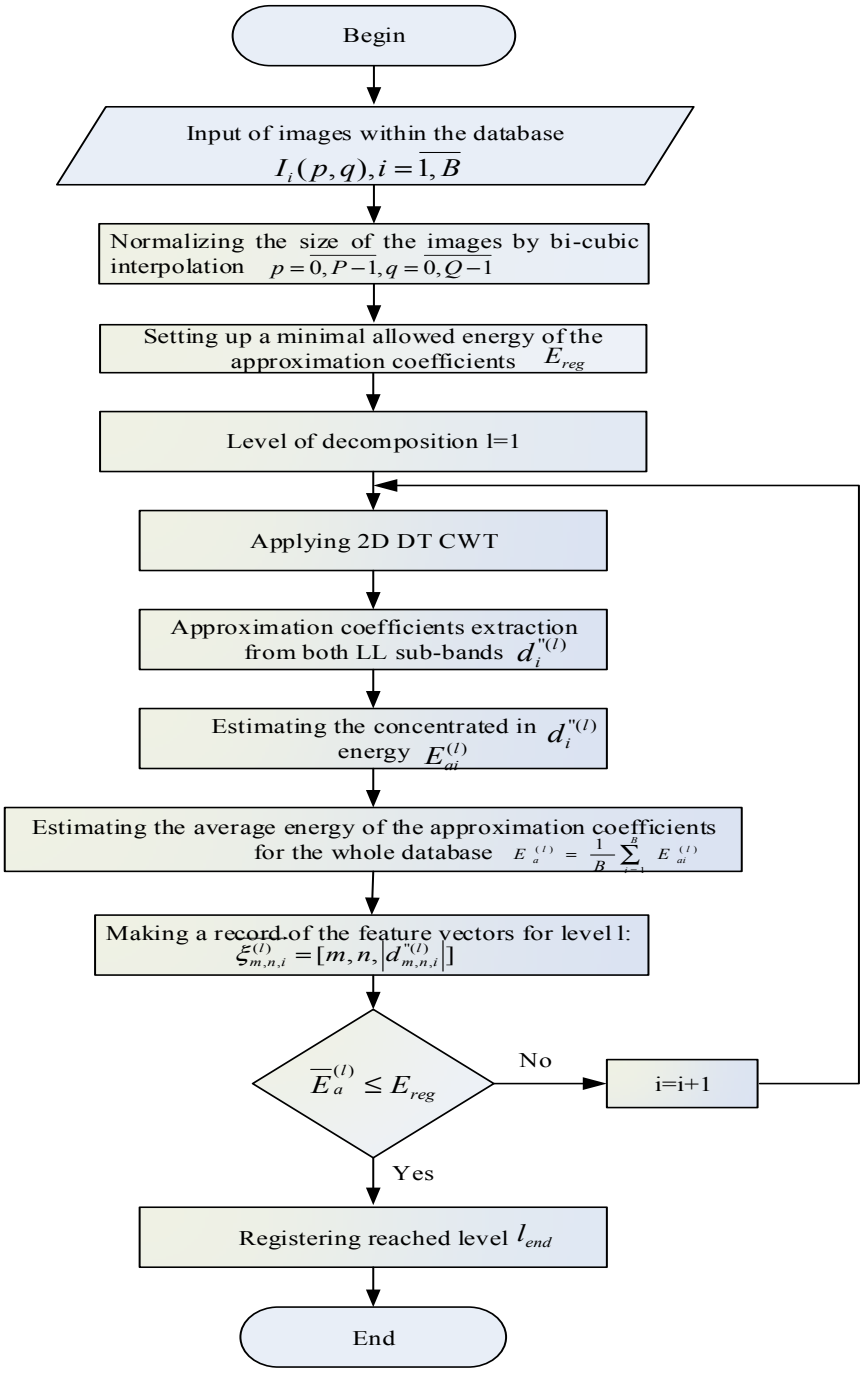

Fig. 3. Indexing algorithm

3.2 Algorithm for CBIR by Set Information Size

The algorithm is given in Fig. 4. It includes query *I(p, q)* size normalization to *PxQ* pixels – the same as all images from the database. By given time t_{req} for similarity comparison based on content with all *B* images, the maximum allowed level of decomposition l_{max} is set.

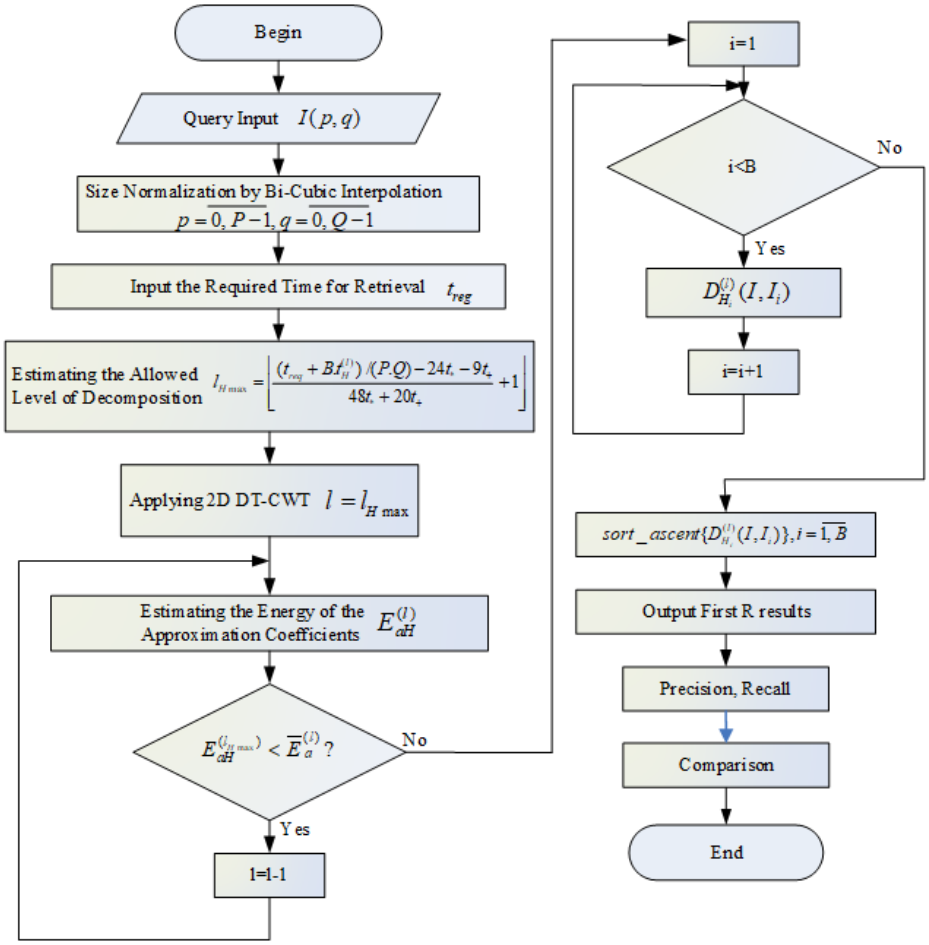

Fig. 4. CBIR algorithm

Then, a check follows for the quantity of energy inside the LL sub-bands of the wavelet spectrum and afterwards a comparison with preliminary defined value $\overline{E}_a^{(l)}$. If $E_a^{(l_{max})} \geq \overline{E}_a^{(l)}$, it follows a comparison with the feature vectors at level l_{max}. Otherwise, feature vectors from upper level *l* such that the condition for the energy to be satisfied. The estimation of $\overline{E}_a^{(l)}$ is done as the average energy from the whole database of images, at such a level that verification of the retrieval accuracy does not seem to produce considerable increase.

While analyzing the accuracy with the use of different metrics and taking into account the rank *R*, the resulting values for *Precision* and *Recall* could be compared for these metrics. This approach may lead to a selection of a better metric for particular application.

4 Experimental Results

The experimental testing of the proposed algorithms is done on a PC compatible workstation with Intel Core 2 Duo CPU running at 2.4 GHz within Matlab R2008b environment. Wang database [10] is used containing 1000 RGB raster images with size 256x384 and 384x256 pixels at 24 bpp. They are equally divided into 10 categories by content. During testing all of them are resized to 256x256 pixels using bicubic interpolation. Fig. 5 shows the mean value of the energy ratio concentrated in the sub-band of the approximation coefficients to the total energy of the image by decomposition levels.

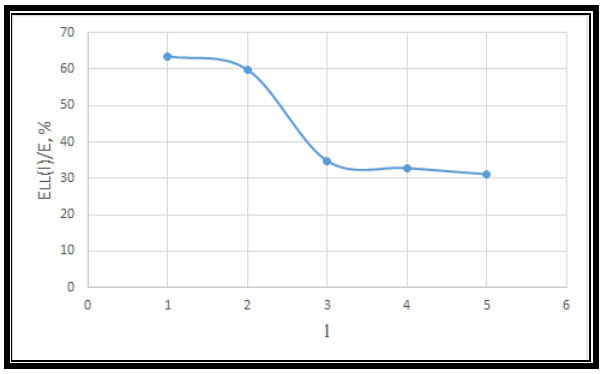

Fig. 5. Relative portion of the energy stored into the approximation coefficients

Given the times for executing a multiplication, addition and comparison between two numbers for the used CPU [11] and using (3), (4), (17) and (18) it becomes possible to predict the time needed to accomplish a similarity search at levels 1-4 between two images from the database. These times are also experimentally measured over the testing platform. Their values along with the absolute and relative errors registered are given in Table 1.

Practically applicable from a user point of view (the average home user) with respect to the retrieval time with the used hardware platform is second, third and fourth decomposition level. The average distance estimation time for them, when full validation is made over all 1000 images from the database, is 3.61 sec, 0.27 sec. and 0.10 sec, respectively. The average times at different ratios of the number of requests decomposed to the second, third and fourth levels are shown in Fig. 6. The dependency shown can be used as a calibration curve to find an appropriate relative number of

images with low S_l for decomposing to a lower level l for more accurate retrieval when tuning a CBIR system.

Table 1. Predicted and measured times for similarity search

l	Measured time, sec	Predicted time, sec	Absolute error, sec	Relative error, %
1	41.6618	26,1724	15,4894	37,18
2	3.6144	1,6358	1,9787	54,74
3	0.2742	0,1022	0,1719	62,71
4	0.1016	0,0639	0,0377	37,12

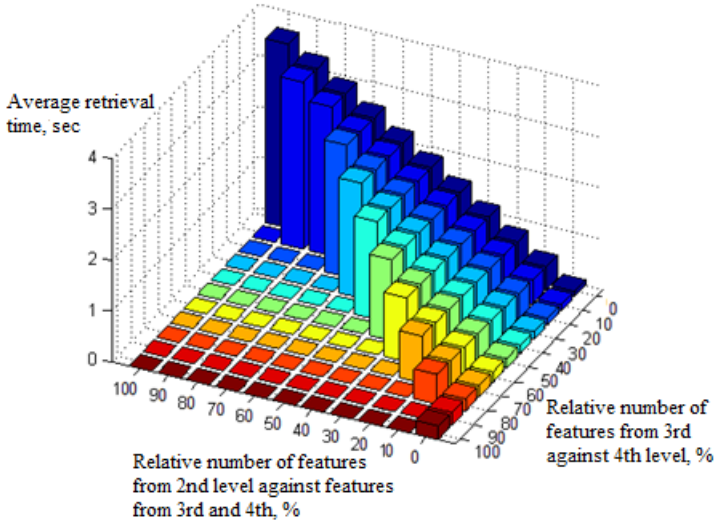

Fig. 6. Influence of the relative number of features by level over the average retrieval time

The overall retrieval accuracy of the proposed algorithm with those of 3 others – R-DT DWT, C-DT DWT and 4-level Curvelet [4] is compared by the average precision achieved (Table 2).

For the categories 'Africa' and 'Buses', for which inferior values of the retrieval accuracy are obtained by giving a lower priority to the retrieval time, it is possible to use coefficients from $l = 3$ for the formation of the features. According to the calibration curve of Fig. 6 and taking into account the relative number of images in these two categories - 20% of the total size of the database, it follows that the average retrieval time would increase from 0.1 sec. 0.5 sec. It remains within the scope of practical relevance to the end user, which proves the applicability of the proposed approach.

As the data from Table 2 shows, for five of the categories from the database - Dinosaurs, Elephants, Horses, Roses and Nature the proposed algorithm has higher or equal average precision compared to the other 3 algorithms. In particular, the images

from the Dinosaurs and Roses groups have lower within-class variability which leads to Average Precision of 100%. For the 'Social Life' and Food' categories, it shows a lower but comparable value with decrease of only 3.25% and 4.5% relative to the 4-level Curvelet algorithm. For the rest three categories higher level of decomposition needs to be implemented to overcome the lower resulting precision. The algorithms applicable in their present form are thought to have potential for navigation applications such as those described in [12].

Table 2. Retrieval accuracy comparison with other algorithms

Category	Average Precision, %			
	R-DT DWT, [4]	C-DT DWT, [4]	4-level Curvelet, [4]	Proposed
Africa	45,5	46	44,75	18,2
Buses	53,5	54,5	55	16,7
Dinosaurs	100	100	100	100
Elephants	66,5	67	68,5	80
Horses	62,5	63,25	65	70
Roses	95	95,5	96	100
Nature	52	52,5	54	60
Social Life	51,5	53	53,25	50
Food	43,5	44,25	44,5	40
Architecture	54,5	54,5	56	30

5 Conclusion

In this paper novel algorithms for image database indexing and content-based retrieval are proposed using the 2D DT-CWT with adaptive selection of the level of decomposition and accounting for the information contained into the approximation coefficients' sub-bands. The indexing and similarity estimation times are derived for arbitrary level given the size of the images and the size of the database. Having a particular retrieval time set as demand by specific application it is possible to select appropriate level for feature vectors construction and to perform the retrieval. If the required processing time for a given retrieval rank allows retrieval accuracy could be increased by using lower level coefficients from the spectrum containing more information which is applicable for certain image categories hard to be retrieved at higher levels of decomposition. Experimental results support the applicability of the proposed algorithms reaching accuracy levels comparable with the R-DT DWT, C-DT DWT and 4-level Curvelet algorithms where for half of the test database the first outperform them.

Acknowledgement

This work was supported by the National Science Fund at the Ministry of Education and Science of Republic of Bulgaria under project KP-06-H27/16 17.12.2018 "Development of efficient methods and algorithms for tensor-based processing and analysis of multidimensional images with application in interdisciplinary areas".

References

1. Patil, S., Talbar, S.: Image retrieval using dual tree complex wavelet transform. In: Choudhary, R, Verma, M., Saini, S. (eds.) 5TH IEEE INTERNATIONAL CONFERENCE ON ADVANCED COMPUTING & COMMUNICATION TECHNOLOGIES [ICACCT-2011], November 5th, Panipat, pp. 210-215, ABC Group of Publication, Karnal, India (2011).
2. Vhanmane, S., Sangve, S.: Comparative analysis of discrete wavelet transform and complex wavelet transform for image retrieval. International Journal of Innovative Science, Engineering & Technology (IJISET) 1(6), 403-408 (2014).
3. Patil, P., Kokare, M.: Interactive semantic image retrieval. J Inf Process Syst 9(3), 349-364 (2013).
4. Patil, S., Talbar, S.: Multi resolution analysis using complex wavelet and curvelet features for content based image retrieval. International Journal of Computer Applications 47(17), 6-10 (2012).
5. Cui, D., Liu, Y., Zuo, J., Xu, B.: A modified image retrieval algorithm based on DT-CWT. Journal of Computational Information Systems 7(3), 896-903 (2011).
6. Selesnick, I., Baraniuk, R., Kingsbury, N.: The dual-tree complex wavelet transform - a coherent framework for multiscale signal and image processing. IEEE Signal Processing Magazine 22(6), 123-151 (2005).
7. Kingsbury, N.: Dual tree complex wavelets, HASSIP Workshop, September 2004, https://www-sigproc.eng.cam.ac.uk/foswiki/pub/Main/NGK/HASSIPtalk04.pdf, last accessed 2018/12/11.
8. Kingsbury, N.: Complex wavelets for shift invariant analysis and filtering of signals. Journal of Applied and Computational Harmonic Analysis 10(3), 234-253, (2001).
9. Vetova, S., Draganov, I., Ivanov, I., Mladenov, V.: CBIR efficiency enhancement using local features algorithm with Hausdorff distance. In Proceedings of the 2017 EUROPEAN CONFERENCE ON ELECTRICAL ENGINEERING AND COMPUTER SCIENCE (EECS'17), vol. 5, pp. 116-123. Bern, Switzerland (2017).
10. Wang, J., Li, J., Wiederhold, G.: SIMPLIcity: Semantics-sensitive integrated matching for picture libraries. IEEE Transactions on Pattern Analysis and Machine Intelligence 23(9), 947-963, (2001).
11. Processor GFlops Compilation, https://www.techpowerup.com/forums/threads/processor-gflops-compilation.94721/, last accessed 2018/12/12.
12. Brassai, S., Iantovics, B., Enachescu, C.: Optimization of robotic mobile agent navigation. Studies in Informatics and Control 21(4), 403-412, (2012).