

Research Article

Object Recognition Based on Dual Tree Complex Wavelet Transform

S. Elakkiya and S. Audithan

PRIST University, Tanjore, Tamilnadu, India

Abstract: Automated recognition of objects from images plays an important role in many computer vision systems such as robot navigation, object manipulation and content based image retrieval. In this study, an approach for object recognition based on Dual Tree Complex Wavelet Transform (DTCWT) is proposed. The proposed approach attempts to extract the detailed information of objects from the multi scale representation by DTCWT. The proposed system is tested on Columbia Object Image Library (COIL-100). All the objects are considered for the classification based on nearest neighbor classifier. The results show that the maximum recognition accuracy achieved by the proposed approach is 97.03%.

Keywords: Dual tree complex wavelet transform, nearest neighbor classifier, object recognition, wavelet transform

INTRODUCTION

Object recognition is to find a given object in an image or video sequence. Large amount of objects in images are recognized by human with little effort, despite the fact that the image of the objects may vary by different viewpoints, different sizes/scale or even when they are translated or rotated. This task is still a challenge for computer vision systems in general. Yang *et al.* (2012) presented a Group Sensitive Multiple Kernel Learning (GSMKL) method for object recognition to accommodate the intra-class diversity and the inter-class correlation. The performance of GSMKL does not significantly vary with different grouping strategies. A simple hybrid grouping strategy can boost GSMKL against other multiple kernel methods.

Choi *et al.* (2012) presented an efficient model that captures the contextual information among more than a hundred object categories using a tree structure. This tree based context model improves object recognition performance and provides a coherent interpretation of a scene, which enables a reliable image querying system by multiple object categories. New data sets with images that contain many instances of different object categories are used. Color descriptions from distinct regions covering multiple segments are considered for object representation in Naik and Murthy (2007). Distinct multicolored regions are detected using edge maps and clustering.

A method for predicting fundamental performance of object recognition is described in Boshra and Bhanu (2000). It considers data distortion factors such as uncertainty, occlusion and clutter, in addition to model similarity. It consistently predicts reasonably tight

bounds on actual performance. A multi-linear supervised neighbourhood embedding for discriminant feature extraction for object recognition is described in Han *et al.* (2012). A local descriptor tensor is used to represent an image and used for subject or scene recognition. A novel approach to measure similarity between shapes and exploit it for object recognition is presented in Belongie *et al.* (2002). The measurement of similarity is preceded by two methods; solving for correspondences between points on the two shapes and using the correspondences to estimate an aligning transform.

New color models, which are analyzed in theory and evaluated in practice for the purpose of recognition of multi colored objects invariant to a substantial change in viewpoint, object geometry and illumination is explained in Gevers and Smeulders (1999). A new representation by sparse function approximation in both spatial dimensions together with the orientation dimension is presented in Pham and Smeulders (2006). This new representation is able to exploit a rich amount of a priori information about the object views.

A new scheme that merges color and shape-invariant information for object recognition is presented in Diplaros *et al.* (2006). To obtain robustness against photometric changes, color-invariant derivatives are computed first. These color invariant derivatives are used to obtain similarity invariant shape descriptors. The matching function of the color-shape context allows for fast recognition even in the presence of object occlusion and cluttering.

A novel approach for parts-based object representation is described in Amores *et al.* (2007). The image is built by a collection of correlograms, where

each one represents specific attributes localized at the same time in several parts of the image holding a specific spatial relationship. A set of composed histogram features of higher dimensionality is explained in Linde and Lindeberg (2004), which give significantly better recognition performance compared to the histogram descriptors of lower dimensionality. The use of histograms of higher dimensionality is made possible by a sparse representation for efficient computation and handling of higher-dimensional histograms.

A validation and rotation invariant object recognition is described in Kim *et al.* (2012). By using difference of gaussian filter and local adaptive binarization, a binary image reserving spotless object boundaries is achieved. An object region from surroundings is extracted with remunerated edges that reserves geometry information of object. Neural network is used to recognize the object. A novel method for object category recognition by improving the popular bag-of-words methods is presented in Wang *et al.* (2010). To obtain the global spatial features, a fast method is proposed to generate the semantic meaningful object parts by exploiting the geometric position distribution of the local salient regions. The multi kernel learning framework is adopted to integrate the extracted features in an optimal way.

METHODOLOGY

The proposed approach for object recognition is built based on Dual Tree Complex Wavelet Transform (DTCWT). A brief description about DTCWT is as follows: Although the Discrete Wavelet Transform (DWT) in its maximally decimated form (Mallat's dyadic filter tree) has established an impressive reputation as a tool for image compression its use for other signal analysis and reconstruction tasks such as image restoration and enhancement has been hampered by two main disadvantages:

- **Lack of shift invariance:** This means that small shifts in the input signal can cause major variations in the distribution of energy between DWT coefficients at different scales.
- Poor directional selectivity for diagonal features because the wavelet filters are separable and real.

A well-known way of providing shift invariance is to use the undecimated form of the dyadic filter tree but this suffers from increased computation requirements and high redundancy in the output information making subsequent processing expensive too. However, the Dual-Tree Complex Wavelet Transform (DT CWT)

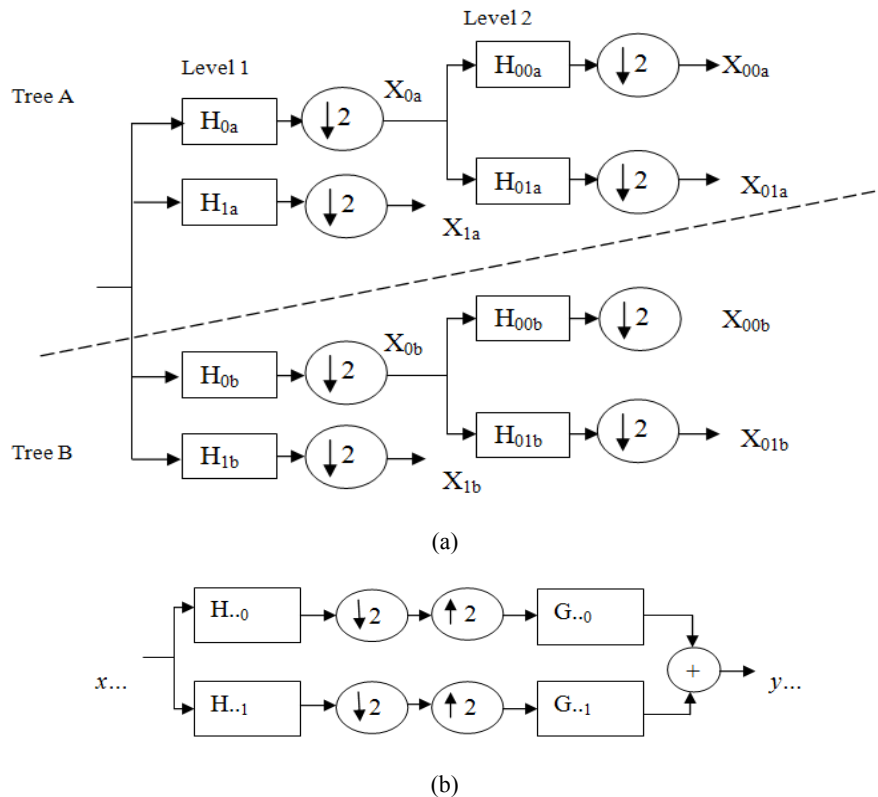


Fig. 1: (a) Band reconstruction block (b) dual tree of filters for the complex wavelet transform

with the following properties (Selesnick *et al.*, 2005; Kingsbury, 1998):

- Approximate shift invariance
- Good selectivity and directionality in 2-Dimensions (2D) with Gabor-like filters (also for higher dimensionality)
- Perfect Reconstruction (PR) using short linear phase filters
- Limited redundancy, independent of the number of scales = 2:1 for 1-D, 2 m:1 for m-D
- Efficient order-N computation-only 2 m times the simple DWT for m-D

DTCWT for one dimensional signal: Unfortunately it was unable to obtain PR and good frequency characteristics using short support complex FIR filters in a single tree (Fig. 1 Tree A). This is because the complex filters in order to be useful should be designed to emphasize positive frequencies and reject negative frequencies (or vice-versa) and it is then not possible for the 2-band reconstruction block to have a flat overall frequency response as required if $y = x$.

It is possible to achieve approximate shift invariance with a real DWT by doubling the sampling rate at each level of the tree. For this to work the samples must be evenly spaced. Hence all the sampling rates in Tree A of Fig. 1 are doubled, by eliminating the down-sampling by 2 after the level-1 filters H_{0a} and H_{1a} . This is equivalent to two parallel fully-decimated trees A and B provided that the delays of H_{0b} and H_{1b} are one sample offset from H_{0a} and H_{1a} . To get uniform intervals between samples from the two trees below level-1, the filters in one tree must provide delays that are half a sample different (at the filter input rate) from

those in the other tree. For linear phase this requires odd-length filters in one tree and even-length filters in the other. This is probably the most novel aspect of the dual-tree transform. Greater symmetry between the two trees occurs if each tree uses odd and even filters alternately from level to level.

Extension to two dimensions: Extension to 2-D is achieved by separable filtering along columns and then rows. However, if column and row filters both suppress negative frequencies then only the first quadrant of the 2-D signal spectrum is retained. Two adjacent quadrants of the spectrum are required to represent fully a real 2-D signal so filter with complex conjugates of the row filters also. This gives 4:1 redundancy in the transformed 2-D signal. A normal 2-D DWT produces three band pass sub images at each level, corresponding to low-high, high-high and high-low filtering. The 2-D CWT produces three sub images in each of spectral quadrants 1 and 2 giving six band pass sub images of complex coefficients at each level which are strongly oriented at angles of $\pm 15^\circ$, $\pm 45^\circ$, $\pm 75^\circ$ as shown from their Gabor-like impulse responses. The strong orientation occurs because complex filters can separate positive from negative frequencies vertically and horizontally.

Proposed method: The proposed object recognition system comprises of two steps. The first one is feature extraction and the next step is classification. In the first step, the dominant features that represents a particular object is extracted using DTCWT and then the extracted features are used later in the classification stage. Feature extraction is very important in the field of machine learning, pattern recognition and data

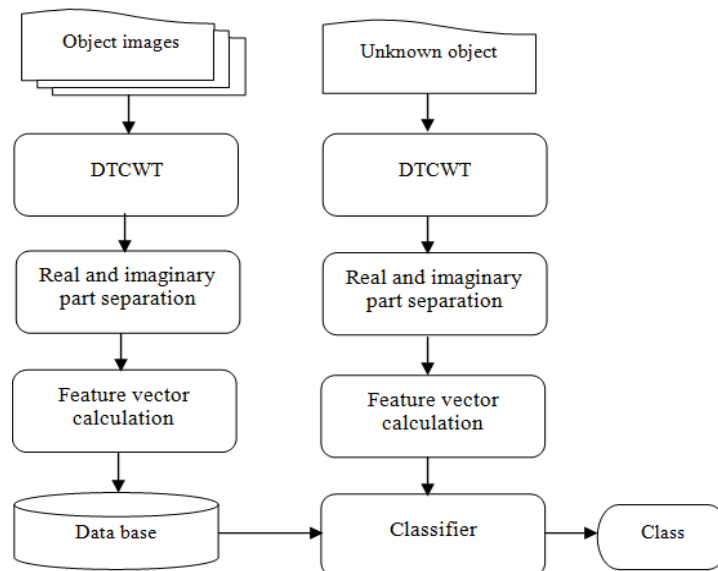


Fig. 2: Proposed object recognition system based on DTCWT

mining. At this stage, the patterns or information that exists in the given image is analyzed. As the performance of the classifier mainly depends on the extracted features, the extracted features must be distinguished between different objects used in the experiment. The fully automated system for object recognition is shown in Fig. 2.

In the proposed approach, the objects are represented by DTCWT at various scales initially. The decomposition of objects by DTCWT creates sub-bands. As DTCWT is a multi-resolutional analysis, each and every sub-band in the decomposed image has various detailed information about the input image. Hence all the sub-bands are considered. The dimension of sub-band is equal to the input image, it is very difficult to identify or extract the dominant features. Hence, the energy of all the sub-bands is used as feature vectors. The energy is calculated by using Eq. (1):

$$E_k = \frac{1}{RC} \sum_{i=1}^R \sum_{j=1}^C |x_k(i, j)| \quad (1)$$

where, $x_k(i, j)$ is the pixel value of the k^{th} sub-band and R, C is width and height of the sub-band respectively. The same procedure is repeated for the training set and the extracted features are stored in the database with their corresponding index/class for retrieval. The classification stage employs a nearest neighbour classifier to compare the extracted feature vectors against those in the database. The classifier gives the class or index of the recognized object.

RESULTS AND DISCUSSION

The proposed object recognition system based on DTCWT is tested on Columbia Object Image Library Dataset (COIL-100) Nene *et al.* (1996). This database is downloadable (Coil Database: <http://www.cs.columbia.edu/CAVE/software/softlib/coil-100.php>). A CCD color camera with a 25 mm lens was fixed to a rigid stand about 1feet from its base. A motorized turntable was placed about 2 feet from the base of the stand. The turntable was rotated through 360° and 72 images were taken per object; one at every 5° of rotation. The size of image in the database is 128×128 and the number of objects in the data base is 100. Figure 3 shows the objects in the COIL database.

In order to analyze the performance of the proposed system, the database is divided into two parts; training and testing set. The training set images are used to create the database and the testing set images are used for testing the accuracy of the proposed approach. The COIL database is divided based on the angles of object rotation. From the database, images captured at multiples of predefined angles of rotation (for example multiples of 10⁰) are grouped to form the training set and all the remaining captured images of the object make the testing set. Table 1 shows the recognition accuracy obtained by the proposed object recognition system. The proposed approach uses up to 6th level of DTCWT decomposition. The nearest neighbor classifier uses Euclidean distance as distance measure for classification.



Fig. 3: COIL database

Table 1: Recognition accuracy obtained by the proposed object recognition system

Level of decomposition	Recognition accuracy (%)					
	10 ⁰	20 ⁰	30 ⁰	45 ⁰	60 ⁰	90 ⁰
2	85.25	77.57	72.18	65.42	59.04	51.51
3	85.38	76.80	71.02	64.84	56.03	49.71
4	87.91	79.72	75.15	68.45	62.17	56.63
5	93.52	87.31	81.30	73.03	66.76	59.51
6	97.03	92.57	87.85	81.63	74.95	67.57

Table 2: Recognition accuracy of individual objects obtained at 6th level decomposition

Object	Accuracy (%)	Object	Accuracy (%)	Object	Accuracy (%)	Object	Accuracy (%)
Obj_1	100	Obj_26	100	Obj_51	100	Obj_76	94.44
Obj_2	100	Obj_27	88.89	Obj_52	100	Obj_77	100
Obj_3	100	Obj_28	100	Obj_53	100	Obj_78	100
Obj_4	100	Obj_29	100	Obj_54	97.22	Obj_79	91.67
Obj_5	100	Obj_30	100	Obj_55	100	Obj_80	97.22
Obj_6	97.22	Obj_31	94.44	Obj_56	100	Obj_81	100
Obj_7	100	Obj_32	100	Obj_57	97.22	Obj_82	94.44
Obj_8	94.44	Obj_33	100	Obj_58	100	Obj_83	100
Obj_9	100	Obj_34	100	Obj_59	100	Obj_84	86.11
Obj_10	100	Obj_35	100	Obj_60	100	Obj_85	100
Obj_11	97.22	Obj_36	100	Obj_61	100	Obj_86	100
Obj_12	100	Obj_37	100	Obj_62	100	Obj_87	100
Obj_13	86.11	Obj_38	100	Obj_63	97.22	Obj_88	100
Obj_14	100	Obj_39	100	Obj_64	100	Obj_89	100
Obj_15	86.11	Obj_40	100	Obj_65	91.67	Obj_90	100
Obj_16	100	Obj_41	100	Obj_66	100	Obj_91	66.67
Obj_17	100	Obj_42	100	Obj_67	80.56	Obj_92	100
Obj_18	100	Obj_43	100	Obj_68	91.67	Obj_93	100
Obj_19	94.44	Obj_44	77.78	Obj_69	66.67	Obj_94	100
Obj_20	100	Obj_45	100	Obj_70	100	Obj_95	100
Obj_21	75	Obj_46	100	Obj_71	100	Obj_96	100
Obj_22	94.44	Obj_47	100	Obj_72	100	Obj_97	100
Obj_23	86.11	Obj_48	100	Obj_73	100	Obj_98	83.33
Obj_24	100	Obj_49	100	Obj_74	100	Obj_99	100
Obj_25	100	Obj_50	100	Obj_75	97.22	Obj_100	97.22
Average							97.03

The results show that the maximum classification accuracy is obtained at 6th level of decomposition. Also, it is clearly noted that the classification accuracy increase as the level of decomposition increases. Table 2 shows the recognition accuracy of individual objects obtained at 6th level decomposition.

Among the 100 objects used in the experiment, 72 objects are classified accurately and only 28 objects are misclassified. The two objects obj_69 and obj_91 are poorly classified and the successful classification accuracy is only 66.67% i.e., among the 36 test samples in each category 12 samples are misclassified. Among the 12 misclassified samples, 5 samples in obj_69 are misclassified to obj_91 and 4 samples in obj_91 to obj_69. These two objects are the same car model with different colors. Hence, the accuracy of the proposed approach may be tuned by incorporating colour features in future.

CONCLUSION

In this study, an automated approach for object recognition based on DTCWT and nearest neighbor classifier is presented. The proposed approach uses sub-band energies of DTCWT as feature vector to represent the COIL database objects. The approach is tested in six different training set separated from the database based on the angle of object rotations. Results show that the proposed approach provides better recognition accuracy of 97.03% for the features extracted at 6th level of DTCWT decomposition and the recognition accuracy of 72% of objects in the COIL database are 100%.

REFERENCES

- Amores, J., N. Sebe and P. Radeva, 2007. Context-based object-class recognition and retrieval by generalized correlograms. *IEEE T. Pattern Anal.*, 29(10): 1818-1833.
- Belongie, S., J. Malik and J. Puzicha, 2002. Shape matching and object recognition using shape contexts. *IEEE T. Pattern Anal.*, 24(4): 509-522.
- Boshra, M. and B. Bhanu, 2000. Predicting performance of object recognition. *IEEE T. Pattern Anal.*, 22(9): 956-969.
- Choi, M.J., T. Antonio and S.W. Alan, 2012. A tree-based context model for object recognition. *IEEE T. Pattern Anal.*, 34(2): 240-252.
- Diplaros, A., T. Gevers and I. Patras, 2006. Combining color and shape information for illumination-viewpoint invariant object recognition. *IEEE T. Image Process.*, 15(1): 1-11.
- Gevers, T. and A.W.M. Smeulders, 1999. Color-based object recognition. *Pattern Recogn.*, 32(3): 453-464.
- Han, X.H., Y.W. Chen and X. Ruan, 2012. Multilinear supervised neighborhood embedding of a local descriptor tensor for scene/object recognition. *IEEE T. Image Process.*, 21(3): 1314-1326.
- Kim, K., K. Joongbae, K. Sangseung, K. Jaehong and L. Jaeyeon, 2012. Object recognition for cell manufacturing system. *Proceedings of the 9th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI)*, pp: 512-514.
- Kingsbury, N.G., 1998. The dual-tree complex wavelet transform: A new technique for shift invariance and directional filters. *Proceeding of the 8th IEEE DSP Work-shop*. Bryce Canyon.

- Linde, O. and T. Lindeberg, 2004. Object recognition using composed receptive field histograms of higher dimensionality. Proceedings of the 17th International Conference on Pattern Recognition, 2: 1-6.
- Naik, S.K. and C.A. Murthy, 2007. Distinct multicolored region descriptors for object recognition. *IEEE T. Pattern Anal.*, 29(7): 1291-1296.
- Nene, S.A., S.K. Nayar and H. Murase, 1996. Columbia object image library (COIL-100). Technical Report CUCS-006-96, February 1996.
- Pham, T.V. and A.W.M. Smeulders, 2006. Sparse representation for coarse and fine object recognition. *IEEE T. Pattern Anal.*, 28(4): 555-567.
- Selesnick, I.W., R.G. Baraniuk and N.G. Kingsbury, 2005. The dual-tree complex wavelet transform. *IEEE Signal Proc. Mag.*, 22(6): 123-151.
- Wang, M., W. Yanling, L. Guangda and Z. Xiang-Dong, 2010. Object recognition via adaptive multi-level feature integration. Proceedings of the 12th International Asia-Pacific on Web Conference (APWEB), pp: 253-259.
- Yang, J., T. Yonghong, D. Ling-Yu, H. Tiejun and G. Wen, 2012. Group-sensitive multiple kernel learning for object recognition. *IEEE T. Image Process.*, 21(5): 2838-2852.