Robust Feature Detection Using 2D Wavelet Transform under Low Light Environment

Youngouk Kim[1,2], Jihoon Lee[1], Woon Cho[1], Changwoo Park[2], Changhan Park[1], Joonki Paik[1]

[1]*Image Processing and Intelligent Systems Laboratory, Department of Image Engineering, Graduate School of Advanced Imaging Science, Multimedia, and Film, Chung-Ang University.*

[2]*Precision Machinery Center, Korea Electronics Technology Institute*
Korea

1. Introduction

Simultaneous localization and mapping (SLAM) requires multi-modal sensors, such as ultrasound, range, infrared (IR), encoder or odometer, and multiple visual sensors. Recognition-based localization is considered as the most promising method of image-based SLAM (Dissanayake, 2001). In practice, we cannot rely on the basic encoder output under kidnapping or shadowing environment. IR-LED cameras are recently used to deal with such complicated conditions. Map building becomes more prone to illumination change and affine variation, when the robot is randomly moving. The most popular solution for the robust recognition method is scale-invariant feature transform (SIFT) approach that transforms an input image into a large collection of local feature vectors, each of which is invariant to image translation, scaling, and rotation (Lowe, 2004). The feature vector is partially invariant to illumination changes and affine (or three-dimensional) projection. Such local descriptor-based approach is generally robust against occlusion and scale variance. In spite of many promising factors, SIFT has many parameters to be controlled, and it requires the optimum Gaussian pyramid for acceptable performance. Intensity-based local feature extraction methods cannot avoid estimation error because of low light-level noise (Lee, 2005). Corner detection and local descriptor-based methods fall into this category. An alternative approach is moment-based invariant feature extraction that is robust against both geometric and photometric changes. This approach is usually effective for still image recognition. While a robot is moving, the moment-based method frequently recognizes non-planar objects, and can hardly extract invariant regions under illumination change. This paper presents a real-time local keypoint extraction method in the two-dimensional wavelet transform domain. The proposed method is robust against illumination change and low light-level noise, and free from manual adjustment of many parameters. Fig 1 displays whole structure of this paper.

Source: Vision Systems: Segmentation and Pattern Recognition, ISBN 987-3-902613-05-9,
Edited by: Goro Obinata and Ashish Dutta, pp.546, I-Tech, Vienna, Austria, June 2007

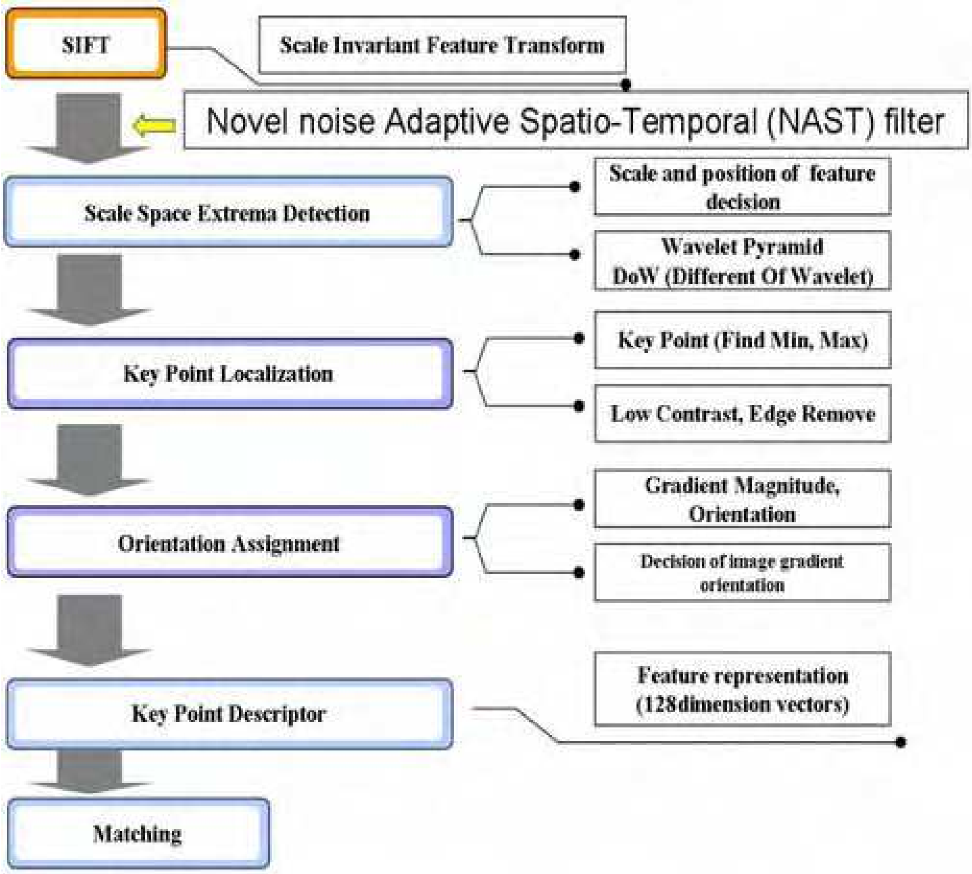


Figure 1. Conceptual flowchart of the whole structure

The paper is organized as follows. In section 2, noise adaptive spatio-temporal filter (NAST) is proposed to remove low light-level noise as a preprocessing step. Section 3 describes the proposed real-time local feature extraction method in the wavelet transform domain. Section 4 summarizes various experimental results by comparing DoW with SIFT methods, and section 5 concludes the paper.

2. Noise Adaptive Spatio-Temporal Filter

The proposed NAST algorithm adaptively processes the acquired image to remove low light level noise. Depending on statistics of the image, information of neighboring pixels, and motion, the NAST algorithm selects a proper filtering algorithm for each type of noise. A conceptual flowchart of the proposed algorithm is illustrated in Fig. 2. The proposed NAST algorithm has four different operations which are applied to the low light images.

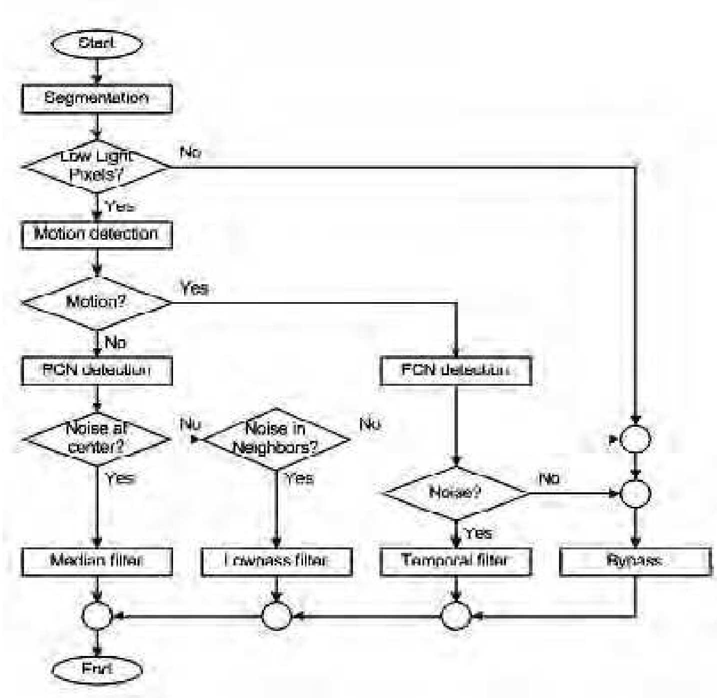


Figure 2. Conceptual flowchart of the proposed algorithm

2.1 Noise Detection Algorithm

The output of the noise detection block determines the operation of filtering blocks. The proposed spatial hybrid filter (SHF) can be represented as

$$y(i,j) = n(i,j) \times \hat{x}(i,j) + (1 - n(i,j)) \times x(i,j) \tag{1}$$

where $\hat{x}(i,j)$ represents a pixel filtered by the SHF and $n(i,j)$, which is the result of the noise detection process, takes 1 for the position of photon counting noise (PCN) pixels and 0 elsewhere. In equation (1), $x(i,j)$ and $y(i,j)$ denote the (i, j)-th pixels in noisy and filtered images, respectively. In the proposed noise detection scheme, $n(i,j)$ forms a binary noise map denoted by N, which is used to filter out uncorrelated noise and to indicate the reference points for the subsequent filtering of correlated noise.

2.2 Filtering Mechanism of SHF

If the central pixel in the window (W) is considered to be noise (i.e., $n(i,j)=1$in the noise map N), it is substituted by the median value of the window as a normal median filter. Then the noise cancellation scheme in SHF is extended to the correlated pixels in the local neighborhood ($x(i,j)$ where $n(i,j) \neq 1$ and at least one $n(k,l) = 1$ in W). In order to identify the correlated noise, the de-noised pixel value $x'(i,j)$ can be defined as

$$x'(i,j) = \frac{\sigma^2(i,j) \times x(i,j) + \overline{x}^2(i,j)}{\sigma(i,j) + \overline{x}(i,j)} \tag{2}$$

where $\overline{x}(i,j)$, and $\sigma^2(i,j)$ respectively represent the mean and variance of the window W.

2.3 Statistical Domain Temporal Filter (SDTF) for False Color Noise (FCN) Detection and Filtering

We use a new SDTF for removing FCN. The sum of the absolute differences (SAD) between the two working windows of consecutive frames is used for motion detection to avoid motion blur due to temporal averaging. Let $\hat{x}(i,j,t)$ and $\hat{x}(i,j,t-1)$ denote intensity values at the (i, j)-th pixel in the spatially filtered frames at time t and $t-1$, respectively, the proposed temporal filter can then be realized as

$$y(i,j,t) = \begin{cases} \hat{x}(i,j,t-1), & S_T(i,j,t) > S_T(i,j,t-1) \\ \hat{x}(i,j,t), & S_T(i,j,t) \leq S_T(i,j,t-1) \end{cases} \tag{3}$$

where $y(i,j,t)$ represents the final result of the proposed NAST and S_T is the local statistics defined as

$$S_T(i,j,t) = \left| \left(x(i,j,t) - \overline{x}(i,j,t) \right)^2 - \sigma^2(i,j,t) \right| \tag{4}$$

3. A New Method for Local Feature Detector Using 2D Discrete Wavelet Transform

In this section 2D discrete wavelet transform is briefly described as a theoretical background (Daubechies, 1998). Based on theory and implementation of 2D discrete wavelet transform, the DoW-based local extrema detection method is presented.

3.1 Characteristics of 2D Wavelet Transform

Human visual characteristics are widely used in image processing. One example is the use of Laplacian pyramid for image coding. SIFT falls into the category that uses Laplacian pyamid for scale-invariant feature extraction [3]. On the other hand wavelet transform is a multiresolution transform that repeatedly decompose the input signal into lowpass and highpass components like subband coding [7,8]. Wavelet-based scale-invariant feature extraction method does not increase the number of samples in the original image, which is the case of the Gaussian pyramid-based SIFT method. Wavelet transform can easily reflect human visual system by multiresolution analysis using orthogonal bases[12]. Because the wavelet-based method does not increase the number of samples, computational redundancy is greatly reduced, and its implementation is suitable for parallel processing.

3.2 Difference of Wavelet in the Scale Space

Most popular wavelet functions include Daubechies [7] and biorthogonal wavelet [10]. Although Daubechies designed a perfect reconstruction wavelet filter, it does not have symmetry. In general image processing applications symmetric biorthogonal filter is particularly suitable[10], but we used Daubechies coefficient set{DB2, DB10, DB18, DB26, DB34, DB42} for just efficient feature extraction purpose.

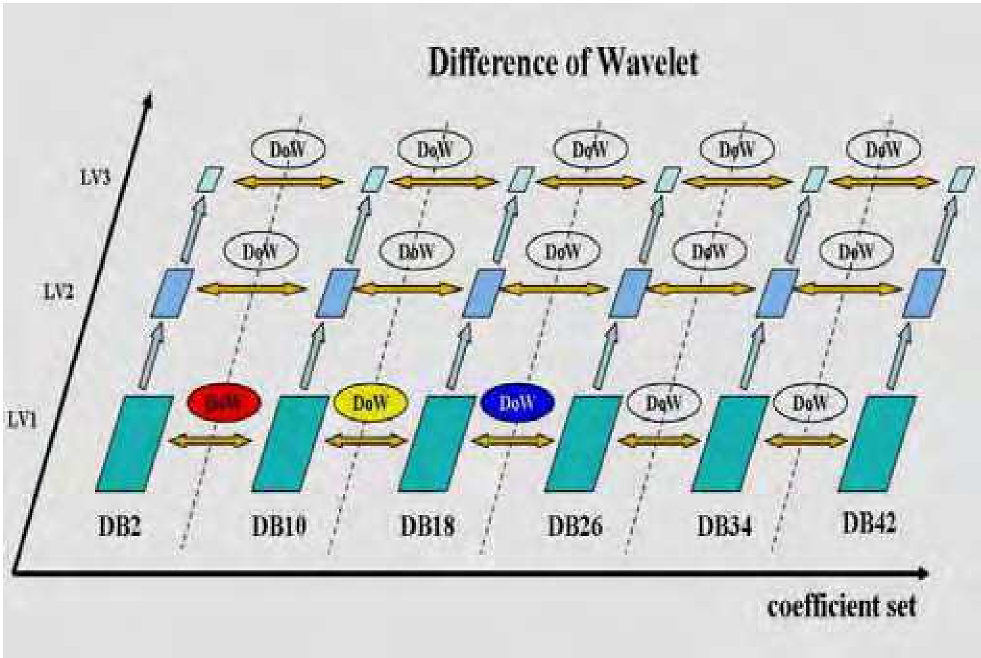


Figure 3. Structure of Difference of Wavelet

A. Parameter Decision for Wavelet Pyramid

In order to construct the wavelet pyramid, we decide the number of Daubechies coefficients and approximation levels, which can be considered as a counterpart of the DoG-based scale expansion. Fig. 4 shows that DB6 provides the optimum local key points, and Fig. 5 shows that approximation level 3 is the most efficient for matching. Although larger coefficients have better decomposition ability, we used DB2 as the first filter, and increased the step by 8. Because all DB filters have even numbered supports, difference between adjacent DB filters' support is recommended to be larger than or equal to 4 for easy alignment. In this work we used difference of 8, because difference of 4 provides almost same filtered images.

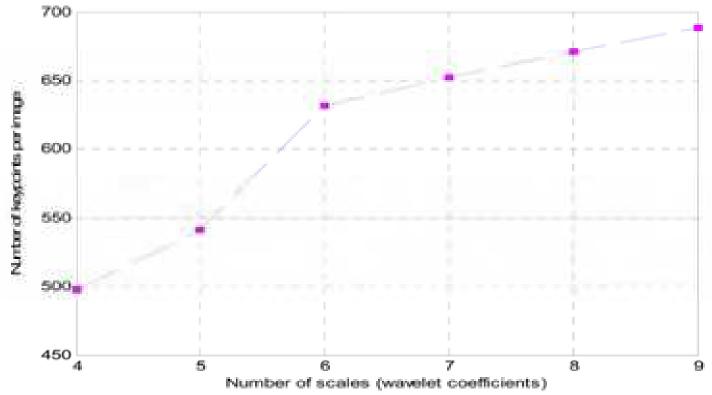


Figure 4. The number of extracted keypoints versus the number of wavelet coefficients

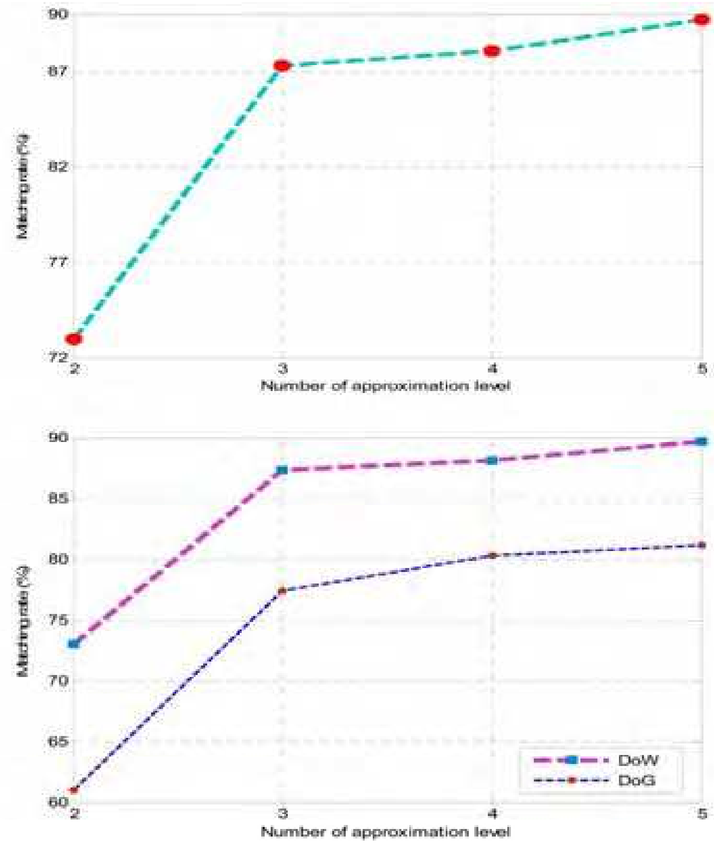


Figure 5. Matching rate versus the number of approximation level

Table 1 summarizes results experimental of processing time and matching rate using different wavelet filters in the SIFT framework. Coefficient set of the first row provides the best keypoint extraction result with significantly reduced computational overhead. The combination given in the second row is the best in the sense of matching time and rate.

Coefficient set \ Comparison factor	Processing time(msec)	Matching rate(%)
DB2, DB6, DB10, DB14, DB18, DB22	121	34.72
DB2, DB10, DB18, DB26, DB34, DB42	130	71.92
DB2, DB14, DB26, DB38, DB50, DB62	173	72.37
DB2, DB18, DB34, DB50, DB68, DB86	213	72.87
SIFT[4] (σ=1.6, k= $\sqrt{2}$, 1D Gaussian kernel size = 11) (Images per octave = 6, Number of octaves = 3)	925	57.68

Table 1. Various coefficient sets of Daubechies coefficients in the SIFT framework for measuring processing time and matching rate under low light(0.05lux) condition.

B. Wavelet-like Subband Transform

As shown in Fig. 3, the proposed wavelet pyramid is constructed using six Daubecies coefficient sets with three approximation levels. Because the length of each filter is even number, we need appropriate alignment method for matching different scales, as shown in Fig. 6, where DB10 is used for 320×240 input images.

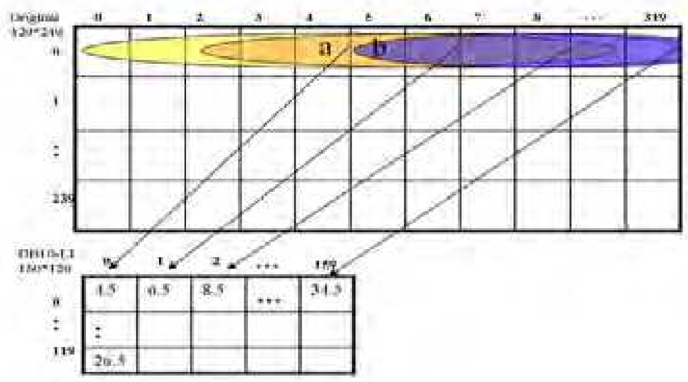


Figure 6. Proposed alignment method for different approximation levels

3.3 Local Extrema Detection and Local Image Descriptors

In the previous subsection we described the detail construction method for wavelet pyramid and DoW. In keypoints extraction step, we used min-max extrema with consideration of aligning asymmetrically filtered scales. In order to extract scale-invariant feature points, we compute DoW in the scale space, and locate the minimum and maximum pixels among the neighboring 8 pixels and 18 pixels in the upper and lower-scale images. Such extrema become scale-invariant features. DoWbased scale space is constructed as shown in Fig. 7.

For each octave of scale space, the initial images are repeatedly convolved with the corresponding wavelet filter to produce the set of scale space images shown in the left. DoW images are shown in the center, and in the right maxima and minima of the difference of wavelet images are detected by comparing a pixel, marked with ×, to its 26 neighbors in three 3 ×3 templates, marked with circle. For discrete wavelet transform, we used six different sets of Daubechies coefficients to generate a single octave, and make each difference image by using three octaves as

$$\begin{aligned} DoW_1 &= DB10\_L1 - DB2\_L1, \qquad DoW_2 = DB18\_L1 - DB10\_L1 \\ DoW_3 &= DB26\_L1 - DB18\_L1, \qquad DoW_4 = DB34\_L1 - DB26\_L1 \\ DoW_5 &= DB42\_L1 - DB34\_L1 \end{aligned} \tag{5}$$

Equation (5) defines how to make a DoW image using two wavelet transformed images. Feature points obtained by the proposed method are mainly located in the neighborhood of strong edges. DoW also has computational advantage to DoG because many octaves can be generated in parallel.

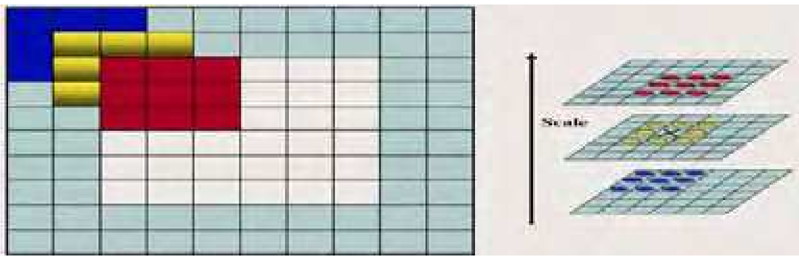


Figure 7. Maxima and minima of the difference of Wavelet images are detected by comparing a pixel (marked with X) to its 26 neighbors in 3×3regions at the current and adjacent scales (marked with circles)

4. Experimental Result

We first enhanced a low light - level image using the proposed NAST filter, as shown in Fig.8.

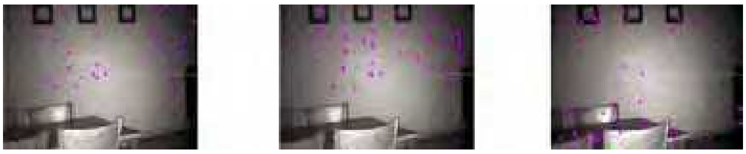
(a) (b)

Figure 8. (a) Input low light-level image with significant noise and (b) NAST filtered image

Comparison between DoG-based SIFT and the proposed DoW methods is shown in Fig. 9.As shown in Fig. 8, the proposed DoW method outperforms the DoG-based SIFT in the sense of both stability of extracted keypoints and computational efficiency.Fig. 10, Compares performance of combined NAST and DoG method with the DoG-based SIFT algorithm.

(a) (b) (c) (d)

Figure 9. Keypoint extraction results: (a) DoG, (b) DoW, and (c, d) translation of (a) and (b), respectively



(a) (b) (c)

Figure 10. Keypoints extraction results under low light-level condition using DoG, (b) DoG with NAST, and (c) DoW with NAST

Table 2 shows performance evaluation for processing time, matching rate and the PSNR in dB is obtained by using pre-filtering algorithm. The low pass filter(LPF)[13] were simulated for comparison with the NAST filter. In order to measure PSNR, we add synthetic noise (20dB PCN, and 15dB FCN) to the acquired low light images. This work was tested using a personal computer with Pentium- 3.0GHz.

Comparison Factor / Type of method	Processing time (msec)	PSNR(dB)	Matching rate (%)
DoG under low light	925	-	68.88
NAST + DoG under low light	1,104	39.48	70.98
LPF + DoW under low light	254	37.13	73.69
NAST + DoW under low light	355	39.50	77.24

Table. 2. Performance evaluation of DoG and DoW with NAST filter

5. Conclusion

The paper presents a local feature detection method for vSLAM-based self-localization of mobile robots. Extraction of strong feature points enables accurate self-localization under various conditions. We first proposed NAST pre-processing filter to enhance low light-level input images. The SIFT algorithm was modified by adopting wavelet transform instead of Gaussian pyramid construction. The wavelet-based pyramid outperformed the original SIFT in the sense of processing time and quality of extracted keypoints. A more efficient local feature detector and a compensation scheme of noise due to the low contrast images are also proposed. The proposed scene recognition method is robust against scale, rotation, and noise in the local feature space.

6. References

Dissanayake, M.W.M.G. et al. (2001) A solution to the simultaneous localization and map building(SLAM)problem, Robotics and Automation, IEEE Transactions, Vol. 17, Issue 3, pp. 229-241, June 2001

Lowe, D. G. (2004) Distinctive image features from scale invariant keypoints, Int'l Conf. on Computer Vision, vol. 60, No. 2, pp. 91-110, 2004

Lee, S.; Maik, V.; Jang, J.; Shin, S. & Paik, J. (2005) Noise-Adaptive Spatio-Temporal filter for real-time noise removal in low light images, IEEE Trans. Consumer Electronics, Vol. 51, No. 2, pp.648-653, May 2005

Daubechies, I. (1998) Orthogonal bases of compactly supported wavelets, Comun, Pure. Appl. Math, Vol. 41, pp. 909-996, Nov 1998 Irie, K. & Kishimoto, R. (1991) A study on perfect reconstructive subband coding, IEEE Trans. On CAS for Video Technology, Vol. 1, no. 1, pp. 42-48, March 1991