FUNCTIONALLY LAYERED CODING FOR RIVER MONITORING

Masahiro IWAHASHI, Sakol UDOMSIRI, Yuji IMAI, Suriyon TANSURIYAVONG

Department of Electrical Engineering
Nagaoka University of Technology
Nagaoka City, JAPAN
E-mail: {iwahashi@vos, sakol_ud@tech, yimai@tech, suriyon@vos}.nagaokaut.ac.jp

ABSTRACT

A new type of layered coding system for use of river monitoring is proposed. The bit stream produced by a sensor node of the system is functionally separated into three layers. The first layer contains band signals, decomposed by the Haar transform, effective for water level detection. The second layer contains band signals for thumb-nail video. These layers are transmitted at low bit rate for regular monitoring. The third layer which contains additional data necessary for decoding original video signal is transmitted when the necessity arises. Experimental results show which band signals should be included into the first layer for water level detection based on the maximum likelihood estimation. It becomes possible to reduce power consumption in a sensor node by distributing computational load via internet and by sharing the temporal and spatial basis function decomposition between "compression" and "recognition".

1. INTRODUCTION

So far, in Japan for example, water level of a river is observed with the "telemeter" installed by the government. The facility employs a water level detector with a probe and makes it possible to monitor water level at any time through website [1,2]. However its installation is small in number and limited to only some principal rivers. On the other hand, a large number of web cameras have been widely installed to various rivers for surveillance. They can transmit video data of a river at high communication bit rate. However it can not automatically detect the water level of a river which needs being monitored regularly and transmitted at low bit rate.

A water level detection algorithm from an image signal has been proposed by Takagi et al. [3,4]. It employs a measuring board with inclined lines painted on it. However, it is strictly controlled and difficult to be permitted to install an object such as the board in the water. Tsunashima et al. proposed a detection method without installing anything in a river by introducing edge detector and frame subtraction [5]. However, it is sensitive to rain and snow drops because of the frame subtraction. It is necessary to check the river scenery with video signal since there is no guarantee that the water level detection is perfect.

This paper proposes a new type of layered coding system specified for river monitoring use. A sensor node of the system produces the bit stream functionally separated into three layers. The first layer includes band signals efficient for water level detection. The band signals are generated by the Haar transform implemented with addition, subtraction and shift operations to reduce power consumption in a sensor node. The second layer contains band signals for thumb-nail video. These priority layers are transmitted at low bit rate for regular monitoring. The third layer contains additional data necessary for decoding original video signal. It is transmitted only when the necessity arises.

This report focuses on discussion on which band signals should be included into the first layer in sense of water level detection namely a "function", not in sense of picture quality as conventional layered coding systems [6-9]. Requirements to the system and the proposed technical solutions are summarized in section **2**. Efficient band signals for water level detection are investigated trough experiments in section **3**. Conclusions are presented in section **4**.

2. FUNCTIONALLY LAYERED CODING

2.1 System Requirement

Requirements to the river monitoring system in this report are summarized as follows.

- Water level is automatically detected and constantly transmitted to users at very low bit rate.
- Thumb-nail video signal of a river scenery is transmitted at low bit rate on demand.
- Video signal with originally obtained resolution is transmitted when the necessity arises.

Also we assume that the video is used for river monitoring in which (1) communication delay is allowed to some extent, (2) monitoring water level of the river has the first priority and (3) disappearance of water or snow drops in the video is preferable. It is our purpose to construct a system under the technical constrains below.

- Power consumption in a sensor node should be minimized since the sensor node is supposed to operate by solar and wind power.
- Data transmission amount should be minimized since the monitoring is supposed to be economic in regular situation and safe in a time of disaster.

2.2 Problem to be Solved

Figure 1 illustrates conventional river monitoring system. "Recognition" processing for water level detection and "compression" processing for video data transmission are simultaneously required. A layered coding for video compression based on MPEG algorithm includes motion compensation (MC) and prediction as temporal processing and discrete cosine transform (DCT) as spatial processing [7,8]. The recognition such as the water detection in [12] includes synchronous frame addition as temporal processing and wavelet transform as spatial processing. It is redundant to prepare two different temporal and spatial signal processing in a sensor node.

2.3 Technical Solutions

Figure 2 illustrates the proposed system. The synchronous frame addition and the MC prediction are synthesized into temporal Haar transform. The wavelet transform and DCT are also synthesized into spatial Haar transform. As a result, it becomes possible to construct low power consumption sensor node since the temporal and spatial basis function decomposition, the three dimensional Haar transform in this case, are shared between "compression" and "recognition".

Note that the two point one dimensional Haar transform is applied vertically and horizontally to produce four band signals LL, HL, LH and HH. The lowest LL band is further decomposed into four band signals in the same way as the wavelet transform in JPEG 2000 [13].

The band signals (transformed coefficients) are layered into a few categories. The system in figure 2 (b) categorizes LL band as the 2nd priority layer for thumb-nail video browsing. The 1st priority layer contains minimum number of bands, discussed in section **4**, necessary for water detection at remote place. All the layers are transmitted for full size video surveillance. There is almost no redundancy in data transmission.

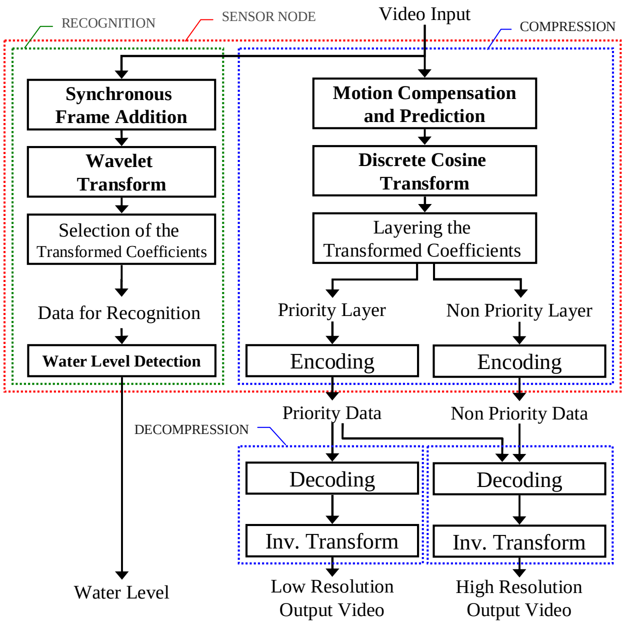
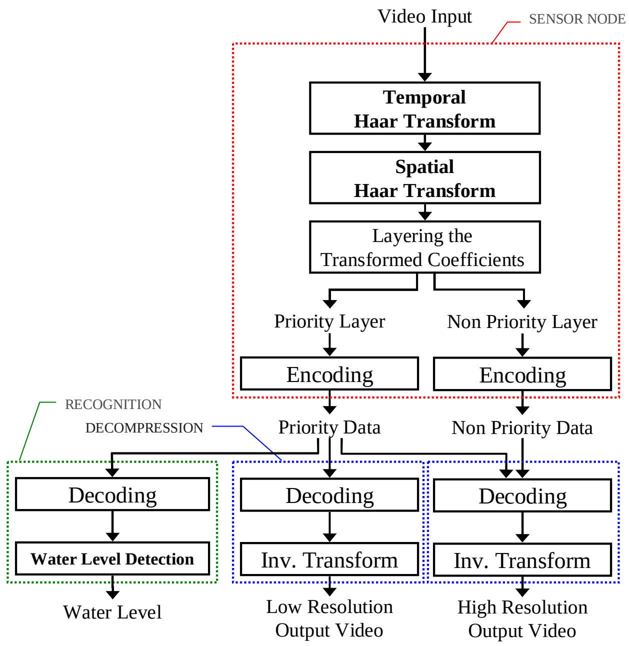


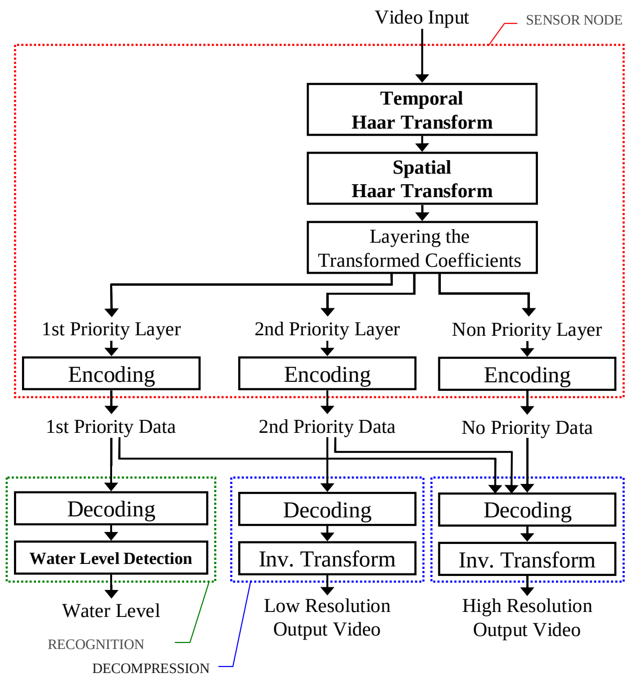
Figure 1 Conventional tandem system.

It is advantageous that the Haar transform is implemented with adders and shifters. Signal processing in the sensor node is simple since the wavelet transform and the MC prediction are excluded. Encoding in the sensor can be Huffman coding or else. It is expected be simpler than arithmetic coding in JPEG 2000 but effective for steady scenery such as river surveillance video signals.

It can be disadvantageous that delay of the video transmission is inevitable due to the three dimensional Haar transform. Coding efficiency will be slightly decreased by not employing MC prediction. Moving objects in the video may be blurred. Application of the proposed system is limited to the case under the assumption described in section **2.1**.



(a) Two layered type



(b) Three layered type

Figure 2 Proposed unified system.

2.4 Water Level Detection

Video signal is assumed to contain "land" region in upper part and running "water" region in lower part as illustrated in figure 5 (a). Teacher region of the "land" and the "water" are extracted from a previous frame and they are used to determine which class a pixel in a current frame belongs to. Figure 5 (f) indicates a discrimination result in which "water" pixel is indicated as black and "land" pixel is white. Based on this result, the water level is detected as a boundary of these two regions.

2.4.1 Temporal & Spatial Filtering

The discrimination method based on Gabor filter bank is proposed for "texture" classification [10,11]. It is illustrated as "Previous research 1" in figure 4. Its performance is precise, however it is not suitable for data compression since it outputs doubled number of pixels. We have replaced the Gabor filter by the wavelet transform in JPEG 2000 for data compression use [13]. It is indicated as "Previous research 2" in the figure. Introducing synchronous frame addition before the wavelet transform, it became possible to steadily detect water level of running water. However the wavelet transform requires huge computational load and power consumption.

The proposed system replaces the synchronous frame addition by the temporal Haar transform and the wavelet transform by the spatial Haar transform respectively as illustrated in figure 4. The lowest band signal in the temporal Haar transform is equivalent to the output signal of the synchronous frame addition as illustrated in figure 3. It is categorized into the priority layer and processed with the spatial Haar transform. This report investigates which band signals produced by the spatial Haar transform are efficient for water level detection.

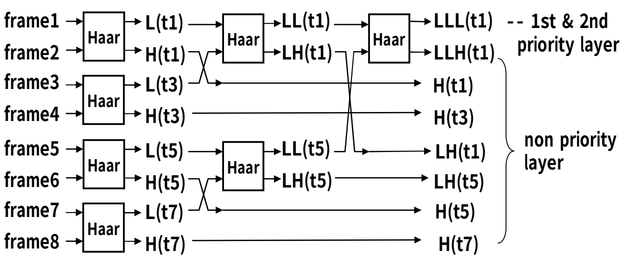


Figure 3 Temporal Haar transform.

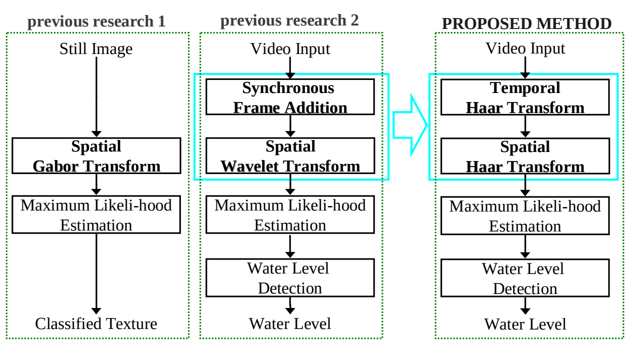


Figure 4 Unification of the temporal and spatial filtering.

2.4.2 Maximum Likelihood (ML) Estimation

Each pixel of the temporally transformed signal illustrated in figure 5 (b) is classified into one of the two classes, "land" (k=0) or "water" (k=1), using some of the spatially transformed signals (band signals) illustrated in figure 5 (d). These band signals are utilized as the feature vectors for "recognition" of the water region so that "compression" and "recognition" can share the same processing to reduce total hardware complexity and electric power consumption of the sensor node.

Assuming that the feature vector $\mathbf{G}(m,n)$ of a pixel at location (m,n) has a Gaussian probability density function

$$P(\mathbf{G}(m,n)|T_k) = \frac{1}{2\pi\sqrt{|\mathbf{C}_k|}}\cdot\exp\left(\frac{-d_k^2(m,n)}{2}\right) \qquad (1)$$

the class T_k , (k=0 or 1) of each pixel is determined according to the Mahalanobis distance between a pixel to be discriminated and centers of the teacher signals

$$d_k^2(m,n) = [\mathbf{G}(m,n)-\mu_k]^T\mathbf{C}_k^{-1}[\mathbf{G}(m,n)-\mu_k] \qquad (2)$$

where μ_k and $\mathbf{C}_k$ denote average vector of the teacher signal and covariance matrix respectively [11]. Dimension of the feature vector, namely the number of band signals to be used for the discrimination, and efficient band signals to be contained in the priority layer are experimentally investigated in the next section.

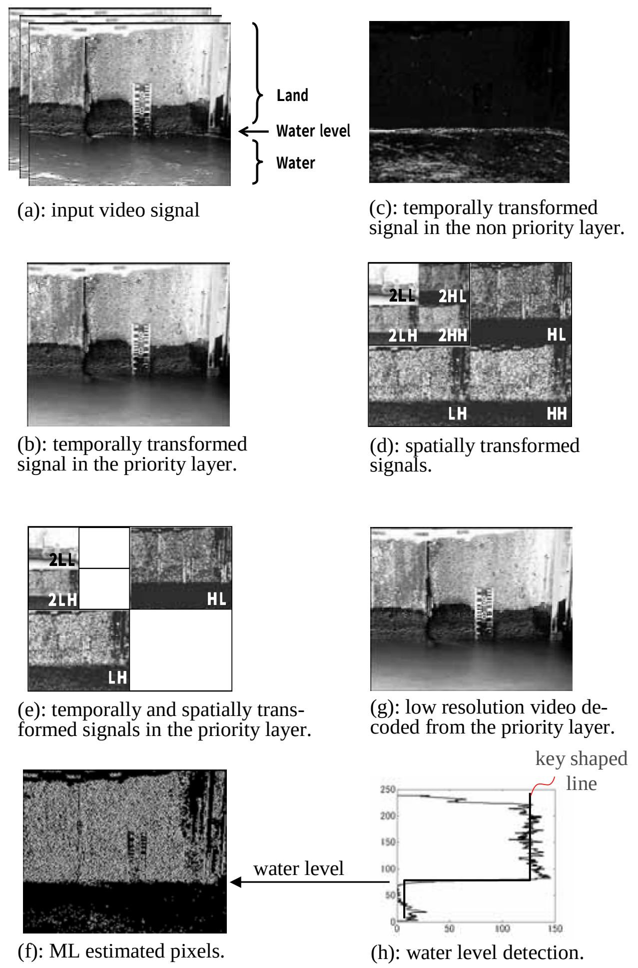


Figure 5 Signal processing example.

3. EXPERIMENTAL RESULTS

The conventional layered coding system selects the lowest LL band for the priority layer so that the decoded low resolution video has the best image quality. On the contrary, the proposed layered coding system selects band signals in respect of function. In this case, it is water level detection. This report focuses on investigating which band should be contained in the priority layer.

3.1 Video Signals to be Tested

Eleven kinds of video signals under various conditions were tested in the experiment. The first frame of some examples are indicated in figure 6. All of them has 320 x 240 pixels and 30 frames per second and taken with SONY "handycam" mounted on a tripod. Infrared light is used at night. It is assumed as same as the previous methods in [5,12] that the water is always running and its surface has waves. Otherwise some other methods with the measuring board reported in [3,4] is sometimes better than our method.

3.2 Wavelet vs. Haar

Since the wavelet transform is replaced by Haar transform in this report, these two transforms are compared in respect of discrimination rate and water level detection error. The temporally lowest band signal from the temporal Haar transform is spatially decomposed into seven bands as illustrated in figure 5 (d) with two stage octave decomposition by the spatial Haar transform. Since it is confirmed that synchronous addition of around thirty frames is efficient for water level detection [12], five-stage temporal octave decomposition is used to produce averaged image of 32 frames. All the combination tested in this report is listed in table 1. The stage of spatial octave decomposition is limited to two since band signals at more than three stages are useless because they decrease spatial resolution of water level detection.

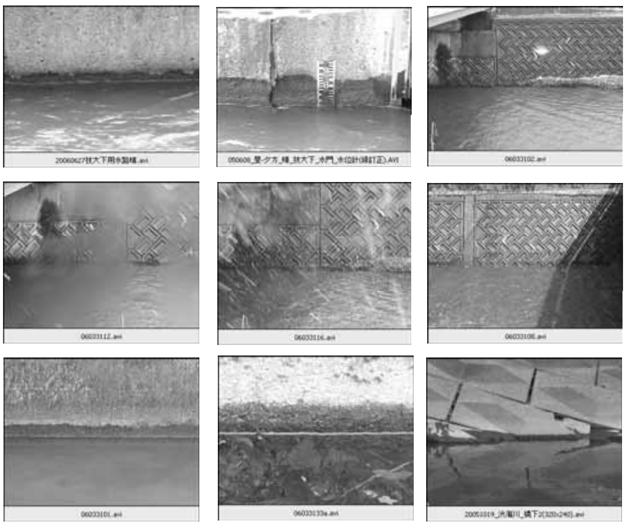


Figure 6 Video examples to be tested.

Figure 7 summarizes a comparison results measured with the discrimination rate averaged over all the tested video signals which is calculated from the number of correctly discriminated pixels in figure 5 (f). A hundred percent means perfect (no error). The number of bands to be used in the ML estimation is also indicated. It means dimension of the feature vector.

Figure 8 illustrates detection error averaged over all the tested video signals. It indicates distance measured in pixel between the detected water level and the ideal level which is specified by hand. It includes inevitable turbulence due to wave on the water surface.

As a result of these experiments, it became clear that more than three bands are necessary for precise water level detection. Since least number of bands for low bit rate transmission of the priority layer, it can be concluded that the number of bands should be three.

In discrimination rate, the wavelet transform is superior to the Haar transform. However there is no significant difference between them in detection error. This is because the discrimination rate around 80 [%] has no difference in detection error as illustrated in figure 9. It can be concluded that the wavelet transform can be replaced by the Haar transform.

Table 1 All the combination of the bands to be tested.

No.	Band
1	LL
2	LH
3	HL
4	HH
5	LL-LH
6	LL-HL
7	LL-HH
8	LH-HL
9	LH-HH
10	HL-HH
11	LL-LH-HL
12	LL-LH-HH
13	LL-HL-HH
14	LH-HL-HH
15	LL-LH-HL-HH
16	2LL
17	2LH
18	2HL
19	2HH
20	2LL-2LH
21	2LL-2HL
22	2LL-2HH
23	2LH-2HL
24	2LH-2HH
25	2HL-2HH
26	2LL-2LH-2HL
27	2LL-2LH-2HH
28	2LL-2HL-2HH
29	2LH-2HL-2HH
30	2LL-2LH-2HL-2HH
31	2LH-LH
32	2LH-HL
33	2LH-HH
34	2HL-LH
35	2HL-HL
36	2HL-HH
37	2HH-LH
38	2HH-HL
39	2HH-HH
40	2LH-2HL-LH
41	2LH-2HL-HL
42	2LH-2HL-HH
43	2LH-2HH-LH
44	2LH-2HH-HL
45	2LH-2HH-HH
46	2HL-2HH-LH
47	2HL-2HH-HL
48	2HL-2HH-HH
49	2LH-LH-HL
50	2LH-LH-HH
51	2LH-HL-HH
52	2HL-LH-HL
53	2HL-LH-HH
54	2HL-HL-HH
55	2HH-LH-HL
56	2HH-LH-HH
57	2HH-HL-HH
58	2LH-2HL-LH-HL
59	2LH-2HL-LH-HH
60	2LH-2HL-HL-HH
61	2LH-2HH-LH-HL
62	2LH-2HH-LH-HH
63	2LH-2HH-HL-HH
64	2HL-2HH-LH-HL
65	2LH-2HH-HL-HH
66	2HL-2HH-HL-HH
67	2LH-2HL-2HH-LH
68	2LH-2HL-2HH-HL
69	2LH-2HL-2HH-HH
70	2LH-LH-HL-HH
71	2HL-LH-HL-HH
72	2HH-LH-HL-HH
73	2LH-2HL-LH-HL-HH
74	2LH-2HH-LH-HL-HH
75	2HL-2HH-LH-HL-HH
76	2LH-2HL-2HH-LH-HL
77	2LH-2HL-2HH-LH-HH
78	2LH-2HL-2HH-HL-HH
79	2LH-2HL-2HH-LH-HL-HH

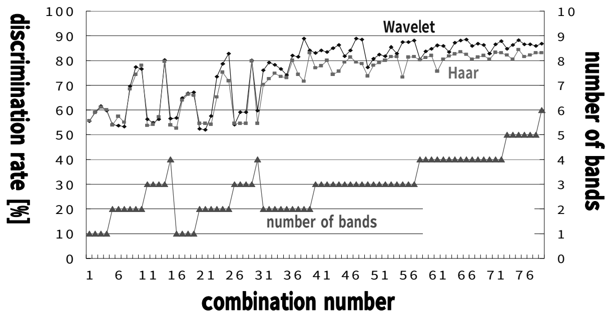


Figure 7 Discrimination rate of each combination.

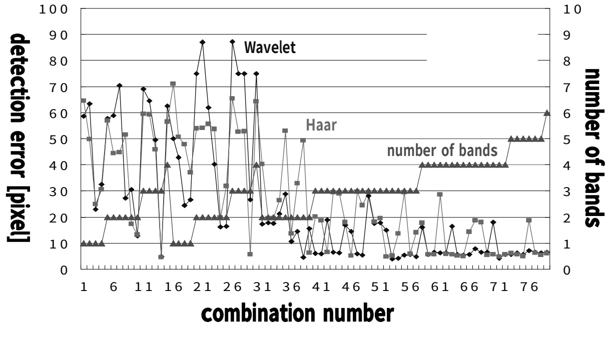


Figure 8 Water level detection error of each combination.

3.3 Band Signals in the Priority Layer

Table 2 summarizes top ten combinations in respect of the water level detection error. As an experimental result, minimum error in average over all the video signals tested in this experiment was found to be 4.6 [pixel]. It is necessary to add one more criterion to determine the bands to be included into the 1st priority layer.

Figure 10 indicates the top ten cases plotted in the detection rate and the data amount plane. The data amount is estimated by entropy rate of each band signal value multiplied by the number of pixels. It means data amount of the 1st priority layer to be transmitted at low bit rate. Now we have four choices: #14, #52, #46 and #29. Difference in detection precision is around one pixel at maximum. This is subtle when the waves on the water surface is considered. On the contrary, difference in data amount is 170k bit (=21 k byte). As described in section **2.1**, the priority layer should be transmitted at low bit rate. Therefore we conclude that combination #29 {2LH, 2HL, 2HH} is the 1st priority layer of the proposed system.

It is also possible to reduce data amount by introducing quantization. In this case, robustness of the water level detection against the quantization should be investigated. This is out of scope of this report but it should be done in the future.

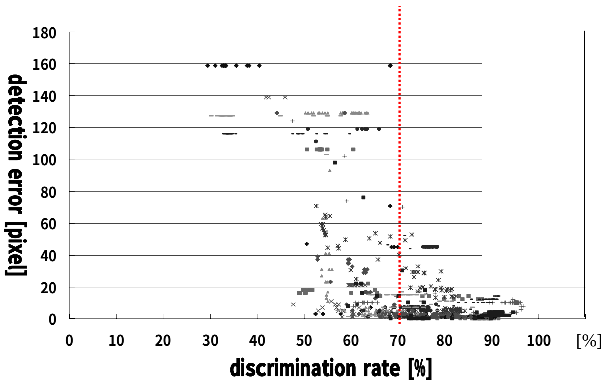


Figure 9 Discrimination rate versus detection error.

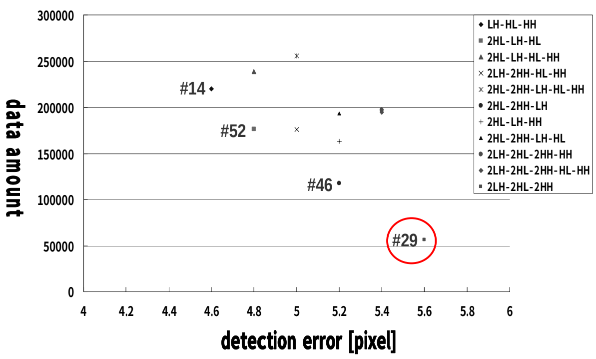


Figure 10 Data amount versus detection error.

3.4 Band Signals in other Layers

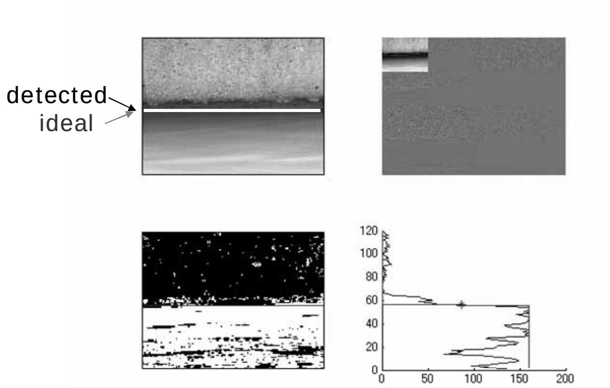
The 2nd layer of the system contains necessary data to reconstruct 2LL band signal. In this case, spatial resolution of the thumb-nail video is "quarter (1/4)" in horizontal and vertical direction. If one wish to have "half (1/2)" as indicated in figure 5 (g), the 1st and the 2nd priority layer is decoded to reconstruct LL band as the thumb-nail video. In case of 1/8 size thumb-nail, the 2nd layer contains 3LL band. Temporal resolution is 1/32 in frame rate for all.

The non priority layer contains all the components necessary to reconstruct the rest of the components in the priority layers. Investigation on data amount of the system is out of scope of this report but it should be done next.

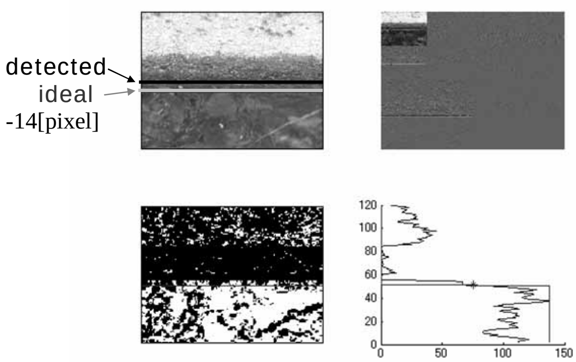
Two representative signal processing examples are illustrated in figure 11 for one's information.

Table 2 Top ten combinations of the band signals.

rank	No.	Band	detection error	number of bands
1	14	LH-**HL**-HH	**4.6**	3
2	52	2HL-LH-**HL**	**4.8**	3
2	71	2HL-LH-**HL**-HH	4.8	4
4	65	2LH-2HH-**HL**-HH	5	4
4	75	2HL-2HH-LH-**HL**-HH	5	5
6	46	2**HL**-2HH-LH	5.2	3
6	53	2**HL**-LH-HH	5.2	3
6	64	2HL-2HH-LH-**HL**	5.2	4
9	69	2LH-2**HL**-2HH-HH	5.4	4
9	78	2LH-2**HL**-2HH-HL-HH	5.4	5



(a) case of success



(b) case of failure

Figure 11 Detection examples.

4. CONCLUSIONS

A functionally layered coding system for river monitoring use is proposed. The bit stream of the system is functionally layered into three categories. The first priority layer contains efficient data for water level detection. It is experimentally determined to be HL, LH and HH band signals of the two stage octave decomposition with the spatial Haar transform applied to transformed coefficients of the five stage temporal Haar transform of the video signal obtained by a camera. The second priority layer contains spatially and temporally lowest band signal for thumb-nail video. These priority layers are transmitted at low bit rate for regular monitoring. The third layer, non priority layer, which contains additional data necessary for decoding original video signal is transmitted when the necessity arises. The band signals to be contained into the 1st priority layer are experimentally determined in respect of water level detection based on the maximum likelihood estimation and data amount to be transmitted.

It becomes possible to reduce power consumption in a sensor node distributing computational load via internet and sharing the temporal and spatial basis function decomposition, three dimensional Haar transform, between "compression" of the video signal and "recognition" of the water level.

Investigation on the data amount or transmission rate associated with quantization is omitted in this report. However it should be done and to be presented in the near future.

ACKNOWLEDGMENT

This research is financially supported by Ministry of Internal Affairs and Communications, Japan as a SCOPE-C research project. Video sensors are installed in cooperation with Ministry of Land Infrastructure and Transport, Japan, Niigata prefecture and Nagaoka city.

5. REFERENCES

[1] Ministry of international affairs and communications, Japan, http://www.soumu.go.jp/english/index.html

[2] http://www.river.go.jp/

[3] Y.Takagi, A.Tsujikawa, M.Takato, T.Saito, M.Kaida, "Development of a non-contact liquid level measuring system using image processing", Water science and technology, vol. 37, no.12, pp.381-387, 1998.

[4] Y.Takagi, T.Yoneoka, H.Mori, M.Yoda, A.Tsujikawa, T.Saito, "Development of a water level measuring system using image processing", The 1st IWA conference on instrumentation, control and automation, pp.309-316, 2001.

[5] N.Tsunashima, M.Shiohara, S.Sasaki, J.Tanahashi, "Water level measurement using image processing", Information processing society of Japan, Research report, Computer vision and image media, vol.121, no.15, pp.111-117, March 2000.

[6] Iwahashi M., Kameyama W., Ohyama K., Kambayashi N.: A motion compensation technique for down-scaled pictures in layered coding; IEICE Trans. on commun., vol.E-77, no.8, pp.1007-1012, Aug. 1994.

[7] Jae-Yong Lee; Hyo-Sub Oh; Sung-Jea Ko, "Motion-compensated layered video coding for playback scalability", IEEE Trans. on circuits and systems for video technology, vol.11, issue 5, pp.619 - 628, May 2001.

[8] Weiping Li, "Overview of fine granularity scalability in MPEG-4 video standard", IEEE Trans. on circuits and systems for video technology, vol. 11, issue 3, pp. 301 - 317, March 2001.

[9] Kondi, L.P.; Srinivasan, D.; Pados, D.A.; Batalama, S.N.; "Layered video transmission over wireless multirate DS-CDMA links", IEEE Trans. on circuits and systems for video technology, vol.15, issue 12, pp.1629 - 1637, Dec. 2005.

[10] D.Dunn, W.E.Higgins, "Optimal Gabor filter for texture segmentation", IEEE Trans. image processing, vol.4, no.7, pp.947-964, 1995.

[11] H.Tanaka, Y.Yoshida, K.Fukami, H.Nakano, "Texture segmentation using amplitude and phase information of Gabor filters", IEICE Trans. D-II, vol.J84 -DII, no.12, pp.2565-2576, Dec. 2001.

[12] Y.Imai, M.Iwahashi, "Water level detection based on wavelet transform of JPEG2000", IEICE circuits and systems "Karuizawa" workshop, pp.531-534 April, 2006.

[13] D.S.Taubman, M.W.Marcellin, "JPEG 2000 - Image compression fundamentals, standards and practice", Kluwer Academic Publishers, 2002.