WATER LEVEL DETECTION FOR FUNCTIONALLY LAYERED VIDEO CODING

*Masahiro IWAHASHI Sakol UDOMSIRI Yuji IMAI Shogo MURAMATSU**

Nagaoka University of Technology
Nagaoka, Niigata, 940-2188, Japan

*Niigata University
Niigata, Niigata, 950-2181, Japan

ABSTRACT

This paper proposes a new type of layered video coding especially for the use of monitoring a river or a water channel. A sensor node of the system decomposes a video signal into some components and produces a bit stream which is functionally separated into three layers. The first layer contains the minimum components effective for detecting water level. The second layer contains signals for thumb-nail video browsing. Each of them is transmitted at very low bit rate for regular monitoring. The third layer contains additional data for decoding the original video signal. It is transmitted in case of necessity. The original video signal is decomposed into band signals as the components by the Haar transform in a sensor node. Experimental results show which band signals should be included into the first layer considering both of water level detection performance and data size to be transmitted.

Index Terms— water, recognition, layered, video, coding

1. INTRODUCTION

So far, in Japan for example, water level of a river has been regularly monitored by a “telemeter” and result is opened to the public on website [1]. However, it is limited to only a few principal rivers and the number of the points is not enough to cover all the dangerous places. A large number of web cameras have been installed to various rivers for video surveillance. They can transmit video signal data at high bit rate. However, it can not automatically detect the water level which is required to be monitored regularly at low bit rate.

Image recognition algorithm for automatic detection of the water level has been proposed by Takagi et al. [2]. It requires 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 the water by introducing edge detection and frame subtraction [3]. However, it is sensitive to rain drops due to the frame subtraction. We have proposed a robust recognition algorithm based on the synchronous frame addition [6]. It was extended from the texture recognition algorithms in [4,5] replacing the Gabor transform by the wavelet transform.

It is necessary to check not only the water level but also scenery of the river with video signals for a proper countermeasure to prevent floods. In this case, both of a video “compression” and water level “recognition” should be installed in a sensor node in tandem. In this case, huge power consumption for the video processing is required.

This paper proposes a functionally layered video coding to construct a river monitoring system with a small DSP board inside a web camera. An input video signal is decomposed into several band signals by the temporal and spatial Haar transform. These components are classified into three layers and separately transmitted from the sensor node. The first, the second and the third layer contain band signals effective for water level detection, browsing a thumb-nail video and a full resolution video, respectively.

The second and the third layer are equivalent to the SNR based conventional layered coding. Band signals to be contained in the first layer, newly introduced in this paper, are determined considering both of data “compression” and water level “recognition” performance. These bands are functionally effective for river monitoring and transmitted at low bit rate regularly.

2. FUNCTIONALLY LAYERED VIDEO CODING

2.1. Overview of the system

Figure 1 illustrates a conventional river monitoring system. “Compression” for video data transmission and “recognition” for water level detection are installed in tandem. The former is based on the motion compensation (MC) and the discrete cosine transform (DCT) widely known as MPEG. The latter includes synchronous frame addition (SFA) and the wavelet transform (WT) [6].

Figure 2 illustrates the proposed system. SFA and MC are synthesized into the temporal Haar transform. WT and DCT are also synthesized into the spatial Haar transform. It becomes possible to construct a low power consumption sensor node since the temporal and spatial video processing: the three dimensional (3D) Haar transform (HT), are shared between “compression” and “recognition”.

1-4244-1437-7/07/$20.00 ©2007 IEEE

It is advantageous to reducing power consumption that HT is implemented with adders and shifters. Signal processing in the sensor node becomes simpler than combination of WT and MC-DCT. Encoding in the sensor node can be Huffman coding or else. It is expected to be simpler than arithmetic coding in JPEG 2000.

It can be disadvantageous that coding efficiency will be slightly decreased compared to MC-DCT. Delay of the video transmission is inevitable due to 3D HT. Moving objects in the thumb-nail video are blurred. Therefore, application of the proposed system is limited to a case where these disadvantages are tolerable.

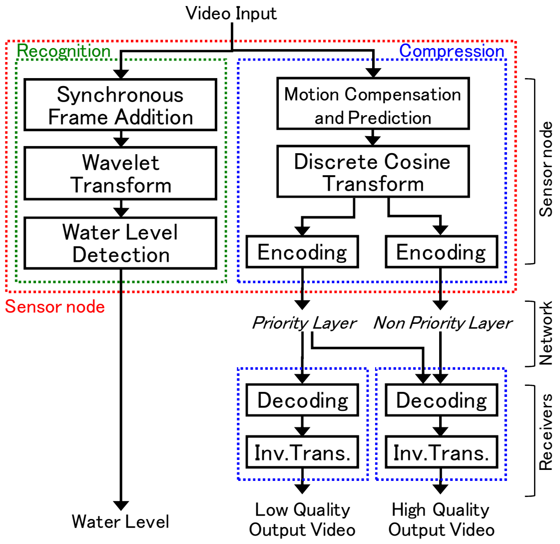


Figure 1 Conventional tandem system.

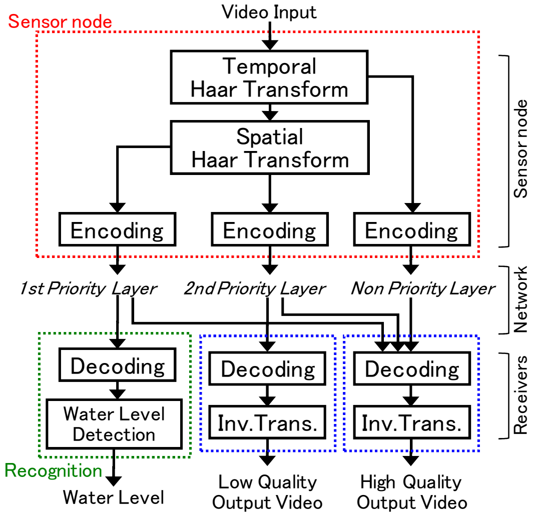


Figure 2 Proposed unified system.

2.2. Water level detection algorithm

Video signal is taken so that it contains "land" region in upper part and running "water" region in lower part as illustrated in figure 4(a). Boundary of these regions is recognized as the water level.

Consecutive frames are averaged so that the running water region contains less energy in high frequency. It is shown by figure 4(b), 5(b), 5(d). This procedure is performed by temporal two-point 1D HT illustrated in figure 3. The lowest band, which is equivalent to SFA processed signal, is categorized into the priority layers.

Furthermore, it is spatially transformed by two-point 1D HT vertically and horizontally to produce four band signals {1LL, 1HL, 1LH, 1HH} at the 1st stage. The lowest band 1LL is decomposed into {2LL, 2HL, 2LH, 2HH} at the 2nd stage. It is repeated to the *S*-th stage by means of the octave decomposition [5,6].

The lowest band signal 2LL (or 3LL) is categorized into the second layer as same as the conventional scalable coding. Some other few band signals are contained in the first priority layer as the minimum band signals effective to detect the water level. It is determined in section **3**.

Figure 4(c) indicates a discrimination result in which "water" pixel is indicated as black and "land" pixel is white. The water level is detected as a boundary of these two regions as in figure 4(d) where the number of "land" pixels is used as a feature value of each line.

2.3. Maximum Likelihood (ML) Estimation

Each of the pixel is classified into "land" (k=0) or "water" (k=1) using one or some of the band signals. These band signals are magnified to the original spatial resolution and utilized as the feature vector components of a pixel located at (n_1, n_2) for discriminating the region.

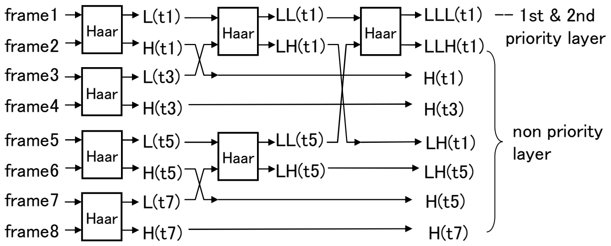
Assuming that the feature vector $G(n_1,n_2)$ has a Gaussian probability density function:

$$P(G(n_1,n_2)|T_k) = \frac{1}{2\pi\sqrt{|C_k|}}\exp\left(\frac{-d_k^2(n_1,n_2)}{2}\right), \qquad (1)$$

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

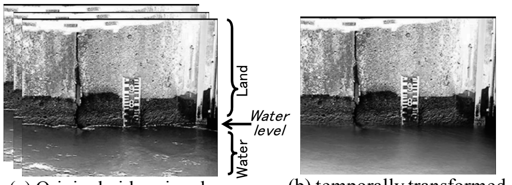
$$d_k^2(n_1,n_2) = \left[G(n_1,n_2) - m_k\right]^T C_k^{-1}\left[G(n_1,n_2) - m_k\right] \qquad (2)$$

where m_k and C_k denote average vector and covariance matrix of the teacher signals respectively [4-6]. 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 first priority layer are experimentally investigated in the next section.

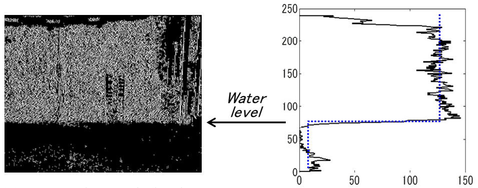


Haar : temporal Haar transform

Figure 3 Octave decomposition with the temporal HT.

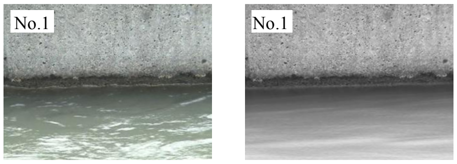


(a) Original video signal. (b) temporally transformed signal in the priority layer.

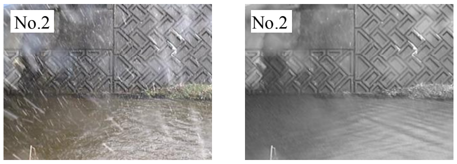


(c) ML estimated pixels. (d) Feature value of each line.

Figure 4 Signal processing of the proposed system.

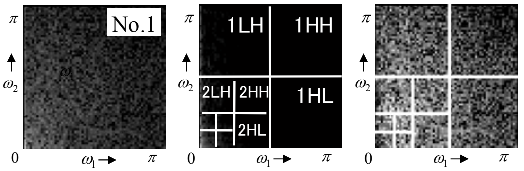


(a) original signal (b) after frame addition

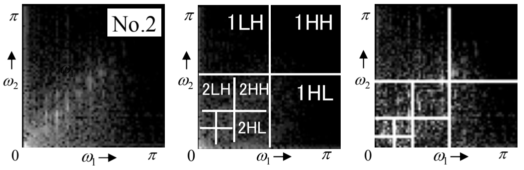


(c) original signal (d) after frame addition

Figure 5 Tested video samples [320x240 pxl, 30 fps].



(a) Land region (b) Water region (c) Difference



(d) Land region (e) Water region (f) Difference

Figure 6 Spectrum of the video samples.

3. EXPERIMENTAL RESULTS

As mentioned earlier, the lowest band of the temporal HT contains less energy in high frequency in the running water region. It is also confirmed by its spectrum in figure 6. Band signals which emphasize the spectrum difference between the two regions are considered to be the best feature vector for the discrimination.

Figure 7 and 8 illustrates experimental results for the video sample no.1 and no.2 in figure 5 respectively. Figure 7(a) indicates percentage of incorrectly discriminated pixels: discrimination error rate "*R*" for using one of the four band signals at 1st, 2nd and 3rd stage of the octave decomposition. In this case, 2HH with *R*=4.6 [%] is the best as the 1D feature vector. 2HL with *R*=4.9 is the 2nd. These are in middle frequency as illustrated in figure 6(b) where the difference between the two regions exists. 1LL and 2LL are not effective for recognition.

Figure 7(b) indicates water level detection error versus discrimination error. Their mutual correlation is 0.853 for sample no.1. On the contrary, correlation is weak for sample no2. It can be explained that the discrimination error indicates "confidence" of recognition, whereas the detection error can be "contingently" close to zero.

In this paper, not only "recognition" but also "compression" is considered. Therefore, one more axis which indicates data amount of the band signal: data size "*B*" is added in figure 7(c). It is estimated by the 1st order entropy rate. 3HH with *B*=3.2 [KB] is the best in respect of data size to be transmitted and 3HL with *B*=3.6 is the 2nd.

A band signal to be included in the 1st priority layer is determined considering both of *R* and *B*. In this paper, it is evaluated by the distance defined by

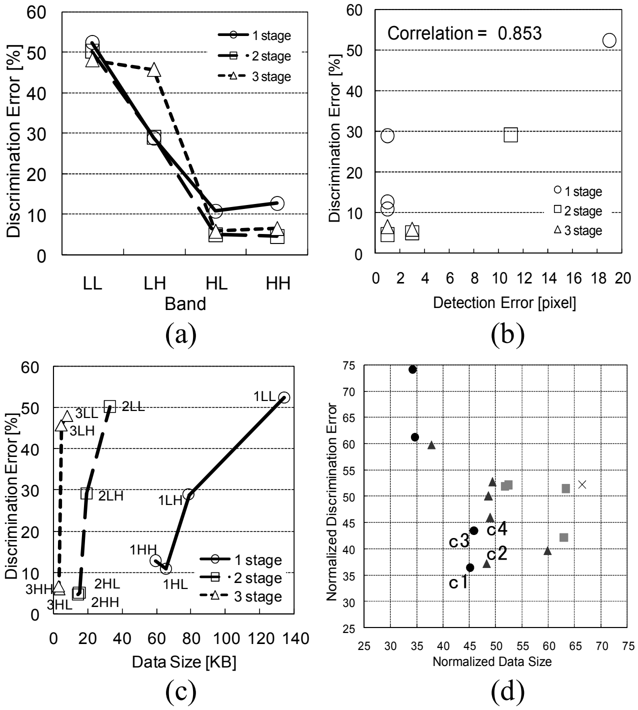
$$D = \sqrt{\frac{NR^2 + NB^2}{2}} \qquad \text{for}$$

$$NX = 50 + 10 \cdot \frac{X - m_X}{\sigma_X}, \quad X \in \{NR, NB\}. \qquad (3)$$

In this case, 2HH is the best and 2HL is the 2nd.

When combination of some band signals are used as the multi dimensional feature vector, result is summarized in figure 7(d) and table 1 (left side). It is found that 2HH with *D*=41.0 is the best and the combination of 3HH and 2HH with *D*=43.1 is the 2nd.

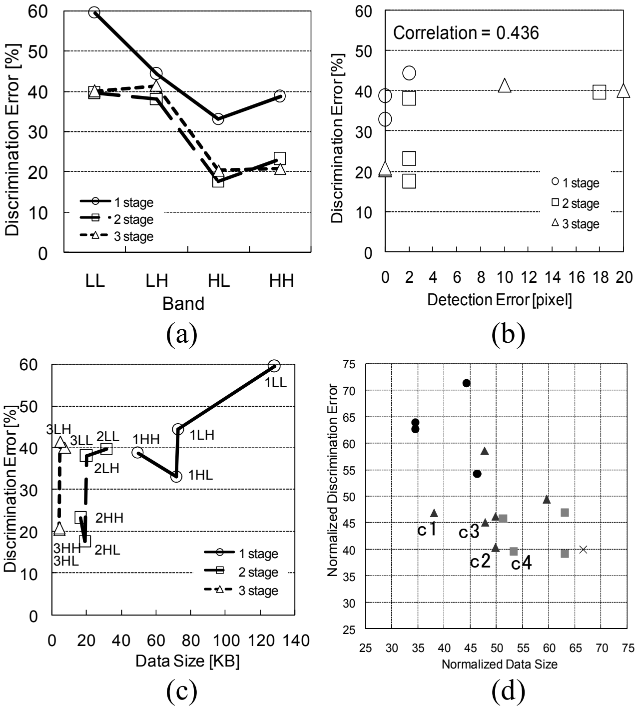
In case of sample no.2, 3HL-3HH with *D*=42.7 is the best and 2HL-3HL with *D*=45.3 is the 2nd. The best combination of the band signals in respect of the distance *D* is encoded as the 1st priority layer. Result varies from sample to sample. This paper indicates how to determine the band signals in the 1st priority layer.



(a) (b)

(c) (d)

Figure 7 Experimental results for sample no.1.



(a) (b)

(c) (d)

Figure 8 Experimental results for sample no.2.

Table 1 Combination of bands for the 1st priority layer.

sample no.1

	bands	R	B	D
c1	2HH	4.6	14.6	41.0
c2	3HH−2HH	4.6	17.8	43.1
c3	2HL	4.9	15.4	44.7
c4	2HL−3HH	5.1	18.6	47.5
c5	3HL−2HH	5.3	18.3	49.3
c6	3HL	5.9	3.6	49.7
c7	3HL−3HH	5.8	6.9	50.0
c8	2HL−2HH	4.7	30.0	50.8
c9	3HL−2HL	5.4	19.0	51.1
c10	3HL−3HH−2HH	5.4	21.5	51.8
c11	3HL−2HL−3HH	5.4	22.2	52.3
c12	2HL−3HH−2HH	4.9	33.2	53.6
c13	3HL−2HL−2HH	5.4	33.6	57.7
c14	3HH	6.5	3.2	57.7
c15	3HL−2HL−3HH−2HH	5.4	36.8	59.8

sample no.2

	bands	R	B	D
c1	3HL−3HH	15.1	8.8	42.7
c2	2HL−3HL	12.9	23.8	45.3
c3	3HL−2HH	14.5	21.2	46.5
c4	2HL−3HL−3HH	12.7	28.1	47.0
c5	2HL−3HH	14.9	23.7	48.0
c6	3HL−3HH−2HH	14.7	25.6	48.6
c7	2HL	17.5	19.3	50.4
c8	3HL	20.3	4.4	50.6
c9	3HH	20.7	4.4	51.4
c10	2HL−3HL−2HH	12.5	40.5	52.5
c11	3HH−2HH	19.0	21.1	53.5
c12	2HL−2HH	15.9	36.1	54.8
c13	2HL−3HL−3HH−2HH	12.8	44.9	54.9
c14	2HL−3HH−2HH	15.1	40.5	55.6
c15	2HH	23.2	16.7	59.4

4. CONCLUSIONS

A functionally layered video coding system for river monitoring is proposed. The bit stream of the system is functionally layered into three categories. Efficient band signals for water level detection are included into the first priority layer. These are experimentally determined according to both of data amount to be transmitted and recognition efficiency to detecting water level. As a result, the band signals in middle frequency, where difference between land region and water region exists, are selected.

It is expected to be able to reduce power consumption in a sensor node by distributing computational load via internet and by sharing the temporal and the spatial transform between "compression" and "recognition".

Investigation on introducing quantization will be done in the near future.

5. REFERENCES

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

[2] 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.

[3] 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, 2000.3.

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

[5] 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, 2001. 12.

[6] Y.Imai, M.Iwahashi, "Water level detection based on wavelet transform of JPEG2000", IEICE circuits and systems "Karuizawa" workshop, pp.531-534, 2006.4.