A Simple Human Visual Weighted Hadamard Transform Image Coding

(단순한 시각적 하중에 의한 아다마르 영상부호화)

黃 在 政* 李 門 浩**

(Jae Jeong Hwang and Moon Ho Lee)

要 約

선택적인 주파수 특성을 가지고 있는 인간의 시각 시스템(HVS)을 왈쉬함수로 표현되는 아다마르 변환 영역에서 적용하였다. 아다마르 기본함수의 주파수 성분을 구하여 시각 시스템을 모델링하였으며 이때 오차측정의 기준으로서 저 전송율 변환 부호화기에서 중요한 요소인 블록 경계상의 오차와 평균 자승 오차를 사용한 결과 인간이 가장 높은 감도를 갖는 공간 주파수는 이산 여현 변환의 경우보다 높게 나타났다.

부호화기에서 HVS 모델을 적용하여 시각적 감도가 높거나 중요한 정보만을 전송하고 복호화기에서는 역하중이 없는 방법을 제시하였다. 블록과 블록이 교차하는 영역에서 특히 크게 나타나는 오차를 감소시키기 위해 시각적 하중 방법을 사용하였으며 0.8bpp이하의 전송율에서 인간의 시각 특성에 만족되는 변환부호화를 시험하였다.

Abstract

Various models incorporating Human Visual System (HVS) with the Hadamard transform (HT) represented by Walsh functions are considered. Using the exact frequency components of HT basis functions, the optimum modulation transfer function (MTF) which has a higher peak frequency than DCT schemes is obtained analytically and visually. The main criterion, for error measurement, is errors at the block boundaries which is an important factor in transform coding. The scheme which has no inverse HVS is proposed. It causes some degradation of image data but it is insignificant. Crossing area of 4 blocks is equalized by the HVS weighting coefficients. The HVS weighted coding results in perceptually higher quality images compared with the unweighted scheme.

*正會員, 群山水産專門大學 通信學科

(Dept. of Telecommun. Kunsan Nat'l Fish. Coll.)

**正會員, 全北大學校 情報通信工學科

(Dept. of Inform. & Telecommun. Chonbuk Nat'l Univ.)

接受日字：1988年 11月 22日

I. Introduction

In recent years, the Human Visual System (HVS) has been employed in reducing the amount of image data. There are many features of visual network according to variations in spatial detail;

Mach band, lateral inhibition, frequency response etc. Frequency response is that the eye is more sensitive to certain spatial frequencies than to others and the eye's spatial response falls to negligible values at a high enough spatial frequency and is substantial over some intermediate range. [1] This is particularly interesting from the point of view of transform coding, since it operates in a similar way, converting the spatial detail in an image into frequency (or sequency) components, In transform coding, we select some coefficients having higher energy or variance using zonal or variance bit allocation methods, but such non adaptive methods are not sufficient to fit with more complicated texture. The optimum quantization is performed so that reconstructed information may be expressed most visually. If we consider the HVS, representing sensitive or insensitive properties of the eye, it is evident that those spectral components to which the eye is more sensitive should be coded more accurately than the others. Subject to the weighting coefficients, coefficients below the threshold are discarded as irrelevant data in a block. Only the relevant and indispensable coefficients therefore are sent to the receiver [2], [11].

The Handamard transform (HT) is attractive in implementation and computational aspects, as it does not involve multiplications, but it is not as efficient as other transforms in data compacting ability [3]. The shapes of the HT basis vectors are rectangular rather than cosinusoidal. But in order to understand how they perform in such applications, we study the frequency spectrum of the Hadamard basis function. In the concept of sequency, Walsh functions can be represented by Cal (even) and Sal (odd) with respect to the midpoint of the unit interval [4]. They are named to emphasize the analogy to sinusoidal functions. The HT takes advantage of them as a transform vector. But the Fourier transform will show the exact features. Using this, we can obtain the most appropriate weighting values with the HT and thus the coding efficiency is increased.

II. Modelling the HVS in Transform Coding

The HVS has been incorporated in transform coding of images by several researchers. Mannos and Sakrison's work [5] may be the first major breakthrough in image coding incorporating the HVS. They carried out extensive experiments to determine the parameters for the HVS model. Using the assumption that the HVS is isotropic, they simulated the optimum encoding for the Gaussian source and modelled the HVS as a nonlinear point transformation followed by the modulation transfer function (MTF) of the form

$$H(f) = a(b+cf)\exp(-cf)^{d} \qquad (1)$$

where f is the radial frequency in cycles/degree of visual angle and a, b, c, and d are constants. By varying these constants, the shape and the peak frequency of the MTF are altered. The 512 x 512 full frame images were viewed at a distance at which 65 pels subtended 1° of vision and then transformed to frequency domain via the Discrete Fourier Transform (DFT). Fourier coefficients are weighted by the MTF at the corresponding frequencies and the coefficients which are greater than the predetermined threshold are selected and transmitted. The reverse process is carried out at the receiver. The simulation results are judged subjectively. The final MTF is of the form

$$H(f) = 2.6(0.0192+0.114f)\exp(-(0.114f)^{1.1}) \qquad (2)$$

This MTF has a peak at f=7.9 cycles/degree.

Recently new MTFs have been proposed for using with the DCT. Nill[6] proposed a multiplicative function A(f) which is multiplied by the following MTF

$$H(f) = (0.2+0.45f)\exp(-0.18f) \qquad (3)$$

This function has a peak value at spatial frequency around 5.1 cycles/degree. In order to incorporate the HVS model into cosine transform coding, an even extension of the original scene has to be created but this causes the loss of physical significance since the human observer is not viewing this altered scene. To overcome this problem, Nill proposed the introduction of a function |A(f)| which takes the form of

$$|A(f)| = \left[\frac{1}{4} + \frac{1}{\pi^2}\right]\left[\log_e\left(\frac{2\pi f}{\alpha} + \sqrt{\frac{4\pi^2 f^2}{\alpha^2}+1}\right)^2\right]^{1/2} \qquad (4)$$

where $\alpha = 11.636$ degree^{-1}.

The modified HVS function thus becomes

$$H'(f) = |A(f)|H(f) \quad (5)$$

which has a peak frequency at around 6.9 cyc./deg.

Ngan et al.[7] used Nill's multiplicative function with their MTF. It has been found in this work that the peak frequency of the transfer function of 3 cycles/degree gives the best results. The transfer function takes the form of

$$H(f) = (0.31+0.69f)\exp(-0.29f) \quad (6)$$

After multiplying H(f) by |A(f)|, the resulting function has the peak frequency around 4.1 cycles/degree. Using zig-zag scanning sequence for the DCT coefficients, they achieved and acceptable reconstructed image at the low bit rates.

The MTFs of Mannos and Sakrison [5], Nill[6], and Ngan[7] are shown in Fig.1. Note that the latter two functions are obtained after multiplying (3) and (6) by (4). In general, incoporating the HVS models improves the coding scheme. So does the proposed Hadamard transform scheme presented in the next section.

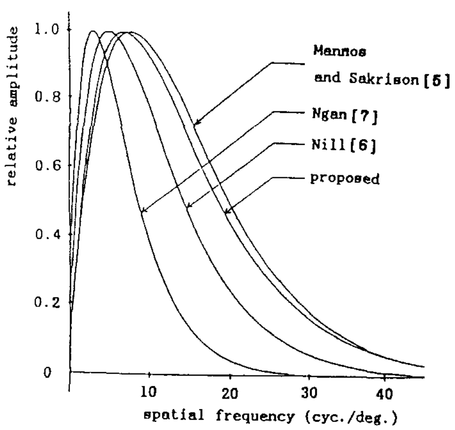


Fig.1. Comparison of various MTFs.

III. Spectrum of the Hadamard Function

Let the Fourier transform of a Walsh function, wal(m,x), be defined by

$$W_m(f) = \int_0^1 \mathrm{wal}(m,x)\exp(-i2\pi fx)\,dx \quad (7)$$

This can also be written as

$$W_m(f) = P_m(f)H_m(f) \quad (8)$$

where $P_m(f)$ is the Fourier transform of the elementary pulse for generating the Hadamard function of index m, and $H_m(f)$ is the transfer function of the system which generates the m-th function by a suitable arrangement of its elementary pulse function.

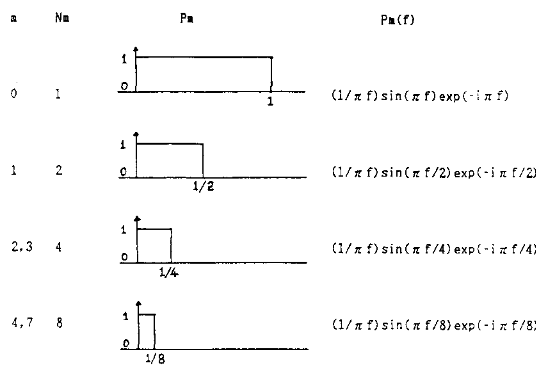


Fig.2. Elementary functions and their Fourier transforms.

Fig.2 shows the Fourier transforms of the elementary pulses, $P_m(f)$, and the number N_m of required elementary pulses to form a sequency ordered Hadamard function of index m from its elementary pulse by using the symmetry method. The transfer function $H_m(f)$ is derived from the pulse transfer function of the system which generates a discrete Walsh function. Therefore the Fourier transform may be obtained simply from inspection of the bits of the Gray code representation as follows [8]

$$W_m(f) = P_m(f)(-1)^m \cdot \prod_q \begin{cases} 2\cos(\pi fq/N_m)\exp[-i(\pi fq/N_m)] \\ \text{or} \\ 2\sin(\pi fq/N_m)\exp[-i(\pi fq/N_m - \pi/2)] \end{cases}$$

$$q = 1, 2, 4, \ldots N_m/2 \quad (9)$$

where the trigonometric function cos or sin are defined from the Gray code representation.

It should be observed, as shown in Fig.3, that the peak frequencies are increasing according to the index m and they are irregular form. When we introduce the concept of frequency to the HVS modelling, this spectrum should be considered as the exact frequency response after normalization to the given order on behalf of incorporation with the spatial frequency.

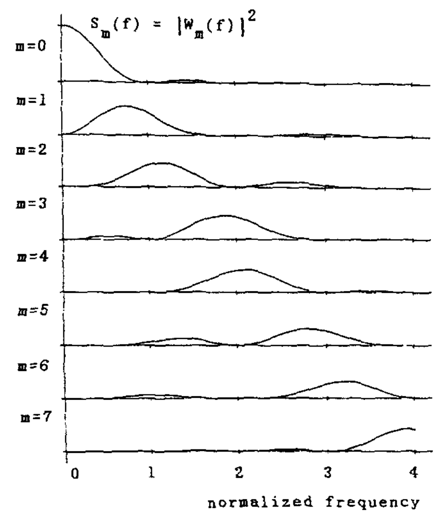


Fig.3. Spectra of Hadamard basis functions.

IV. Derivation of the HT Weighting

Various HVS models have been proposed as above. The generalized model can be represented by

$$H(f) = a(b+cf)\exp(-cf) \tag{10}$$

Constants a,b and c are defined by following three criteria. Firstly, calculation of maximum frequency of this equation is done as

$$f_{max} = \frac{1-b}{c} \tag{11}$$

Next, for a normalized curve, we define the peak response of H(f) as 1, which results in

$$a = \exp(1-b) \tag{12}$$

Last, in the low frequency portion of the model, there are two conflicting phenomena. First as the intercept value is raised, more emphasis is placed on the low frequencies and the image becomes more blurred or blocked. However, the lower intercept below a certain point causes to lose the mean value of the image. Mannos and Sakrison observed, if the peak height is normalized to 1, then the best compromise occurred for an intercept value of 0.05. [5], that is

$$H(0) = 0.05 \tag{13}$$

Thus, by using the above relations, we can derive the values of a, b and c according to f_{max} analytically. We simulate various images by means of mean square error and BMSE, which is a change of errors at the block boundaries and defined as

$$BMSE = [E\{\Delta e(m, jN)\}^2 + E[\Delta e(kN, n)\}^2]^{1/2}$$
$$\Delta e(m, jN) = e(m, jN) - e(m, jN+1)$$
$$\Delta e(kN, n) = e(kN, n) - e(kN+1, n) \tag{14}$$

where (m,n) is an address of pixels, (j,k) is of blocks and N is the block size. It is found that the best HVS function should peak at around 7 cycles/degree. This leads to

$$H(f) = 2.667(.0187 + .140f)e^{-0.14f} \tag{15}$$

and the proposed weighting matrix is shown in Table 1.

Table 1. Proposed weighting matrix for the HT.

.050	.766	.929	.999	.985	.883	.804	.656
.766	.900	.974	.992	.973	.865	.787	.642
.929	.974	.998	.976	.952	.840	.764	.623
.999	.992	.976	.914	.885	.773	.702	.574
.985	.973	.952	.885	.855	.746	.678	.555
.883	.865	.840	.773	.746	.650	.592	.487
.804	.787	.764	.702	.678	.592	.539	.446
.656	.642	.623	.574	.555	.487	.446	.372

The frequency variable f of the model needs to be changed to the normalized spatial frequency f_n and requires a conversion factor f_s as follows

$$f[\text{cycles/deg.}] = f_n[\text{cycles/pel}]\; f_s[\text{pels/deg.}] \quad (16)$$

where f_n, frequency in the HT domain is defined from the Fourier transform of the basis function in Ch. III, respectively, which are not the same as the DCT domain. The f_s depends on the viewing distance, horizontally, f_x, and vertically, f_y. According to Mannos and Sakrison [5], simulated 512 x 512 images were viewed at a distance at which 65 pels subtended 1 degree of vision. For 256 x 256 images, corresponding number of pels is one half of it. The HT basis images are outer products between two one dimensional HT functions. Hence, the weight coefficients, W(u,v), are developed at the corresponding radial frequency from the obtained HVS model. The weighting matrix has the property of symmetric and isotropic model, since we select the same vertical sampling frequency as horizontal one. So the maximum radial spatial frequency, $f_r=(f_x^2+f_y^2)^{1/2}$, is 23 cycles/deg. for the 8 x 8 subblock. Similar weighting matrix for the DCT has been developed in [9]

V. Coding Scheme

Fig.4 describes block diagram of the proposed Hadamard transform coder. At first, thresholding is performed relating to the HVS and activities. There are three activities in a block, direction and surrounding region.

(a) Block activity is an absolute sum of all AC coefficients

$$A_{blk}(j,k) = \sum_{u=1}^{N}\sum_{v=1}^{N}[\,|H(u,v)|\,] - H(1,1) \quad (17)$$

where (j,k) is address of blocks and H(u,v) is the HT coefficient.

(b) Directional activity (A_{dir}) is defined from the coefficients of horizontal (A_{hor}), vertical (A_{ver}) and diagonal activity (A_{dia}) as shown in Fig.5.

(c) Regional activity (A_{reg}) is an absolute sum of the surrounding four blocks.

Measured activities are used for reducing the irrelevant information in that block by thresholding. Thresholding is defined as follows

$$H(u,v) = \begin{cases} 0 & , |H(u,v)| < T \\ H(u,v) & , |H(u,v)| \geq T \end{cases} \quad (18)$$

Following three thresholds are defined.

(a) Block threshold is the HVS weighted block activity

$$T_1 = A_{blk}(j,k) * W(u,v) \quad (19)$$

where W(u,v) is the weighting coefficient as shown in Table 1.

(b) Directional threshold uses directional activity. The coefficients having lower activity are reduced relatively as follows

$$T_2 = T_1 * \left(1 + \frac{A_{dir}}{\text{the largest } A_{dir}}\right) \quad (20)$$

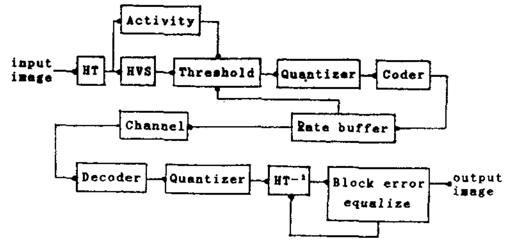


Fig.4. Blockdiagram of the HVS weighted HT image coding.

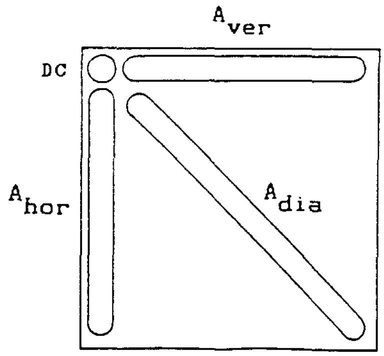


Fig.5. Directional activities.

(c) Regional threshold is based on contrast sensitivity. Regions of high contrast are coarsely coded whereas regions of low contrast are finely coded

$$T_3 = T_2 * \left(1 + \mathrm{Log}\left(1 + \frac{|(A_{reg} - A_{blk})|}{A_{reg}}\right)\right) \tag{21}$$

Uniform 256 level quantizer is designed and used as truncated 32 level version. Variable length coding is performed at coder and decoder by means of Huffman and Run-length coding [10]. To adjust the bit rate despite of more or less active image, rate buffer status affects the threshold and quantizer.

HVS weighting means that only the relevant components should be transmitted and could be visible. So we can remove the inverse weighting of the AC components. This causes increase of mean square error as shown in Table 2.

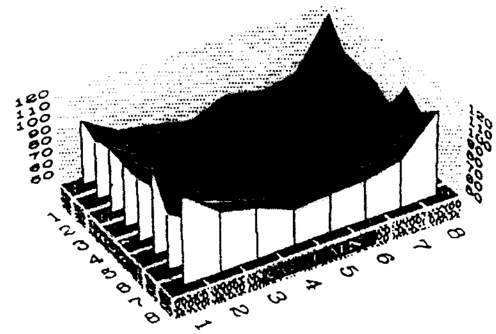
Table 2. Performance comparison of HT coding with and without HVS.

		with HVS	without HVS
'GIRL'	mse[%]	0.092	0.107
	SNR[dB]	30.32	29.65
'COUPLE'	mse[%]	0.076	0.089
	SNR[dB]	31.16	30.44

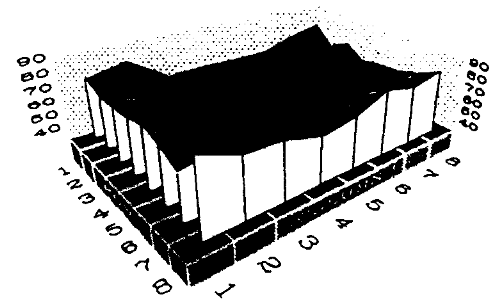
In transform coding, errors in edge area are more significant. It is shown that the larger errors exist in the outer area and the largest one is in corner in Fig.6 a) unequalized distribution. Therefore errors in crossing area of 4 blocks in Fig. 7, must be equalized at receiver. With a view to overcome these errors 4∗4 HT is performed to the crossing areas and multiplied by the weighting coefficients. After equalization, errors are reduced in the edge area as in Fig. 6 b).

VI. Results and Discussions

Various models incorporating HVS have been published in the literature. The peak frequency is an important factor and it varies according to coding scheme used and characteristics of the unitary transform. In the proposed HT scheme we



a) unequalized errors



b) equalized errors

Fig.6. Error distributions in 8x8 subblocks (GIRL. DAT).

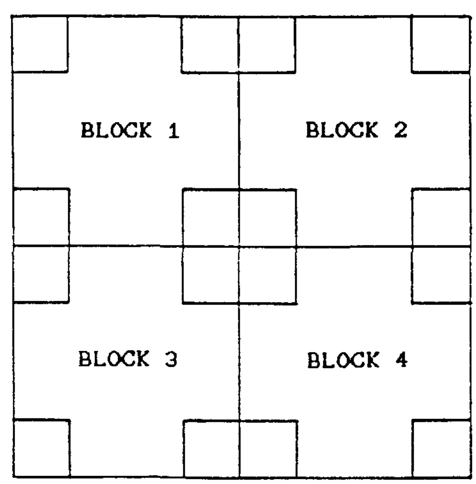


Fig.7. Crossing area of 4 blocks.

find that the HVS weighting is suitable with the HT and maximum sensitive frequency becomes higher than with the other DCT schemes [6] [7] [9], owing to the less energy compaction ability of the HT.

Incorporating HVS with HT, we computed the Fourier spectrum of HT function instead of regular Cal, Sal notation. We obtained the optimum MTF by means of error change at block boundary and conventional MSE.

Since HVS weighting reduces quantization range and bit rate, variable length coding is simplified. HVS weighting implies that only the relevant component can be visible. Non inverse weighting reduces dynamic range and increases mean square error by 0.01% which is insignificant. In Fig.8, reconstructed image by non inverse weighting has more blurred edge area than the HVS weighted image, but the weighting in crossing area of 4 blocks reduced edge errors as shown in Fig.6 and improves the picture quality as shown in Fig.8 d). Consequently, the HVS weighted method results in perceptually more pleasing images and leads to acceptable images at low bit rates.

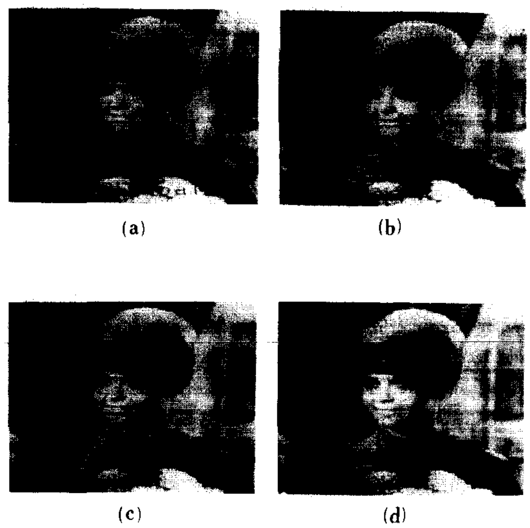


Fig.8. Reconstructed 'GIRL' images at 0.8 bpp
a) Original image.
b) HVS weighted image.
c) Non inverse weighted image.
d) HVS equalized image.

References

[1] J.J. Depalma and E.M. Lowry, "Sine-wave response of the visual system. II. Sine-wave and square-wave contrast," *J. Opt. Soc. Am.*, vol. 52, pp. 328-335, Mar. 1962.

[2] J.J. Hwang and M.H. Lee, "A visual model weighted Hadamard transform," *Proc. Picture Coding Symposium '88, Torino Italy*, paper no. 1. 7, Sept. 12-14, 1988.

[3] M.H. Lee and M. Kaveh, "Fast Hadamard transform based on a simple matrix factorization," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, pp. 1666-1667, Dec. 1986.

[4] N. Ahmed and K.R. Rao, Orthogonal Transforms for Digital Signal Processing, Springer-Verlag Berlin, 1975.

[5] J.L. Mannos and D.J. Sakrison, 'The effects of a fidelity criterion on the encoding of images," *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 525-536, July 1974.

[6] N.B. Nill, "A visual model weighted cosine transform for image compression and quality assessment," *IEEE Trans. Commun.*, vol. COM-33, pp. 551-557, June 1985.

[7] K.N. Ngan, K.S. Leong and H. Singh, "Cosine transform coding incorporating human visual system model," *SPIE Fiber LASE 86*, Cambridge, MA, vol. 707, pp. 165-177, Sep. 1986.

[8] G.S. Robinson, "Fourier transforms of Walsh functions," *IEEE Trans. Electromag. Compat.*, vol. EMC-16, pp. 183-185, Aug. 1974.

[9] B. Chitprasert and K.R. Rao, "Human visual weghted progressive image transmission," *ICCS '88, Intl. Conf. on Commun. Systems*, Singapore, paper no. 17. 1, Nov. 1988.

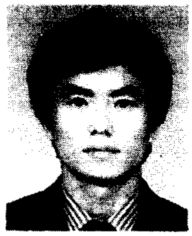
[10] W. Chen and W.K. Pratt, "Scene adaptive coder," *IEEE Trans. Commun.*, vol. COM-32, no. 3, pp. 225-232, Mar. 1984.

[11] Moon Hohee, "The center weighted Hadamand transform," To be appear, *IEEE, Trans. on Circuits and Systems.* *

著 者 紹 介

李 門 浩(正會員)

1967年 전북대학교 전기과 공학사 학위 취득. 1984年 전남대학교 대학원 전자공학과 박사학위 취득. 1985年~1986年 미네소타 주립대 전기과 포스트닥터. 1988年 1月 일본 동경대학 전자과 초청연구원. 1974年 유선 및 무선설비 기사 1급. 1982年 전기통신기술사. 1971年~1980年 남양MBC(주) 부장(송신소장). 1986年 11月 한국통신학회 및 1987年 11月, 대한전자공학회에서 각각 학술상 수상. 1980年~현재 전북대학교 정보통신공학과 부교수(학과장). 주관심분야는 데타통신 및 영상신호처리, 시각신경회로망 등임.



黃 在 政(正會員)

1960年 8月 10日生. 1983年 2月 전북대학교 전자공학과 졸업. 1986年 2月 전북대학교 대학원 전자공학과 석사과정 졸업. 1987年 3月~현재 전북대학교 대학원 박사과정. 1983年~1987年 전주문화방송(주) 기술국, 1987年 3月~현재 군산수산전문대학 통신과 전임강사. 주관심분야는 디지탈 영상신호처리, 시각 신경회로망 등임.