다해상도 영상과 시각 효과를 결합한 영상 부호화에 관한 연구

김 진 태[†]·구 하 성[†]·김 동 욱[††]

요 약

본 논문에서는 HDTV, 디지털 TV, 영상전화, 멀티미디어 영상 등의 다양한 영상에 적용할 수 있는 영상 부호화 기법을 제안한다. 효율적인 영상 압축을 위하여 움직임에 따른 인간의 시각 효과 특성을 결합한 영상 부호화 기법이다. 입력 영상은 웨이브릿 변환을 통하여 다해상도 영상으로 구성된다. 데이터 압축은 저주파수 대역의 움직임 벡터를 움직임 정도에 따라 고주파수 대역 성분들을 처리함에 의해 얻어진다. 즉, 공간과 방향 주파수에 감도에 따라 인간의 눈에 민감하지 않은 대역을 제거한다. 대역내의 모든 블록을 균일하게 처리하는 기존의 방법에 비해 제안한 방법은 같은 전송율에서 재생 영상의 우수한 화질을 얻었다.

A Study on Image Coding with Multiresolution Image and Visual Effects

Jin-Tae Kim[†]·Hasung Koo[†]·Dong-Wook Kim[††]

ABSTRACT

In this paper, an image coding scheme, which can be applied to various kinds of images such as HDTV, digital TV, videophone, and multimedia image is proposed. A technique for an efficient video compression based on characteristics of human visual effects in relation to motion is described. An input image is constructed to multiresolution image by discrete wavelet transform. Data compression is achieved that motion vectors in lower band are utilized to process components in higher band according to the degree of movement. The non-sensitivity parts of the segmented bands are removed according to spatial and directional frequency sensitivity. Compared with the existing methods, which uniformly process all blocks in a band, the quality of the image reconstructed by the proposed method is shown to be superior at the same bit rate.

키워드 : 멀티미디어(multimedia), 시각 효과(visual effects), 다해상도(multiresolution), 웨이브릿 변환(wavelet transformation)

1. Introduction

The purpose of image coding is to express and reconstruct the original image fully with the least bits by the elimination of the redundancy among image data, while the quality of image should be guaranteed. Image data can be efficiently compressed because of their temporal and spatial correlation. In the case of moving images, the temporal correlation coefficient is larger than the spatial one, so the interframe coding using temporal correlation is suitable. Interframe coding is consisted of the part of the motion estimation and compensation (ME/MC) and the part of the prediction error coding.

As one of the techniques to compress such as image and speech signals, discrete wavelet transform (DWT) was first introduced in middle 1980's and has been widely used for data compression since then. DWT decomposes an original signal into components in N independent bands by the analysis and synthesis filters, and compresses the data by applying an appropriate coding technique to each band. DWT has the following merits. First, analysis and synthesis for the band of the signal is easy. Second, the effect of the error on the overall image after reconstruction is small because an error occurring in coding a signal in a band is independent of the other bands. Third, a satisfactory reconstruction of the signal can be made in the subjective point of view by controlling the number of bits allocated to each band. Fourth, parallel-

※ 본 연구는 한국과학재단 목적기초연구(2000-1-30300-003-2) 지원으로 수행되었음.
†정 회 원 : 한서대학교 컴퓨터정보학과 교수
††정 회 원 : 전주대학교 정보기술컴퓨터공학부 교수
논문접수 : 2001년 2월 20일, 심사완료 : 2001년 5월 15일

processing techniques can be applied

Human eyes judge the quality of image. Several studies taking into account human visual system (HVS) have attempted to improve the subjective quality since the coding algorithm is closely associated [1-3]. Compared with other transform coding, DWT utilizing the information on HVS was shown to have good quality in still image coding. In general, HVS applied to spatial frequency is widely used to model the modulation transfer function [3-5]. However, in cases of sequential images such as TV signals, the results may be against the characteristics of HVS for the moving images since the HVS model on spatial frequency domain is applied to each frame.

In this paper, we propose an image coding method based on characteristics of visual effects, which can be applied to various applications such as videophone and HDTV images. Since we are concerned with the compression of moving images, it will be appropriate to take into account human visual characteristics in the spatial and temporal domain. For this reason, the proposed approach incorporates the function of the perceived resolution loss as a result of movement in scenes in order to compress video signals with wavelet transform. Simulation results show the proposed scheme can compress image signals without subjective degradation of the reconstructed images.

The paper is organized as follows. Expression of multiresolution techniques is presented in Section Ⅱ. In Section Ⅲ, the characteristics of human visual effects concerning moving images are examined. In Section Ⅳ, a new coder utilizing the human visual effects described in Section Ⅲ is designed. In Section Ⅴ, computer simulations are carried out to demonstrate the effectiveness of the proposed algorithm. And the results are discussed in Section Ⅴ. Conclusion of this paper is given in Section Ⅵ.

2. Expression of Multiresolution

Wavelet theory has been developed in mathematics and engineering fields in order to study and analyze the signals with various scale and resolution [6]. Morlet, Grossmann, and Meyer in middle 1980s had accomplished its general mathematical system, after then Daubechies and Mallat developed it to discrete signal processing.

Fourier transform is used for the analysis of stational signals, but it does not satisfy local characteristics because it integrates whole spatial region. For nonstational signals, their spectrum has the characteristics of time varying, while short time Fourier transform has uniform time resolution independent on frequency, so it is a disadvantage for the analysis of signals. Wavelet transform has frequency-varying time-resolution and it favors analyzing of signals.

Expression of multiresolution offers useful hierarchical processing techniques for the analysis of image information. The signals of low resolution have larger structure representing contour of images. So it is more efficient that the analysis of image signals is gradually carried out from low resolution to higher resolution. These multiresolution techniques are studied at the parts of low-level signal processing, and the hierarchical motion estimation method is also similar to them.

First, the general concept of wavelet transform is considered as the following equation,

$$V_{2^j} \subset V_{2^{j+1}} \forall j \in Z \; Z : \textit{integerset.} \tag{1}$$

It represents that the approximate signals $V_{2^{j+1}}$ at the resolution 2^{j+1} include all of the signals V_{2^j} at the lower resolution 2^j. When the approximate values of $f(x)$ are calculated at resolution 2^j, several information of original signal $f(x)$ is lost. And approximate signals converse to original signals as resolution is increased, while the approximate signals lose the information gradually and after all, converse to zero as resolution is reduced.

$$\lim_{j \to +\infty} V_{2^j} = \bigcup_{j \to -\infty}^{+\infty} V_{2^j} \quad \textit{is dense in } L^2(R) \tag{2}$$

$$\lim_{j \to -\infty} V_{2^j} = \bigcup_{j \to -\infty}^{+\infty} V_{2^j} = \{0\} \tag{3}$$

Approximating operation for all resolutions are similar, and the space of approximate function is obtained by scaling of each approximate function as follows :

$$f(x) \in V_{2^j} \Leftrightarrow f(2x) \in V_{2^{j+1}} \quad \forall j \in Z. \tag{4}$$

Translation of signals is

$$f(x) \in V_{2^j} \Rightarrow f(x - 2^{-j}k) \in V_{2^j} \quad \forall k \in Z \tag{5}$$

and it does not affect the variation of resolution. If R is real number and $L^2(R)$ is a vector space with finite energy, the approximation of multiresolution is the sequence of the closed subspace $(V_j)_{j \in Z}$ satisfying the above conditions [7].

Projection components of original signal $f(x)$ on vector space are calculated by dividing $f(x)$ into normal projection bases $\phi_{2^j}(x-2^{-j}n)$. That is, if the projection operator on vector space V_{2^j} is named $A_{2^j}f(x)$, the following relation is formed.

$$A_{2^j}f(x)=2^{-j}\sum_n(\langle f(u),\varphi_{2^j}(u-2^{-j}n)\rangle\cdot\phi_{2^j}(x-2^jn),\quad \forall f(x)\in L^2(R) \tag{6}$$

where discrete approximation of $f(x)$ at resolution 2^j is represented as inner product.

$$A_{2^j}^d f=(\langle f(u),\varphi_{2^j}(u-2^{-j}n)\rangle)_{n\in Z} \tag{7}$$

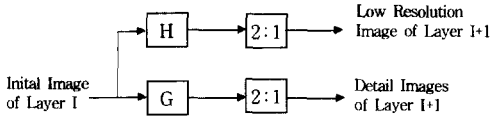
Actually it is represented as pyramidal transform.

$$A_{2^j}^d f=\sum_{k=-\infty}^{\infty}\tilde{h}(2n-k)\,A_{2^{j+1}}^d f\quad where,\ \tilde{h}(n)=h(-n). \tag{8}$$

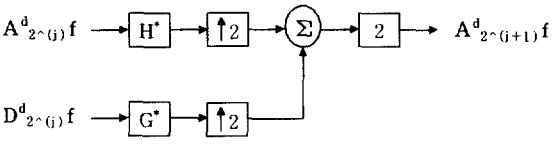
When $j<0$, every discrete approximation $A_{2^j}^d f$ is calculated by repeating of transform of $A_1^d f$. And $D_{2^j}f=(\langle f(u),\Psi_{2^j}(u-2^{-j}n)\rangle)_{n\in Z}$ is named the discrete detail at resolution 2^j. It indicates the information difference of $A_{2^{j+1}}^d f$ and $A_{2^j}^d f$.

$$D_{2^j}f=\sum_{k=-\infty}^{\infty}\tilde{g}(2n-k)\,A_{2^{j+1}}^d f\quad where,\ \tilde{g}(n)=g(-n). \tag{9}$$

Original signals are analyzed to discrete approximation and detail signals by filtering and subsampling, and synthesized conversely. (Figure 1) shows the analysis and synthesis of multiresolution images.



(a) Analysis



(b) Synthesi

(Figure 1) Multiresolution model

3. Characteristics of Visual Effects

The human eye behaves as a two-dimensional low-pass filter for spatial patterns, with a high-frequency cut-off of about 60 cycles per degree of foveal vision and significant low-frequency attenuation below about 0.5 cycles. Thus, high spatial frequencies in the image are not seen and need not be transmitted. The eye also acts as a temporal band-pass filter having a high-frequency cut-off 50 and 70 Hz depending on viewing conditions. Flicker is more disturbing at high display luminance and low spatial frequencies [8].

Noise and amplitude distortion are generally less visible at high luminance levels than at mid and low luminance values, again depending on viewing conditions such as overall scene brightness and ambient room lighting. High and low frequency noise is less visible than mid-frequency noise. Amplitude distortions are also less visible near color transitions ; such as occur at boundaries of objects in a scene. This is termed spatial masking, since the transitions mask the distortions. Although amplitude distortions are less visible near color transitions, small spatial shifts in the transitions themselves are easily visible and annoying.

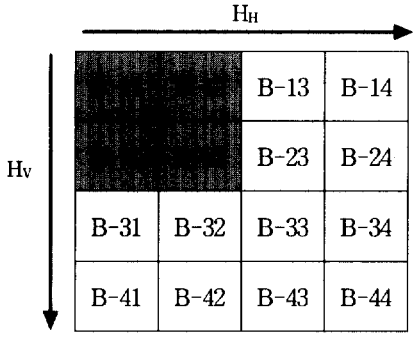
Temporal masking also occurs. For example, shortly after a television scene change, the viewer is relatively insensitive to distortion and loss of resolution. This is also true of objects in a scene that are moving in an erratic and unpredictable fashion. However, if a viewer is able to track a moving object, then resolution and distortion requirements are the same as for stationary areas of a picture.

Temporal masking and perception of temporally changing stimuli are extremely important in interframe coding. However, temporal masking is complicated by at least two facts ; 1) television cameras integrate the image of any object on the target and, thus there is motion-related blurring and resolution loss ; 2) perception of a moving object depends heavily on whether or not the object is tracked by the eye. The psychophysical literature contains many facts about the perception of temporally changing stimuli. However, their application to coding is still in its infancy. Instead, several applied studies have attempted to evaluate the loss of perceive resolution (spatial and amplitude) as a result of movement in scenes. If movement is drastic, such as with a scene change (when TV cameras are switched), the perceived spatial resolution is reduced significantly immediately after the scene change. In fact, the perceived spatial resolution of the new scene may be reduced down to only one-tenth of normal without detection provided that full resolution is restored gradually within about half a second. Experiments also show that if the eye tracks a moving

object, then perceived resolution due to camera integration dominates any reduction in resolution introduced by the visual system. However, when erratically moving objects be not tracked, the loss of perceived spatial resolution due to the visual system is significant. In practical television viewing, most displayed movement is not easily tracked an object and how accurately he tracks it. Also, since in many visual communication systems a transmitted picture may be viewed by many observers (e.g. broadcast TV), it is not clear how the resolution loss of nontracked objects can be used to improve the coding efficiency.

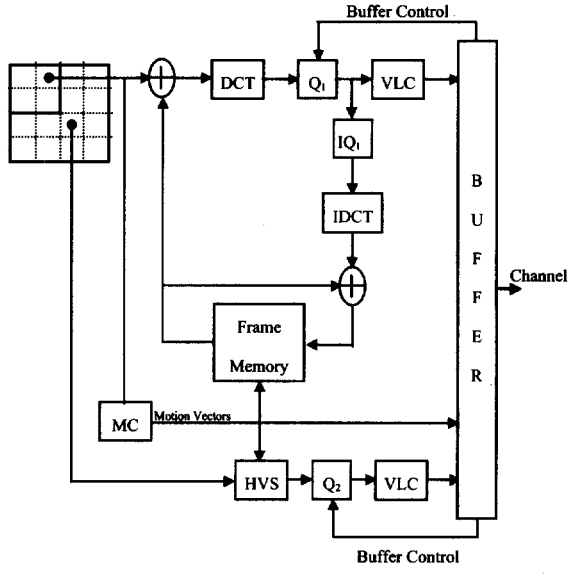
4. A New Image Coding Scheme

In this paper, the function of the perceived resolution loss as a result of movement in scenes is used to compress video signal with wavelet transform. The human visual characteristics on moving images are used in deciding efficient bands of moving parts according to its motion. The resolution of each image part depends on motion and the number of efficient bands. This may be regarded as a filtering process.



(Figure 2) 16-band of images

Each frame has been decomposed into 16 equal bands as shown in (Figure 2). The bands of B-11, 12, 21, and 22 have much influence upon the quality of the reconstructed images, and the 12 remaining bands have little influence on them. The former will be referred to as *lower band* and the latter *higher band*. Now, we can apply the characteristics of low-pass filter in human vision systems along with temporal axis and those of bandpass filter in the human vision system, which is dependent on the speed of motion in the spatial frequency domain, to these bands. In other words, the image is processed differently depending on the moving area or background area, which is composed of flat and detail regions. Also, the components in lower bands are processed more cautiously than the ones in higher bands. By doing this, the reconstructed image can retain high quality in the subjective sense.



(Figure 3) Encoder of the proposed scheme

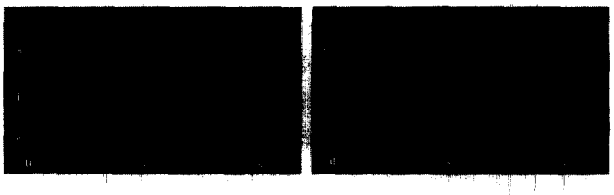
(Figure 3) is an encoder of the proposed scheme in this paper. For interframe processing, each band signals are divided into lower band and higher band signals. Motion compensation is applied only to lower band, resulting in the construction of motion vectors that will be used to process higher band signals. Each band is composed of several nonoverlapped blocks. ME/MC are performed in the unit of block. The prediction errors of the lower band are quantized uniformly after discrete cosine transform (DCT). The quantized coefficients are scanned zigzagly and coded with variable length. The variable length codes (VLC) are determined according to the probability of occurrence of the combination of continued zeros followed by nonzero values. The blocks in the higher band are classified into moving and background blocks, according to the corresponding motion vectors of the lower band. If the magnitude of motion vector of the lower band is less than the threshold, the block gets background block ; otherwise it gets moving block. Since the components of higher band for moving blocks are insignificant, they can be eliminated without subjective degradation of image quality. These blocks are skipped in the processing of components of higher band. Background blocks are quantized uniformly after DCT. The quantized coefficients are scanned zigzagly and coded with variable length. In the case of intraframe processing, it removes the ME/MC and HVS parts of interframe processing.

That is, in lower band, signals in B-11, 12, 21, and 22 are

compensated using the motion vector of B-11. The motion vectors of B-11 contain motion information for the higher bands coding. Size of the motion estimation block is set to 4×4. Maximum displacement of the motion vector is ±3. The size of the DCT block is 4×4. While, in higher band, signals are not motion compensated even though the motion information for their coding is utilized. The size of a block is 4×4. The implementation of this procedure is determined by the magnitude of corresponding to motion vectors. And signals of B-23, 32, 33, 34, 43, and 44 in these bands are not processed due to ineffectiveness for the better quality of reconstructed image.

5. Computer Simulations and Discussion

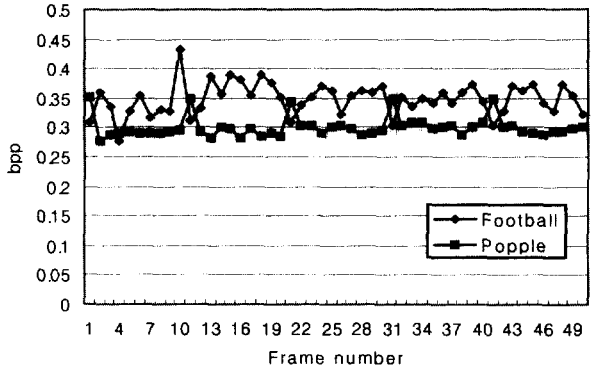
We used 'Football' and 'Popple', which are standard images for the moving picture experts group (MPEG). 'Football' contains the regions of large motion and has a lot of high-frequency components in the background. On the other hand, 'Popple' contains vary static background and rotational motions of objects at fixed positions. The image signals are composed of luminance components and chrominance components. In this simulation, we considered only the luminance component. The size of a test image is 720×480. The frame rate of a sequence image is 30 frames/second, and the interframe processing of nine frames follows the intra-frame processing of one frame. In this experiment, we were processing 50 frames of an image and used IPPP structure of group of picture (GOP). Biorthogonal analysis and synthesis filters are used for 16-band decomposition. The taps of filters are 7 and 9 [9-10]. Figure 4 is scaling and wavelet function of the used filters. We set the quantization step sizes of the lower and higher bands to 2 and 8, respectively. The results of computer simulation are shown in (Figure 5) and (Figure 6).



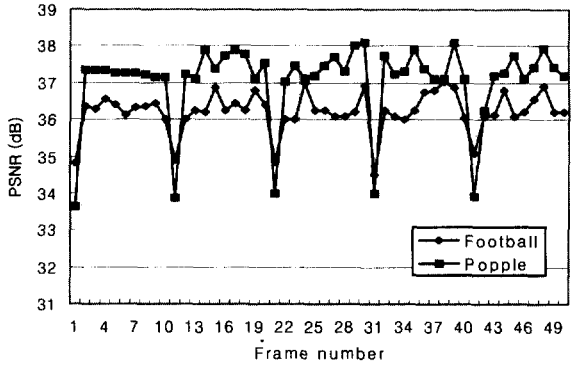
(a) Scaling function (b) Wavelet function

(Figure 4) Filter characteristics

From (Figure 5) and (Figure 6), we can observe that the images processed by the proposed method have high PSNR and good quality in overall respects. Bit allocations to the lower and higher bands were set differently, depending on the characteristics of the images. When we were applied to the Football image, the bit allocation ratio in the higher and lower bands was about 3 : 1 and the compression ratio of the image was about 24 : 1. In the Popple image, the bits were allocated at about 1 : 1 in lower bands, and the compression ratio was about 27 : 1. In this simulation, we controlled the step size of the uniform quantizer for steady transmission rate, and the control range was limited by the image quality. However, one can increase the performance of the proposed method by use of a quantizer with non-uniform step size.



(Figure 5) Results of computer simulation (b/p)



(Figure 6) Results of computer simulation (PSNR)

The proposed method depends on the motion information in images ; however, it shows good performance regardless of the degree of motion. The Football image, which has a lot of moving parts, was obtained at high data compression ratio in the higher bands. On the other hand, the number of bits to be transmitted increased in the lower bands because of the large prediction error values. In the case of 'Popple' image, we did not obtain high coding gain in the higher bands because of the small motion. However, we obtained high coding gain in the lower bands because of the

small prediction error. The PSNR's we obtained in this simulation are different depending on the images, however, the quality of the reconstructed images are subjectively good because the images are processed by taking into account the characteristics of a human vision system. We obtained a relatively high quality of image over all frames.

6. Conclusion

In this paper, we proposed a new image coding scheme based on human visual effects. Human eyes have the characteristics of a lowpass filter in the temporal domain, and those of a bandpass filter in the spatial domain. These visual characteristics were applied to the design of efficient image data compression. By simulation, we confirmed that the proposed method based on characteristics of visual effects could compress image signals without subjective degradation of the reconstructed image by adaptively processing depending upon the motion. The proposed method was developed to be applied to full digital HDTV images. The proposed method, with the combination of techniques taking the cost of ME, frame memory, and less complexity into account, can be easily applied to HDTV images. Further study for the detection of real motion and buffer control will improve the performance of the proposed scheme.

References

[1] T. G. Stockham, "Image processing in the context of a visual model," Proc. IEEE, Vol.60, No.7, pp.828-842, 1972.

[2] D. J. Granrath, "The role of human visual models in image processing," Proc. IEEE, Vol.69, No.5, pp.552-560, 1981.

[3] J. L. Marros and D. J. Sakirson, "The effect of a visual fidelity criterion on the encoding of images," IEEE Trans. Inform. Theory, Vol.IT-20, No.4, pp.525-536, 1974.

[4] N. B. Nill, "A visual model weighted cosine transform for image compression and quality assessment," IEEE Trans. Communications, Vol. COM-33, No.6, pp.551-557, 1985.

[5] K. N. Ngan, K. S. Leong, and H. singh, "A HVS-weighted cosine transform coding scheme with adaptive quantization," in Proc. Visual Communications and Image Processing, Vol.1001, pp.702-708, 1988.

[6] O. Rioul and M. Vetteri, "Wavelets and signal processing," IEEE Signal Processing Magazine, Vol.8, No.4, pp.14-38, 1991.

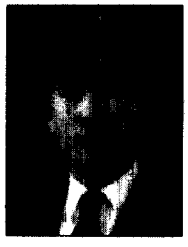
[7] J. C. Feauveau, P. Mathieu, M. Barlaud, and M. Antonini, "Recursive biorthogonal wavelet transform for image coding," in Proc. ICASSP, pp.2649-2652, 1991.

[8] A. N. Netravali and B. G. Haskell, Digital Pictures, New York ; Plenum, 1988.

[9] C. H. Keon, "Subband image coding with biorthogonal wavelets," IEICE Trans. Fundamentals, Vol.E75-A, pp.871-881, 1992.

[10] M. Antonini, M. Barlaud, P. Marthieu, and I. Daubechies, "Image coding using vector quantization in the wavelet transform domain," in Proc. ICASSP, pp.2297-2300, 1992.

[11] J. T. Kim, H. Koo, and D. W. Kim, "A discrete wavelet transform coder based on human visual characteristics," in Proc. CISST, pp.475-481, 1999.



김 진 태

e-mail : jtkim@hanseo.ac.kr

1987년 중앙대학교 전자공학과 졸업 (학사)

1989년 중앙대학교 대학원 전자공학과 졸업(석사)

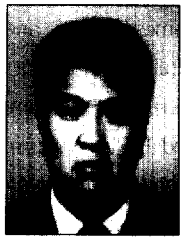
1993년 중앙대학교 대학원 전자공학과 졸업(박사)

1993년~1994년 중앙대학교 기술과학연구소 선임연구원

1994년~1995년 서울대학교 자동제어특화연구센터 선임연구원

1995년~현재 한서대학교 컴퓨터정보학과 조교수

관심분야 : 영상통신, 얼굴인식, 디지털 워터마킹 등



구 하 성

e-mail : hskoo@hanseo.ac.kr

1989년 광운대학교 전자통신공학과 졸업 (학사)

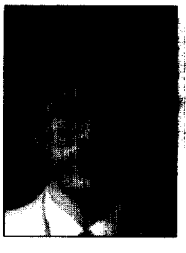
1991년 광운대학교 대학원 전자통신공학과 졸업(석사)

1995년 광운대학교 대학원 전자통신공학과 졸업(박사)

1995년~1997년 기아정보시스템 근무

1997년~현재 한서대학교 컴퓨터정보학과 조교수

관심분야 : 지문인식, 웹을 이용한 영상 보안 등



김 동 욱

e-mail : dwkim@jeonju.ac.kr

1987년 성균관대학교 전자공학과 졸업 (학사)

1992년 중앙대학교 대학원 전자공학과 졸업(석사)

1996년 중앙대학교 대학원 전자공학과 졸업(박사)

1997년~1998년 청운대학교 전자공학과 전임강사

1998년~현재 전주대학교 정보기술컴퓨터공학부 조교수

관심분야 : 영상통신, 통신신호처리, MPEG 등