비전 기반의 손 동작 검출 및 추적 시스템

정회원 박 호 식*, 종신회원 배 철 수*

Vision-based hand Gesture Detection and Tracking System

Ho-Sik Park* ***Regular Member***, Cheol-soo Bae* ***Lifelong Member***

요 약

본 논문에서는 비전 기반의 손동작 검출 및 추적 시스템을 제안하고자 한다. 기존의 손동작 인식 시스템은 정적인 관측 환경에서 배경을 제거함으로 손을 검출하는 단순한 방법을 사용함으로써, 카메라의 움직임, 조명의 변화 등에 의해 견실하지 못하였다. 그러므로 본 논문에서는 기하학적 구조에 의하여 손의 외형을 인식하여 검출할 수 있는 통계적 방법을 제안하였다. 또한 카메라의 각도에 의한 손이 겹쳐 보이는 문제를 줄이기 위하여 다중 카메라를 사용하였으며 비동기식 다중 관측으로 시스템의 범용성을 향상시키었다. 실험 결과 제안된 방법이 기존의 외관을 이용한 방법보다 3.91% 개선된 99.28%의 인식률을 나타내어 제안한 방법의 효율성을 입증하였다.

Key Words : hand tracking, gesture interface, object recognition, computer vision

ABSTRACT

We present a vision-based hand gesture detection and tracking system. Most conventional hand gesture recognition systems utilize a simpler method for hand detection such as background subtractions with assumed static observation conditions and those methods are not robust against camera motions, illumination changes, and so on. Therefore, we propose a statistical method to recognize and detect hand regions in images using geometrical structures. Also, Our hand tracking system employs multiple cameras to reduce occlusion problems and non-synchronous multiple observations enhance system scalability. In this experiment, the proposed method has recognition rate of 99.28% that shows more improved 3.91% than the conventional appearance method.

Ⅰ. 서 론

기존의 컴퓨터 인터페이스 방법인 마우스나 키보드로는 가상 공간에서 객체를 자연스럽게 조작하기란 쉽지 않으므로, 자연스럽고 지적인 새로운 인터페이스가 필요하다. 그 중 손동작 인식은 빠른 의사 전달 특징과 함축적 의미를 지닌 동작을 통해 다양한 정보를 전달할 수 있으므로 최근 활발한 연구가 진행되고 있다.

손동작 인식은 크게 글로브 기반 방법(glove-based method)[1]과 비전에 기반한 방법(vision-based method)[2]으로 나눌 수 있는데, 전자는 실시간으로 손의 모양과 손가락의 움직임을 검출할 수 있으나, 장치 착용에 따른 불편함과 손의 운동 범위와 같은 여러 가지 제약 조건을 수반하고, 후자의 방법은 카메라를 통한 입력 영상을 사용하는 방법으로 전자의 방법보다 장비가 간단하며, 행동반경이 자유롭고, 사용자 불편함 없이 자연스러운 손동작 인식이 가능하다는 장점이 있다. 그러나 정적인 관측 환경에서 배경을 제거하여 손을 검출하는 방법[3,4,5]을 사용함으로써 카메라의 움직임, 조명의 변화 등에 견실하지 못하다.

본 논문에서는 그림 1과 같이 기하학적 구조에 의하여 손 외형을 인식한 후 검출하는 계층적인 통

* 관동대학교 전자통신공학과 영상처리연구실 (mediana@netsgo.com)
논문번호 : KICS2005-11-482, 접수일자 : 2005년 11월 29일

계적 방법을 사용하였다. 먼저 화소값 분포에 의해 초기 검출로 형태 모델을 구성하였고, 초기 검출에서 추출한 저차원 정보와 본 검출에서 얻은 정보로 색상과 움직임 모델을 정의하였다. 또한 카메라 각도에 의한 손이 겹쳐 보이는 문제를 줄이기 위하여 다중 카메라를 사용하였다.

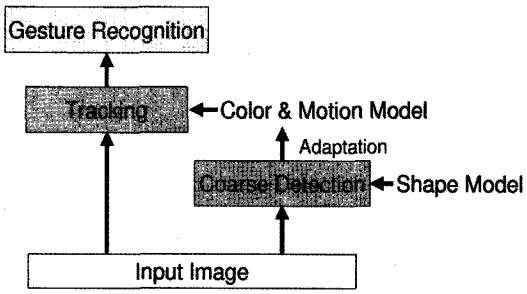


그림 1. 계층 적 인식
Fig. 1. Hierarchical Recognition.

Ⅱ. 손영역 추적

본 논문에서는 비동기 다중 관측에 의한 손 영역 추적 방법을 제안한다. 제안된 시스템의 구성은 그림 2와 같이 다중 관찰로 얻은 순차 영상을 이용하여 양 손의 움직임을 추적한다. 여기서 각각의 카메라 시스템(관측 노드)은 각각의 처리기에서 입력 영상을 독립적으로 처리한다. 추적 노드에서는 2차원 위치를 추정하여 추적 모델을 생성하고, 각각의 관측 노드에서는 영상 특징과 추적 모델을 정합한다. 정합 과정이후에 목표와 정합되는 특징 정보(위치, 손 모양)는 추적 노드로 되돌아가고 추적 모델을 갱신한다. 정합되지 않은 나머지 특징은 발견 노드로 보내진다.

발견 노드는 관측 노드에서 보내진 정합되지 않은 영상 특징을 근거로 하여 새로운 외형의 손을 검출한다. 이러한 정보는 추적 노드로 보내져 새로운 추적 절차를 이룬다. 추적 모드는 발견과 관측 모드에서 보내진 정보로 모델을 갱신하고 객체를 추적한다.

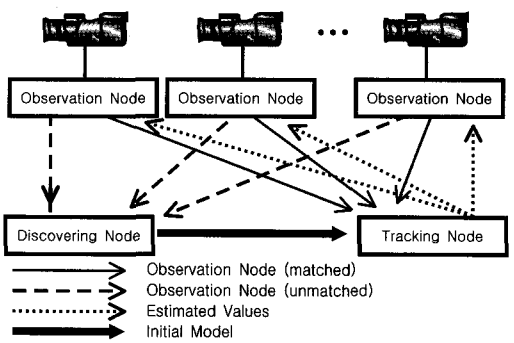


그림 2. 제안된 시스템의 구성도
Fig. 2. Proposed system configuration

Ⅲ. 손 영상의 기하학적 구조

본 논문에서는 독립적인 객체 색상 변화에 따른 공통의 영상 특징을 구하여 손을 검출하는 방법을 제안한다. 손 영역 추적에서 색상/휘도가 일정하다면 외형적 추적 모델은 매우 효과적이지만, 실제적으로는 기대한 만큼의 효과를 얻을 수 없다. 그러므로 제안된 방법은 각각의 영상간의 차이로부터 목표 객체에 대한 기하학적 구조를 추출하여 사용한다.

3.1 입력 영상 변환

대다수의 동작 인식시스템은 손 외형 방향에 대해 자유롭지 못하다. 그러므로 이러한 문제를 해결하고자 방향에 대해 자유로운 형태 모델은 구성하고자 그림 3과 같은 영상 변환을 수행하였다. 모델 구성을 위한 원영상 영역을 각도를 3도씩, 반지름이 1부터 60 화소가 되도록 영상을 변환하여 60×120 화소의 영상을 생성하였다.

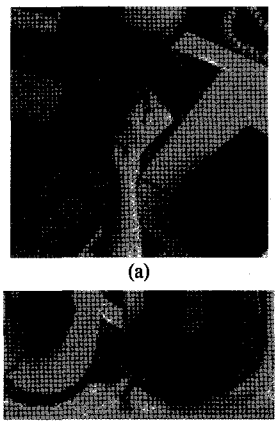


그림 3. (a) 원 영상 (b) (r,θ) 영상 변환
Fig. 3. (a) Original Image (b) (r,θ) Image Translation

3.2 거리맵(distance map)연산

제안된 방법은 영상에서 화소값 분산의 차이를 측정하여 마할라노비스(Mahalanobis) 거리를 사용하였다. 영상을 일련의 블록(작은 영역)으로 나누고 두 블록 사이의 거리를 계산하였다. 그리고 하나의 지역에 대한 거리의 조합인 거리맵을 구성하였다.

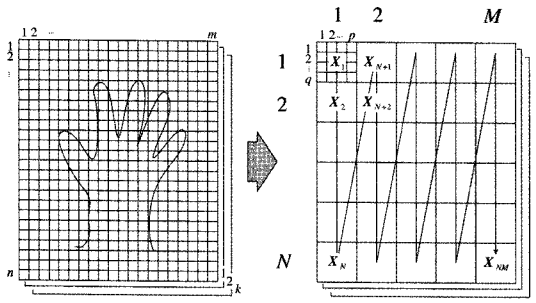


그림 4. 블록 분리
Fig. 4. Block Separation

그림 4의 왼편 그림과 같이 $m \times n$ 의 입력 영상이 있다면, 각 화소 $x_{s,t}$ $(1 \le s \le m,\ 1 \le t \le n)$는 k-차원의 벡터로 나타낼 수 있다.

$$x_{s,t} = [y_1 \quad y_2 \cdots y_k]' \tag{1}$$

그레이 영상을 위해서는 k=1을 칼라 영상을 위해서는 k=3을 사용한다.

각 블록은 $p \times q$ 화소로 되어있다. 수평으로 M개의 블록을 수직으로 N개의 블록을 얻을 수가 있다. 각각의 블록에 대해 1,…, MN까지의 고유번호를 부여하였다. 각각의 블록 X_l에 대한 공분산 행렬과 평균 벡터를 구하였다.

$$\bar{x}_l = \frac{1}{pq} \sum_{(s,t) \in X_l} x_{s,t} \tag{2}$$

$$\Sigma_l = \frac{1}{pq} \sum_{(s,t) \in X_l} x_{s,t} (x_{s,t} - \bar{x}_l)(x_{s,t} - \bar{x}_l)' \tag{3}$$

결국 매 두 블록 사이의 마할라노비스 거리는 연산하여 거리맵 D가 결정된다.

$$D = \begin{bmatrix} 0 & d_{(1,2)} & \cdots & d_{(1,MN)} \\ d_{(2,1)} & 0 & \cdots & d_{(2,MN)} \\ \vdots & \vdots & \ddots & \vdots \\ d_{(MN,1)} & d_{(MN,2)} & \cdots & 0 \end{bmatrix} \tag{4}$$

여기서 $d_{(i,j)} = (\bar{x}_i - \bar{x}_j)'(\Sigma_i + \Sigma_j)^{-1}(\bar{x}_i - \bar{x}_j)$ 이다.

그림 5에 계산된 거리맵의 예를 나타내었다. 연산을 위하여 12×24 블록을 가진 영상에서 5×5 블록을 사용하였다.

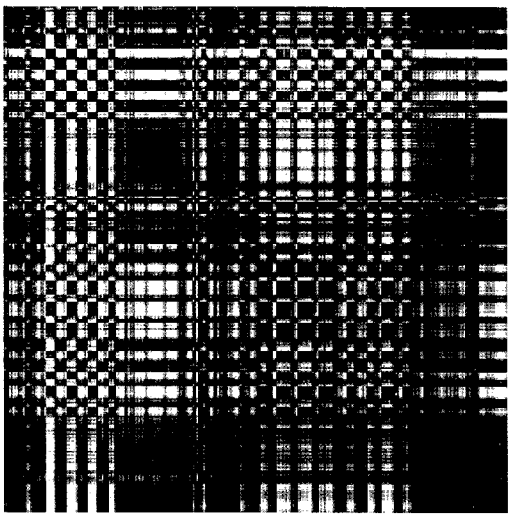


그림 5. 거리맵 연산의 예
Fig. 5. Example of Distance Map Calculation

3.3 통계적 표현

목표 객체를 통계적 표현으로 구성하였다. 먼저, 손 방위 차이에 대한 영향을 줄이기 위해, 맵 D로부터 주면 영역에 대한 평균 맵 A를 연산하였다.

$$A = \begin{bmatrix} a_{(1,1)} & \cdots & a_{(1,MN)} \\ \vdots & \ddots & \vdots \\ a_{(N,1)} & \cdots & a_{(N,MN)} \end{bmatrix} \tag{5}$$

여기서 $a_{(j,k)} = \sum_{i=0}^{M-1} d(j + i \times N,\ u + (v+1 \bmod M) \times N)$ 이다.

$u = k \bmod N$, $v = (k-u)/N (1 \le j \le N, 1 \le k \le MN)$.

그림 6에 정규화 된 거리맵을 나타내었다.

그림 6. 정규화된 거리맵
Fig. 6. Normalized Distance Map

목표 개체에 대한 샘플 영상과 샘플이 아닌 영상의 거리맵을 계산하였다. 수식 (4)에 의하여 A_1^{obj},..., A_K^{obj}는 객체 영상에 대한 거리맵으로, A_1^{bck},..., A_K^{bck}는 객체가 아닌 영상의 거리맵으로 구성하였다. 여기서 K는 양 분류에 대한 샘플 수이다. 그리고 A_K^{obj} 와 A_K^{bck} 거리 분산을 나타내었다.

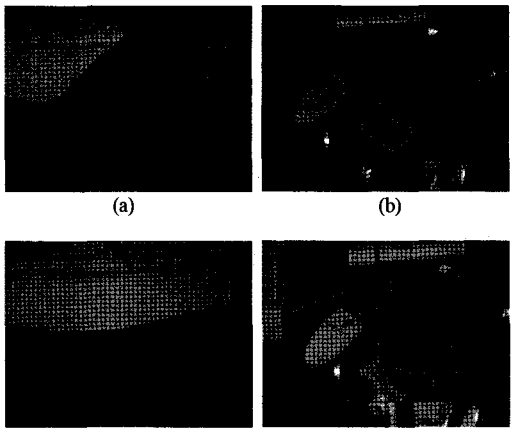


그림 7. 모델 구성에 사용한 샘플 영상
(a)(b) 손 영상 (c)(d) 배경 영상
Fig. 7. Sample Images used for Model Construction
(a)(b) Hand Image (c)(d) Background Image

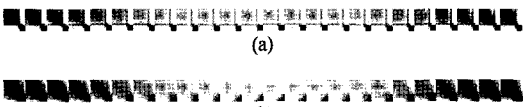


그림 8. 평균 거리맵
(a) 손 영상 (b) 배경 영상
Fig. 8.Averaged Distance Maps
(a) Hand Image (b) Background Image

1750개의 손영상과 손이 아닌 배경 영상에 대한 통계적 표현을 구성하였다. 그림 7에 손영상과 배경 영상의 일부를 나타내었고, 그림 8에 정규화된 거리맵 평균을 나타내었다.

A^{obj}와 A^{bck}의 두 개의 거리맵 분산을 비교하기 위하여, 각 요소 (i, j)에 대한 두개의 분산 사이의 마할라노비스 거리를 구하였다.

$$w_{(i,j)} = \frac{(\bar{a}_{obj(i,j)} - \bar{a}_{bck(i,j)})^2}{\sigma^2_{obj(i,j)} + \sigma^2_{bck(i,j)}} \quad (6)$$

거리맵은 약 26,000개의 요소들로 구성되어있으므로 다음과 같은 방법으로 요소의 수를 줄였다. 먼저 수식 6에서 얻은 요소 $((u_{r,1},v_{r,1}), (u_{r,2},v_{r,2}),..., (u_{r,r},v_{r,r}))$중 큰 요소인 r을 선택하였다.

거리맵 A_k에서 선택된 요소 r의 조합을 r차원 거리 벡터인 A'_k ($A'_k=[a_k(u_{r,1},v_{r,1}), \ldots, a_k(u_{r,r},j_{r,r})]'$로 표현하였다. 선택 후에 손 영상 식별을 위해 LDM (linear discriminant method)을 사용하였다.

손과 배경 영상에 대한 r 차원 벡터의 분산으로부터 식별공간에 대한 거리 벡터의 변환인 투영 행렬 B_r을 정의한다. 여기서는 손 영상에 대한 거리 벡터 $A'_{r,k}$의 분산과 가우시안 분산 w으로 구성된 가우시안 혼합모델로 표현하였다. 각 분산에 대한 매개변수는 EM (Expectation Maximization) 알고리즘에 의해 정의하였다,

$A'^{obj}_{r,k}$ (손 영상)와 분산에 대한 평균과 공분산 행렬을 $\bar{A}'^{obj}_r$와 $\Sigma_{\bar{A}'^{obj}_r}$으로, $A'^{bck}_{r,k}$ (배경 영상)의 분산에 대한 평균과 공분산 행렬을 $\bar{A}'^{bck}_r$와 $\Sigma_{\bar{A}'^{bck}_r}$이라 하면 투영행렬 B_r은 다음과 같다.

$$B_r = \left(\Sigma_{\bar{A}'^{obj}_r} + \Sigma_{\bar{A}'^{bck}_r}\right)^{-1}\left(\bar{A}'^{obj}_r - \bar{A}'^{bck}_r\right) \quad (7)$$

3.4 인식 과정

그림 9에 인식과정을 나타내었다. 인식과정에서는 입력 영상의 각 부분마다 거리맵을 계산하여 이미 구성한 모델과 비교하여야 한다.

각 영역에 대한 비교 값은 다음에 의하여 구할 수 있다.

$$A_r B'_{r,i} > \text{threshold} \qquad (i \in 1,\ldots,w) \quad (8)$$

영역 $i(i=1,...,w)$에 대한 비교값이 문턱치보다 크다면 이 영역 안에 목표 객체가 있는 것이다.

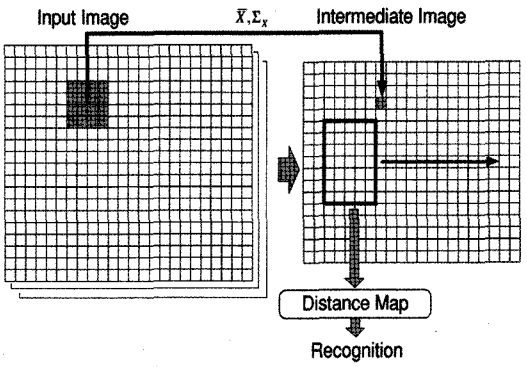


그림 9. 인식 과정
Fig. 9. Recognition Process

Ⅳ. 실험 결과 및 고찰

제안된 시스템의 성능을 인식하고자 일련의 실험을 수행하였다. 초당 15프레임의 120×120 화소의 2분간의 영상을 가지고 영상모델을 구성하였다. 이중 50%는 손이 포함된 영상이고 나머지는 50%는 배경 영상이다.

4.1 손 추적 실험

각기 다른 10명의 피검자에 대해 각기 5분의 데이터를 가지고 손의 위치 추적 실험을 진행하였다. 그 결과를 그림 10과 표 1에 나타내었다. 손 위치 추적에 있어서는 거의 100% 추적하였지만 일반적인 손의 움직임보다 매우 빠르게 위치가 변하는 경우 간혹 인식하지 못하는 경우도 발생하였다.

표 1. 손 위치 추적 결과
Table 1. Result of hand position tracking

	인식	실패	인식률
X 좌표	4476	24	99.4%
Y 좌표	4487	13	99.7%
평균			99.6%

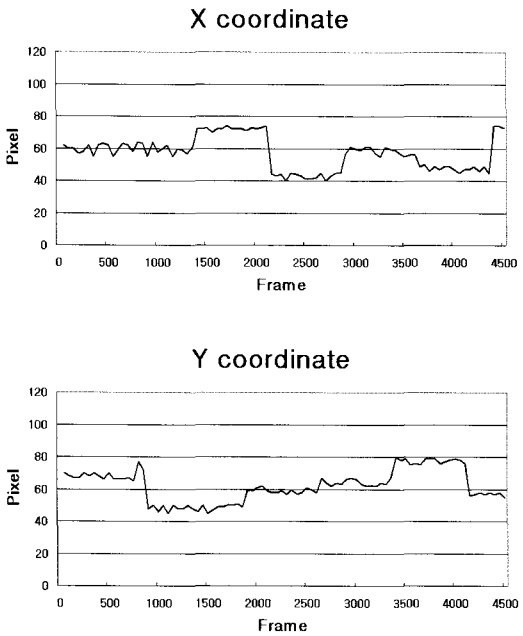


그림 10. 손 위치 추적 결과
Fig. 10. Result of hand position tracking

4.2 동작 인식 실험

객체 조작을 위한 손동작을 정의하기 위하여 그림 11과 같이 일곱 가지 손 형태를 정의하였다. 그리고 표 2와 같이 손 형태가 변화함에 따른 손동작과 객체 제어를 위한 명령어를 정의하였다. 제안된 시스템의 동작 인식을 성능을 입증하기 위해 제안된 방법과 외형을 이용한 방법[2]에 대하여 손동작 인식 성능을 비교하였다.

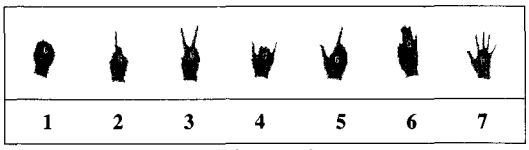


그림 11. 일곱 가지 형태의 손 모양
Fig. 11. Seven hand shapes

표 2. 객체 조작을 위한 명령어 목록
Table 2. Command list for object manipulation

Gesture	Command	Shape Transition
1	Grab & Move	7 → 1
2	Resize	7 → 2
3	Change Color	7 → 3
4	Delete	7 → 4
5	Connect	7 → 5
6	Separate	7 → 6

표 3에 전체적인 비교 결과를 나타내었다. 제안된 방법은 평균 99.28%의 인식률을 나타내어 기존의 외관을 이용한 방법의 평균 인식률인 95.37%보다 높게 나타났으며, 형태 변화가 뚜렷한 동작 1번보다 형태의 변화가 불분명한 6번의 경우가 양쪽 모두 인식률이 낮았다.

표 3. 동작 인식 결과
Table 3. Result of gesture recognition

Gesture	Proposed method (%)	appearance method[2] (%)
1	100	97.6
2	100	96.6
3	99.5	95.2
4	98.7	94.7
5	99.1	94.6
6	98.4	93.5
Avg.	99.28	95.37

V. 결 론

본 논문에서는 가상 공간에서 객체 조작을 위한 비전에 근거한 동작 상호작용 시스템을 제안하였다. 제안한 방법은 비동기 다중 관찰을 사용함으로서 폭넓은 시스템의 확장성을 이룰 수 있었고, 통계적 방법을 이용하여 기하학적 구조의 외관을 이용하여 손을 검출할 수 있었다. 또한, 모델의 수가 적은 경우에도 영상에서 손을 견실하게 인식하였다. 실험 결과 제안된 방법은 기존의 외관을 이용한 방법보

다 3.91% 개선된 99.28%의 인식률을 나타내어 효율성을 입증하였다.

앞으로는 다른 손 자세에 대한 다중 모델구성과 다양한 샘플 영상을 이용한 확장된 실험을 진행할 계획이며, 또한 연산 비용을 줄이고 연산 복잡도에 따른 동적 제어를 증가시켜야 할 것이다.

참 고 문 헌

[1] Ishikawa. M, Matsumura. H, "Recognition of a hand-gesture based on self- organization using a DataGlove.," ICONIP '99. Vol.2, pp.739-745, 1999.

[2] V. I. Pavlovic, R. Sharma, and T. S. Huang. "Visual interpretation of hand gestures for human-computer interaction: A review.," IEEE PAMI, vol.19, no.7, pp.677-695, 1997.

[3] Baback Moghaddam and Alex Pentland. "Maximum likelihood detection of faces and hands." in Proc. of International Workshop on Automatic Face and Gesture Recognition, pp.122-128, 1995.

[4] James M. Rehg and Takeo Kanade. "Visual tracking of high dof articulated structures: an application to human hand tracking." in Computer Vision(ECCV), pp.35-46, 1994.

[5] James Davis and Mubarak Shah. "Determining 3-d hand motion." in Asilomar Conference in Signals, Systems and Computers, pp.1262-1266, 1994.

[6] Tominaga, M.; Hongo, H.; Koshimizu, H.; Niwa, Y.; Yamamoto, K. "Estimation of human motion from multiple cameras for gesture recognition." Pattern Recognition, pp.401-404, Aug. 2002.

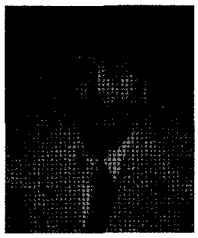
[7] Kumar, S.; Kumar, D.K.; Sharma, A.; McLachlan, N. "Visual hand gestures classification using temporal motion templates and wavelet transforms." Multimedia Modelling Conference, 2004. Jan. 2004.

[8] Huang, T., "Vision-based hand gesture tracking and recognition.," ISSCS 2005. pp.403-404 July, 2005.

박 호 식(Ho-Sik Park) 정회원

1994년 2월 연세대학교 의용전자공학과 졸업(공학사)
2001년 2월 관동대학교 대학원 전자통신공학과 졸업(공학석사)
2005년 2월 관동대학교 대학원 전자통신공학과 졸업(공학박사)
<관심분야> 영상처리, 신호처리시스템, 영상압축

배 철 수(Cheol-soo Bae) 종신회원



1979년 2월 명지대학교 전자공학과 졸업(공학사)
1981년 2월 명지대학교 대학원 전자공학과졸업(공학석사)
1988년 8월 명지대학교 대학원 전자공학과졸업(공학박사)
1999년 3월~2001년 5월 관동대학교공과대학 학장
2001년 6월~2003년 8월 관동대학교 평생교육원장
2003년 1월~현재 한국통신학회 국내저널 편집부위원장
1981년~현재 관동대학교 전자정보통신공학부 교수
<관심분야> 영상처리, 신호처리시스템, 영상압축