

수화 인식을 위한 얼굴과 손 추적 알고리즘

정회원 박 호 식*, 종신회원 배 철 수*

Face and Hand Tracking Algorithm for Sign Language Recognition

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

요 약

본 논문에서는 수화 인식을 위한 얼굴 및 손 추적시스템을 제안한다. 제안된 시스템은 검출 및 추적 단계로 구분된다. 검출 단계에서는 신호의 주체인 얼굴과 손에 위치한 피부 특징을 이용하였다. CbCr 공간에서의 타원 모델을 구성하여 피부 색상을 검출하고 피부 영역을 분할한다. 그리고 크기와 얼굴 특징을 이용하여 얼굴과 손 영역을 정의한다. 추적 단계에서는 동작 추정을 위하여 첫 번째 손 영역으로 예측된 다음의 손위치를 연산함으로써 두 번째 손의 영역을 유도해낸다. 그러나 갑작스런 움직임의 속도 변화가 있을 경우 연속된 프레임에서 추적된 위치는 부정확하였다. 이러한 점을 해결하고자 손 영역에 대하여 반복적인 재연산을 수행하여 적응적으로 영역을 찾음으로써 오차를 보정하도록 하였다. 실험 결과 제안된 방법은 기존의 방법보다 4%의 처리 시간이 증가된 반면, 예측 오차는 96.87%까지 감소 시킬 수 있었다.

Key Words : Hand Tracking, Sign Language, Ellipse Model, Skin-Color Ssgmentation, Motion Estimation

ABSTRACT

In this paper, we develop face and hand tracking for sign language recognition system. The system is divided into two stages; the initial and tracking stages. In initial stage, we use the skin feature to localize face and hands of signer. The ellipse model on CbCr space is constructed and used to detect skin color. After the skin regions have been segmented, face and hand blobs are defined by using size and facial feature with the assumption that the movement of face is less than that of hands in this signing scenario. In tracking stage, the motion estimation is applied only hand blobs, in which first and second derivative are used to compute the position of prediction of hands. We observed that there are errors in the value of tracking position between two consecutive frames in which velocity has changed abruptly. To improve the tracking performance, our proposed algorithm compensates the error of tracking position by using adaptive search area to re-compute the hand blobs. The experimental results indicate that our proposed method is able to decrease the prediction error up to 96.87% with negligible increase in computational complexity of up to 4%.

I. 서 론

오랫동안 인간의 손을 이용한 여러 가지 신호가 상호간의 정보를 교환하기 위한 수단으로 사용되어

왔다. 특히, 청각 장애우들은 손동작을 이용한 수화를 사용하여 대다수의 의사소통을 하고 있다. 그러나 수화 습득의 어려움으로 인하여 일반인과의 의사소통에서 많은 어려움을 겪고 있다. 그러므로 청

* 관동대학교 전자통신공학과 영상처리연구실 (mediana@netsgo.com)

논문번호 : KICS2006-09-386, 접수일자 : 2006년 9월 18일, 최종논문접수일자 : 2006년 10월 17일

각 장애우와 일반인 사이의 원만한 의사소통이 가능 하도록 하는 수화 인식 장치가 필요하다.

수화 인식을 위해 동작 데이터를 취득하는 방법은 크게 2가지로 구분된다. 첫 번째 방법은 기기 기반의 측정 방식으로 손동작에 대한 정확한 위치를 획득 할 수 있는 데이터 글로브와 같은 장비로 손동작을 측정¹⁾한다. 그러나 장비 착용에 따른 불편함과 손의 운동 범위와 같은 여러 가지 제약 조건을 수반하고 얼굴 동작은 획득 할 수 없다. 두 번째 방법은 비전에 근거한 방법^{2,3)}이다. 이 방법은 데이터 글로브 같은 장치의 착용 없이 얼굴과 손동작을 동시에 획득할 수 있다. 이 방법은 카메라를 통한 입력 영상을 사용하는 방법으로 첫 번째 방법 보다 장비가 간단하며 행동변경이 자유롭고 사용자 불편함 없이 자연스러운 동작 인식이 가능하다는 장점이 있다

본 논문에서는 수화 인식을 위한 얼굴과 손 검출 및 추적 시스템을 제안한다. 제안된 시스템은 검출과 추적 단계로 구분할 수 있다. 검출 단계에서는 피부 색상 특징을 추출하여 피부 영역을 분할하고, 크기 및 얼굴 특징을 이용하여 얼굴과 손을 분리해 낸다. 추적 단계에서는 손 추적을 위해 움직임 추정 모델을, 얼굴 추적을 위하여 이전 프레임의 위치를 기반으로 추적하도록 하였다. 추적의 정밀도를 개선하기 위하여 손 영역에 대하여 반복적인 재연산을 수행하여 적응적으로 영역을 찾음으로써 오차를 보정하도록 하였다. 실험 결과 제안된 알고리즘이 얼굴과 손 추적에 있어 높은 정확도를 나타냄을 증명하였다.

II. 얼굴 및 손 검출

검출 단계의 목표는 피부 색상 모델을 이용하여 연속된 수화 영상으로부터 영역을 분할하여 신호의 주체인 얼굴과 손의 위치를 검출하는 것이다.

2.1 피부 색상 분할

본 논문에서는 얼굴 및 손 검출을 위하여 우선적으로 각 화소를 피부 화소와 피부 외 화소로 분류하는 방법을 사용하였다. 피부 색상 분할은 조명 조건에 민감하게 반응함으로 피부 분할을 위해 적당한 색상 공간을 선택하는 것이 매우 중요하다. 색상 공간으로 RGB[4], 정규화 된 RGB[5], HSV[6] 그리고 YCbCr^{17,8)} 등이 있으나, 피부 색상 검출 성능을 고려하여 YCbCr을 사용 하였다.

피부 색상 모델 구성을 위해 수동으로 피부 영역을 잘라내어 CbCr 평면에 나타내었다. 그리고 타원 피부 색상모델을 이용하여 얼굴과 손을 분할하였다. 타원 모델은 식 (1)-(4)에 의해 유도 할 수 있다. 그림 1(c)에 피부 마스크 결과를 나타내었다.

$$\begin{bmatrix} \alpha(Cb, Cr) \\ \beta(Cb, Cr) \end{bmatrix} = \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} Cb \\ Cr \end{bmatrix} \quad (1)$$

여기서 α β 는 변환 영역이고, θ 는 수식 (2)에 의하여 구할 수 있다. 또한 수식 (2)의 μ_{pq} 는 수식 (3)에서 얻을 수 있다.

$$\theta = \frac{1}{2} \tan^{-1} \left(\frac{2\mu_{11}}{\mu_{20} + \mu_{02}} \right) \quad (2)$$

$$\mu_{pq} = \sum_{i=1}^n \sum_{j=1}^n (Cb_i - \bar{Cb})^p (Cr_j - \bar{Cr})^q \quad (3)$$

$$\frac{L_1^2}{a^2} + \frac{L_2^2}{b^2} = 1 \quad (4)$$

여기서 a 는 주축의 b 는 보조축의 타원의 크기이고, L_1 와 L_2 는 각각 타원 주축과 보조축이다.

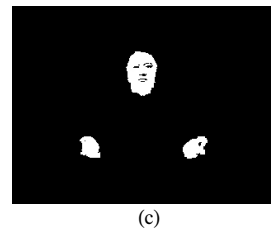
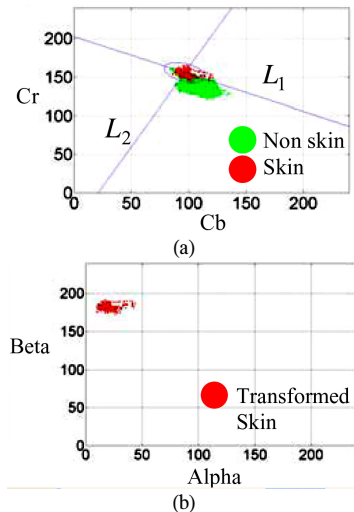


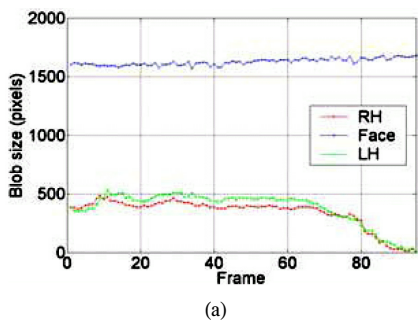
그림 1. (a) 피부-화소 분포 (b) 피부 화소를 α, β 로 변환 (c) 피부 색상 마스크
Fig. 1. (a) Skin pixels distribution (b) Transform skin pixels to α, β (c) Skin color mask

평균 (Cb, Cr) 에서 수직이 되고 L_1 와 L_2 가 교차 되는 지점에서 (a, β) 의 평균 및 표준 편차로 a, b 를 얻을 수 있고, θ 의 탄젠트는 L_1 의 경사도를 사용하였으며, 최적화된 문턱치 값은 특징점의 위치 정보에서 유도할 수가 있다.

2.2 얼굴 및 손 검출

피부 색상을 분할하고 난 후 형태학상의 연산으로 잡음을 제거하였다. 크기와 얼굴 특징을 사용하여 얼굴과 손 영역을 분리한다.

그림 2(a)에 얼굴 과 왼손 및 오른손 영역의 화소 크기 차이를 나타내고 있다. 검출 초기에는 목을 포함한 얼굴 영역은 1,600개의 화소를 포함하지만 양 손 영역은 각각 약 400개의 화소만을 포함하고 있다. 그림 2(a)에서와 같이 얼굴과 손 영역의 크기 차이가 확연히 나타난다. 그러므로 이에 상응하는 문턱치 값을 선택하였다.



(a)



Zero-cross Edge Fill in boundary Only Face (biggest blob) Facial feature

(b)

그림 2. (a) 오른손, 왼손 및 얼굴 영역 크기 (b) 얼굴 특징
Fig. 2. (a) RH(Right Hand), LH(Left Hand), Face Blobs size. (b) Facial features

그러나 사용자가 반팔을 입은 경우와 같이 얼굴과 손의 영역 크기가 유사해질 수 있으므로 얼굴 특징점을 추가로 사용하였다. 그림 2(b)는 얼굴 특징점 검출 과정을 나타내고 있다. 먼저 제로 교차 모서리 검출 방법으로 얼굴 영역을 분리한 후 경계선으로 둘러싸인 부분을 채워 넣었다. 그 중 가장 큰 영역을 얼굴 영역으로 정의 하였다. 그리고 눈, 코, 입과 같은 얼굴 특징들을 다른 영역에 비해 보

다 휘도가 높다는 사실에 근거하여 검출할 수 있다.

그림 3에 검출 단계의 순서도를 나타내었다. RGB 순차 영상이 입력되면 YCbCr 색상 공간으로 변환한다.

피부 영역을 CbCr 의 타원 모델에서 피부 색상 특징을 이용하여 분할한다. 분할된 피부 색상은 팽창 및 침식에 의해 근접하게 되고 잡음 공간을 제거한다. 그리고 특징을 검출하여 손 및 얼굴 영역을 검출한 후 추적을 시작한다.

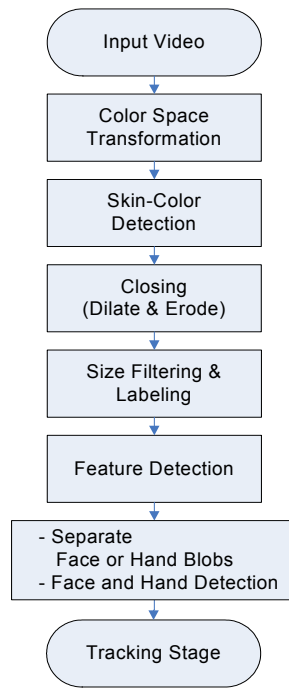


그림 3. 검출 단계의 순서도
Fig. 3. Flowchart of Detection

Ⅲ. 얼굴 및 손 추적

본 논문에서는 새로운 움직임 추정 모델을 제안한다. 검출 단계에서 얼굴과 손의 초기 위치를 얻은 후에 추정 모델의 할당에 의해 다음 프레임의 얼굴과 손의 새로운 위치를 예측한다.

3.1 얼굴 추적

얼굴 영역은 위치의 변화가 거의 없으므로 이전 프레임의 위치로 손쉽게 얼굴 영역을 정할 수가 있다. 식 (5)로 이전 위치에서 검색 영역을 한정 할 수 있다.

$$\begin{bmatrix} x_{search}(i+1) \\ y_{search}(i+1) \end{bmatrix} = \begin{bmatrix} x(i) \pm W_i \\ y(i) \pm W_i \end{bmatrix} \quad (5)$$

여기서 W_i 는 다음 프레임에서 검색할 창의 범위이다. 본 논문에서는 초당 15프레임의 320×240 영상을 사용하여 실험한 결과 프레임간 화소 변화가 최대 8개 화소를 나타내었기에 약간의 여유를 두어 W_i 를 10으로 사용하였다.

3.2 손 추적

손 추적을 위하여 연속된 프레임에서 속도와 가속도를 이용하여 움직임 추정을 하였다. 손의 이전 위치에서 다음의 위치를 예측하였다. 각 영역의 모든 화소 위치는 각 영역의 중심으로 계산한다. 초기 3프레임에서 손 영역의 위치를 보전하여 속도 및 가속도를 연산한다.

식 (6)-(7)로부터 $i-1$ 과 i 프레임 사이의 수평 및 수직 속도 및 가속도를 구한다.

$$\begin{bmatrix} V_x(i) \\ V_y(i) \end{bmatrix} = \begin{bmatrix} x(i) \\ y(i) \end{bmatrix} - \begin{bmatrix} x(i-1) \\ y(i-1) \end{bmatrix} \quad (6)$$

$$\begin{bmatrix} A_x(i) \\ A_y(i) \end{bmatrix} = \begin{bmatrix} V_x(i) \\ V_y(i) \end{bmatrix} - \begin{bmatrix} V_x(i-1) \\ V_y(i-1) \end{bmatrix} \quad (7)$$

식 (8)에 의해 움직임 추정을 한다.

$$\begin{bmatrix} x(i+1) \\ y(i+1) \end{bmatrix} = \begin{bmatrix} x(i) + V_x(i) \\ y(i) + V_y(i) \end{bmatrix} + \frac{1}{2} \begin{bmatrix} A_x(i) \\ A_y(i) \end{bmatrix} \quad (8)$$

본 논문에서는 식 (9)와 같은 움직임 추정 모델을 제안한다. 위치 추정에 있어 기존 식 (8)의 방법이 현재 손의 위치와 속도, 가속도만을 이용하였으

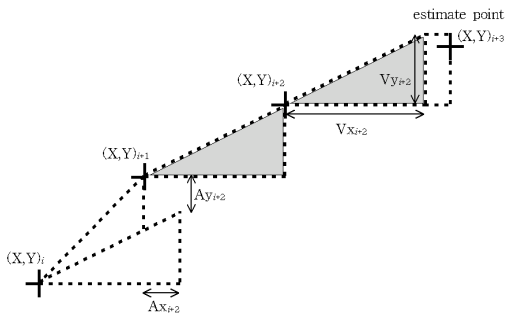


그림 4. 움직임 추정 방법
Fig. 4. The method of motion estimation

나 제안된 모델은 이전 손의 속도 정보 또한 이용함으로써 손의 위치 예측 오차를 줄일 수 있었다. 그림 4에 제안된 방법을 나타내었다.

$$\begin{bmatrix} x(i+1) \\ y(i+1) \end{bmatrix} = \begin{bmatrix} x(i) \\ y(i) \end{bmatrix} + \frac{1}{2} \begin{bmatrix} V_x(i) + V_x(i-1) + A_x(i) \\ V_y(i) + V_y(i-1) + A_y(i) \end{bmatrix} \quad (9)$$

3.3 검색 영역 확장

갑작스런 움직임의 속도 변화가 있을 경우 연속된 프레임에서 추적된 위치는 오차가 발생 함을 관찰할 수 있었다. 그러므로 본 논문에서는 손 영역에 대하여 반복적인 재연산을 수행하여 적응적으로 검색 영역을 확장함으로써 오차를 보정하도록 하였다. 식 (10)과 같이 문턱치 값을 초과하는 가속을 가지는 경우에는 검색 창을 확장하도록 하였다. 그림 5에 추적방법의 순서도를 나타내었다.

$$W_i = \begin{cases} W_{small} & \text{if } A_{x,y} < \varepsilon \\ W_{large} & \text{if } A_{x,y} \geq \varepsilon \end{cases} \quad (10)$$

- W_{small} : 검색을 위한 작은 창
- W_{large} : 검색을 위한 큰 창
- ε : 문턱치 값

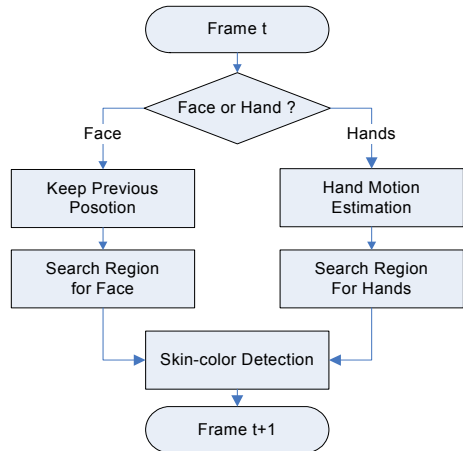


그림 5. 추적 단계의 순서도
Fig. 5. Flowchart of Tracking

IV. 실험 결과 및 고찰

제한된 알고리즘의 효율성을 입증하기 위하여 초당 15 프레임 320×240 화소의 각기 다른 수화 동

영상을 각각 5분씩 획득하였다.

그림 6은 획득한 영상 일부에 대한 손의 프레임 별 속도의 변화와 식(8)에 의한 위치오차 그리고 제안된 방법에 대한 위치오차를 나타내고 있다. 그림 6(a)는 그 중 속도의 변화를 나타낸 것으로 손의 위치가 급격하게 변하는 지점을 원으로 표기하였다. 그리고 그림 6(b)는 식 (8)에 의한 예측 오차를 나타낸 것으로 속도의 변화가 심한 곳에서 오차가 많이 나타남을 확인할 수있다. 그러나 제안된 방법을 사용한 그림 6(c) 경우에는 예측 오차가 현격하게 감소 한 것을 확인 할 수가 있었다. 또한 그림 7에 2개의 실험 영상에서 제안된 방법을 이용하여 추적된 얼굴과 손 영역의 예를 나타내었다.

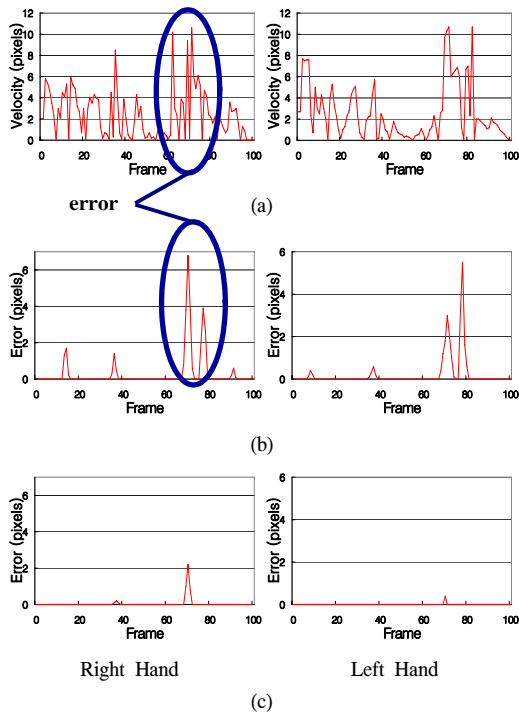


그림 6. 위치 오차의 예 (a) 속도 (b) 수식(8)에 의한 위치 오차 (c) 제안된 방법에 의한 위치 오차
 Fig. 6. Sample of Euclidean error (a) Velocity (b) Position Error eq. (8) (c) Position Error (Proposed)

표 1에 개선된 예측 오차를 나타내었고, 표 2에는 수식(8)의 기존의 추적방법과 제안된 검색 영역 확장 방법을 사이의 예측오차와 처리시간을 비교하여 나타내었다. 실험 결과 제안된 방법은 기존의 방법보다 4%의 처리 시간이 증가된 반면, 예측 오차는 96.87%까지 감소 시킬 수 있었다.

표 1. 두 개의 연속된 수화 영상에 대한 움직임 추정 모델의 성능 비교

Table 1. Comparative performance of motion estimation model in two Thai sign language sequences

Video Sequence	Method 1 Eq. 8	Method 2 (proposed)
	Position error (pixels)	Position error (pixels)
Seq.1	7.06 (RH) 9.97 (LH)	0.80 (RH) 1.80 (LH)
Seq.2	18.37 (RH) 11.50 (LH)	2.47 (RH) 0.36 (LH)

(참고) RH: Right Hand, LH: Left Hand

표 2. 수식(8)의 기존의 추적 방법과 제안된 방법과의 위치 오차와 처리 시간 비교

Table 2. Position error and processing time between normal tracking, eq. (8), versus proposed adaptive window mode

Video	Decrease Position Error (%)	Increase Processing Time (%)	Number of iterative computation
Seq.1	88.67 (RH) 81.95 (LH)	2.45	8
Seq.2	86.55 (RH) 96.87 (LH)	4.05	16

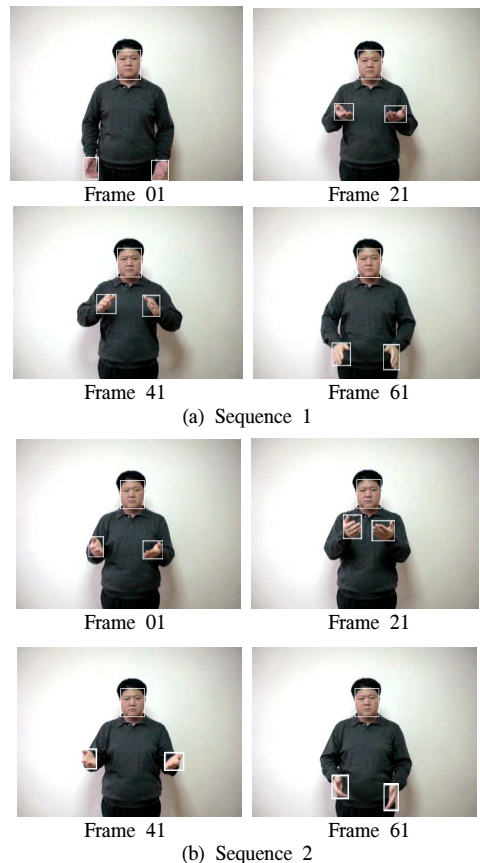


그림 7. 추적 결과의 예
 Fig. 7. Sample of Tracking Result

V. 결론

본 논문에서는 수화 인식을 위한 얼굴 및 손 검출 방법 대하여 제안하였다. 제안된 방법은 피부 색상 특징을 이용하여 크기와 얼굴 특징을 이용하여 얼굴과 손 영역을 분리하여 얼굴과 손영역을 검출하고 연속적으로 위치를 추적하도록 하였다. 추적시 얼굴은 위치의 변화가 적어 손쉽게 다음의 위치를 예측할 수 있었지만, 손은 움직임 추정 모델을 구성하여 다음의 위치를 연산하여 예측하였으나 속도의 변화가 심할 경우 위치 오차가 발생하였다. 그러므로 본 논문에서는 손 영역에 대하여 반복적인 재연산을 수행하여 적응적으로 검색 영역을 확장함으로써 오차를 보정하는 방법을 사용하였다. 실험 결과 제안된 알고리즘은 얼굴 및 손 추적에 있어 기존의 방법보다 4%의 처리 시간이 증가된 반면, 예측 오차는 96.87%까지 감소 시킬 수 있었다.

참 고 문 헌

[1] V. I. Pavlovic, R. Sharma, T. S. Huang, "Visual interpretation of hand gestures for human-computer interaction: a review" IEEE transactions on pattern analysis and machine intelligence, v.19 no.7, pp.677-695, 1997.

[2] D. M. Gavrilu, "The Visual Analysis of Human Movement: A Survey" CVIU, v.73 no.1, pp.82-98, 1999.

[3] A. Shamaie and A. Sutherland, "A dynamic model for real-time tracking of hands in bimanual movements," in 5th International Gesture Workshop, Geneva, April 2003.

[4] M. J. Jones and J. M. Regh, "Statistical color models with application to skin detection," in Proc. Computer Vision and Pattern Recognition, pp.274-280, June 1999.

[5] J. Fritsch, S. Lang, M. Kleinhagenbrock, G. A. Fink and G. Sagerer, "Improving Adaptive Skin Color Segmentation by Incorporating Results from Face Detection," IEEE Int. Workshop on Robot and Human Interactive Communication, September 2002.

[6] N. Tanibata, N. Shimada and Y. Shirai, "Extraction of Hand Features for Recognition of Sign Language Words," in Proc. of Int. Conf on Vision Interface, pp.391-398, 2002.

[7] N. Soontranon, S. Aramvith and T.H. Chalidabhongse, "Face and Hand Localization and Tracking for Sign Language Recognition," in Proc. International Symposium on Communication and Information Technologies (ISCIT'04), October 2004.

[8] R. L. Hsu, M. Abdel-Mottaleb and A. K. Jain, "Face Detection in Color Images," IEEE Trans. on Pattern Analysis and Machine Intelligence, vol.2, No.5, pp.696-706, May 2002.

박 호 식 (Ho-sik Park)

정회원



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

시스템, 영상압축

배 철 수 (Cheol-soo Bae)

종신회원



1979년 2월 명지대학교 전자공학과(공학사)
1981년 2월 명지대학교 대학원 전자공학과(공학석사)
1988년 8월 명지대학교 대학원 전자공학과(공학박사)
1999년 3월~2001년 5월 관동대학교공과대학 학장

2000년 3월~2002년 2월 관동대학교 양양캠퍼스 창업보육센터 소장
2001년 6월~2003년 8월 관동대학교 평생교육원장
2003년 1월~현재 한국통신학회 국내저널 편집부위원장
1981년~현재 관동대학교 전자정보통신공학부 교수
<관심분야> 영상처리, 신호처리시스템, 영상압축