

제스처 기반 인터페이스를 위한 손영역 획득 시스템

The Hand Region Acquisition System for Gesture-based Interface

양선옥 · 고일주 · 최형일

Seon-Ok Yang, Il-Ju Ko and Hyung-Il Choi

*숭실대학교 컴퓨터학부

요 약

인간의 시각이 물체를 구분할 수 있는 가장 큰 특징 중의 하나가 색상 정보이므로 본 논문에서는 칼라 정보를 이용하여 손영역을 획득한다. 카메라를 통하여 입력된 영상의 칼라 값은 동일한 물체에 대해서도 조명의 강도와 방향에 따라 변하기 때문에 사전 지식이 없이는 정확한 손영역을 획득하기가 어렵다. 따라서 본 논문에서는 손영역을 획득하기 전에 학습단계를 통하여 손색상 모델을 생성한다. 뿐만 아니라 사용자의 움직임에 따른 칼라 값의 변화를 손색상 모델에 반영하기 위하여 칼만필터를 이용한다. 이 때, 칼만필터는 다음 시점의 손영역을 예측하여 탐색 공간을 최소화 하므로 시스템의 처리 속도를 향상시킨다. 실험에서는 손영역 획득 모듈을 구현함으로써 제안된 방법의 타당성을 검증하였다.

ABSTRACT

We extract a hand region by using color information, which is an important feature for human vision to distinguish objects. Because pixel values in images are changed according to the luminance and lighting source, it is difficult to extract a hand region exactly without previous knowledge. We generate a hand skin model at learning stage, and extract a hand region from images by using the model. We also use a Kalman filter to consider changes of pixel values in a hand skin model. A Kalman filter restricts a search area for extracting a hand region at next frame also. The validity of the proposed method is proved by implementing the hand-region acquisition module.

1. 서 론

인간이 의사를 전달하기 위하여 가장 많이 사용하는 수단은 음성이다. 그러나 인간의 제스처에 대한 이해 능력과 직관력으로 인하여, 우리 인간은 대화할 때에 보조적인 의사 전달 수단으로 제스처를 많이 이용하게 되며, 때에 따라서는 제스처만으로 의사를 전달하기도 한다. 즉, 사람들의 관심을 어떤 물체로 집중시키고자 할 때의 가리키는 동작, 친구에게 인사할 때의 손 흔드는 동작, 개수를 표시할 때의 손가락을 이용하는 동작 등이 이러한 예가 될 수 있다. 이러한 특성을 사용자와 컴퓨터간의 의사소통에도 적용시키고자 하는 연구가 최근 많이 수행되어 왔다[1,2].

제스처를 인식하기 위해서는 먼저 3차원 공간에서 사용자의 제스처를 입력받아야 하는데, 어떻게 입력을 받아들이어야 하는가에 대한 문제가 중요한 연구분

야로 알려져 있다. 현재까지 널리 알려진 방법으로는 전자 장갑을 입력 장치로 하는 방법과 컴퓨터 시각 시스템을 입력 장치로 하는 방법이 있다. 전자 장갑은 정확한 손의 움직임을 입력받을 수 있기 때문에 가상 현실 시스템에 많이 사용되지만, 전자 장갑을 착용하여야만 한다는 부담감은 자연스럽고 편안한 인터페이스를 원하는 사용자의 요구를 만족시킬 수 없었다. 컴퓨터 시각 시스템은 사용자가 다른 부가적인 장치를 착용하지 않도록 하기 때문에 보다 자연스러운 인터페이스를 제공할 수 있다. 그러나 컴퓨터 시각 인식 기술의 한계로 인하여 아직은 만족할만한 수준으로 제스처를 인식하지 못하고 있다.

본 논문은 이러한 컴퓨터 시각 인식 기술의 한계를 극복하여 사용자의 제스처를 인식함으로써 현재의 키보드나 마우스와 같은 입력 장치보다 편리한 제스처 기반 인터페이스의 구현을 목표로 하고 있다. 그림

* 본 연구는 1997년도 숭실대학교 학술연구비에 의하여 연구되었음.

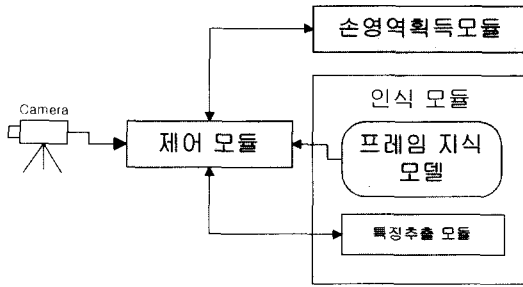


그림 1. 시스템 구조.

1은 본 논문에서 제안하고자 하는 시스템에 대한 개략적인 구조를 보인다. 제안하는 시스템은 영상 특징에 기반한 컴퓨터 시각 인식 기법을 이용하여 복잡한 배경 영상으로부터 손영역을 획득하고, 다음 단계에서는 획득된 손영역에서 특징을 추출하여 프레임 지식과 정합을 수행하고 적절한 스카마를 활성화하여 제스처를 인식하는 과정으로 이루어진다.

그림 1에서와 같이 시스템은 크게 제어모듈과 손영역 획득 모듈, 제스처 인식 모듈로 구성된다. 본 논문에서는 손영역 획득할 때의 문제점을 기술하고, 이에 대한 해결 방안을 모색하기 위한 기법을 제안한다. 손영역을 획득하는 단계는 영역 분리 모듈, 초기 손영역 추출 모듈, 손영역 추적 모듈, 손영역 획득 모듈로 이루어진다.

컴퓨터 시각에 의한 영상을 처리하기 위해서 본 논문에서는 YIQ 칼라공간을 사용하는 손색상 모델과 손색상 결정함수를 정의하고, 칼라영상의 이진화 기법을 제안한다. 그리고 칼만필터를 이용하여 연속한 프레임에서의 손영역을 예측하고, 손색상 모델을 갱신한다. 따라서 손색상 모델은 상황에 적응적인 특징을 갖게 되며, 손영역의 예측은 탐색 공간을 최소화하므로 시스템의 처리 속도를 향상시킨다.

본 논문은 다음과 같이 구성된다. 제 II 장에서는 칼라영상의 이진화를 위한 손색상 모델에 대한 정의와 학습 방법을 기술하고, 제 III 장에서는 배경으로부터 물체 영역을 분리하는 과정인 이진화 및 레이블링 기법과 칼만필터를 이용한 손색상 모델의 갱신과 예측을 통한 탐색 영역의 축소, 그리고 영역 추적에 의한 손영역의 획득에 대하여 기술한다. 마지막으로 제 IV 장에서는 실험 및 결론에 대하여 기술한다.

2. 손색상 모델의 학습

2.1 칼라공간

인간의 시각이 물체를 구분할 수 있는 가장 큰 특징 요소 중의 하나가 색상이므로 본 논문에서는 칼라정

보를 이용하여 손영역을 획득한다. 일반적으로 칼라 영상을 코딩하기 위해서 RGB 칼라공간을 많이 사용하지만, 영상처리를 수행하려 할 때 세 가지의 RGB 칼라 값을 모두 처리해야 하기 때문에 RGB 칼라공간을 사용하여 영상처리를 수행하는 것은 비효율적이다. 그리고 RGB 칼라공간은 영상의 밝기 변화에 민감하게 작용할 뿐 아니라 동일한 영상의 동일한 영역에 대하여서도 조명의 크기나 방향에 따라 칼라 값이 차이가 크게 발생할 수 있다. 이러한 이유 때문에 많은 이들은 영상처리에 다른 칼라공간을 적용한다.

인간의 시각 시스템은 칼라를 구분하기 위해 밝기(Luminance), 색상(Hue), 채도(Saturation)의 세 가지 특징을 사용한다. 전통적으로 칼라영상의 처리를 위해 이 세 가지 특징을 잘 표현하는 LHS 칼라공간이나 YIQ 칼라공간을 많이 사용한다. 그 중에서도 특히 YIQ 칼라공간은 밝기의 변화에 대하여 색상이나 채도가 변하지 않고 동양인의 손색상을 구분하기에 적합한 방법이다. 따라서 칼라영상을 NTSC(National Television System Committee)에서 정한 합성 칼라 비디오 표준 방식인 YIQ 칼라공간으로 변환하여 영상 처리에 이용한다[3,4].

YIQ 칼라공간은 YUV 칼라공간에 의해 유도된 것으로서 Y(Luminance)요소는 밝기 값을 나타내고, I(InPhase)요소와 Q(Quadrature)요소는 두 요소를 합성하여 색상과 채도를 나타낸다. YIQ 칼라공간에서 색상과 채도는 각각의 밝기 수준에서 위상(phase)과 진폭(amplitude)으로 표현되며, 다음의 식 (1)과 식 (2)와 같이 정의한다.

$$Hue = \tan^{-1} \left(\frac{Q}{I} \right) \quad (1)$$

$$Saturation = \sqrt{I^2 + Q^2} \quad (2)$$

NTSC는 식 (3)과 같이 Y(Luminance)요소와 C(Chrominance)요소로 구성되어 있고, C요소는 식 (4)와 같이 Q요소와 I요소로 구성되어 있다. 이 두 식을 통해서 YIQ 칼라공간은 R-Y와 B-Y축에서 33도 회전시킴으로써 YUV 칼라공간보다 색상을 더 정확히 구분할 수 있다.

$$NTSC = Y (luminance) + C (chrominance) \quad (3)$$

$$C = Q \sin(\omega + 33^\circ) + I \cos(\omega + 33^\circ) \quad (4)$$

YIQ 칼라공간은 RGB 칼라공간보다 손색상을 갖는 영역에 보다 민감하게 반응을 한다. YIQ 칼라공간은 흰색과 검은색의 축을 표현하는 Y 요소, 적색과 청록색의 축을 표현하는 I 요소, 그리고 초록색과 자홍

색의 축을 표현하는 Q 요소로 구성되어 있다. 그림 4는 YIQ 칼라공간의 축을 시각적으로 보인다.

식 (5)는 RGB 칼라공간을 YIQ 칼라공간으로 변환하는 식을 보인다[5].

$$\begin{bmatrix} Y \\ I \\ Q \end{bmatrix} = \begin{bmatrix} 0.30 & 0.59 & 0.11 \\ 0.60 & -0.27 & -0.32 \\ 21 & -0.52 & 0.31 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (5)$$

2.2 손색상 모델

손색상 모델은 사용자와의 상호작용에 의한 학습을 통해 획득된 I 요소와 Q 요소의 평균과 표준편차를 이용하여 정의한다. 식 (6)은 손색상 모델을 보이고, 식 (7)과 식 (8)은 평균과 표준편차를 구하는 식을 보인다.

$$C = [\mu_I \ \sigma_I \ \mu_Q \ \sigma_Q] \quad (6)$$

$$\mu_I = \frac{\sum_{x=0}^M \sum_{y=0}^N f_I(x, y)}{M \times N}, \mu_Q = \frac{\sum_{x=0}^M \sum_{y=0}^N f_Q(x, y)}{M \times N} \quad (7)$$

$$\sigma_I = \sqrt{\frac{\sum_{x=0}^M \sum_{y=0}^N (f_I(x, y) - \mu_I)^2}{M \times N}},$$

$$\sigma_Q = \sqrt{\frac{\sum_{x=0}^M \sum_{y=0}^N (f_Q(x, y) - \mu_Q)^2}{M \times N}} \quad (8)$$

손색상 모델은 칼라영상에서 손색상을 추출하기 위해 정의한다. 학습된 I 요소와 Q 요소의 평균과 표준편차가 정규분포를 갖는다고 가정할 때, 화소의 I 요소 값과 Q 요소의 값이 $\mu + 3\sigma$ 에서 $\mu - 3\sigma$ 까지의 구간에 속하는 것을 판별하여 손색상에 해당하는 화소를 획득한다.

손색상 결정함수는 식 (9)와 같이 손색상 모델을 이용하여 칼라영상을 이진화한다. 이 함수는 Mahalanobis 거리를 구하는 식을 이용하여 정의한다. 평균을 중심으로 3σ 만큼의 범위에 속하는 것을 판별하여 손색상을 결정한다.

$$D(X) = \sqrt{[M - X]^T \Sigma^{-1} [M - X]} \quad (9)$$

where, $M = \begin{bmatrix} \mu_I \\ \mu_Q \end{bmatrix}$

$$X = \begin{bmatrix} f_I(x, y) \\ f_Q(x, y) \end{bmatrix}$$

$$\Sigma^{-1} = \begin{bmatrix} 1/\sigma_I^2 & 0 \\ 0 & 1/\sigma_Q^2 \end{bmatrix}$$

$$\begin{cases} D(X) \leq 3 & X \in C \\ otherwise & X \notin C \end{cases}$$

M 은 손색상 모델의 I 요소와 Q 요소의 평균을 말하며 X 는 입력화소의 값을 나타내고 Σ^{-1} 은 손색상 함수의 표준편차의 제곱으로 구성된 공분산 행렬의 역행렬이다. $D(X)$ 의 값이 3보다 작은 경우에는 입력 화소가 손색상을 갖는 것으로 판정하고 그렇지 않은 경우에는 손색상을 갖지 않는 것으로 판정한다. 손색상 결정함수의 임계값이 3인 것은 $\mu \pm 3\sigma$ 의 거리를 Mahalanobis 거리를 이용하여 거리에 대한 정규화를 수행함으로써 얻을 수 있다.

2.3 사용자와의 상호작용에 의한 학습

손색상 모델의 학습은 환경에 적응적인 시스템을 구축하기 위해 필요하다. 조명과 배경의 변화에 따른 사전 지식이 주어지지 않은 영상에 대하여 영역 분리를 정확하게 수행한다는 것은 어려운 문제이다. 또한 같은 의미를 갖는 제스처라 해도 사용자에게 따라서 그 움직임의 특징이 조금씩 다를 수 있다. 이러한 입력 영상들은 결과적으로 고정된 인수에 의한 제스처의 인식 성능을 저하시키는 원인이 된다. 이러한 문제를 해결하기 위해 시스템은 변하는 조명 및 배경으로부터 제스처의 영역을 분할하기 위한 정보를 사전에 학습할 수 있는 기능을 갖추어야 한다[6,7].

사용자와의 상호작용에 의한 학습 방법은 시스템이 제시하는 손모양의 템플릿에 사용자가 자신의 손을 올려놓으면 시스템이 연속한 영상을 획득하여 분석한 후 손영역을 검출하여 손색상 모델을 학습하는 방법이다. 학습 단계는 준비 단계, 진입 단계, 검증 단계, 완료 단계의 4단계로 되어 있다. 첫 번째 단계인 준비 단계에는 시스템이 사용자에게 손모양의 템플릿을 제공하여 시스템을 초기화한다. 진입 단계에서는 사용자가 템플릿 위에 손을 진입시키는 시도를 하는 단계이다. 그리고 검증 단계에서는 사용자가 템플릿 위에 손을 고정시키고 시스템이 템플릿을 분석하여 손을 확인하는 단계이다. 마지막으로 완료 단계는 템플릿으로부터 손색상 모델인 평균과 표준편차를 획득하는 단계이다. 그림 2는 각 단계별로 진행 과정을 보인다.

검증 단계에서 템플릿 위에 손이 있는 것을 판별하기 위한 방법은 준비 단계에서 획득된 템플릿 내부의 배경과 현 시점의 템플릿 내부의 영역에 대한 유사성과 현 시점과 전 시점의 템플릿 내부의 영역에 대한 유사성을 측정하여 수행된다. 두 영역의 유사성은 일반적으로 많이 알려져 있는 유사성 측정 방법인 형판정합(template matching) 기법을 이용하여 계산한다.



그림 2. 상호작용에 의한 손색상 모델의 학습.

획득된 유사성은 적절한 임계값을 적용하여 학습 모델 추출 조건에 만족하는 가를 결정한다.

학습 모델 추출 조건은 현시점의 영역을 초기 시점의 영역과 비교하는 추출 조건 (1)과 전시점의 영역과 비교하는 추출 조건 (2)가 있다. 학습 모듈은 추출 조건 (1)의 유사성이 높고 추출 조건 (2)의 유사성이 낮은 경우가 일정시간 동안 지속이 되면 검증을 완료하고 완료 단계를 수행한다. 완료 단계에서는 템플릿 내부에 있는 화소들을 대상으로 I 요소와 Q 요소의 평균과 표준편차를 구하여 손색상 모델을 작성한다.

3. 영역 추적에 의한 손영역 획득

3.1 손영역 레이블링

본 논문에서는 칼라영상을 이진화하여 배경으로부터 손색상을 갖는 화소를 분리한다. RGB 칼라공간을 사용하는 입력 영상을 YIQ 칼라공간으로 변화하여 I 요소와 Q 요소의 값을 계산하고 2.2절에서 정의한 손색상 모델과 손색상 결정함수를 사용하여 이진화를 수행한다. 이진화된 영상은 아직 영역이 분리된 상태가 아니라 단지 각각의 화소들이 배경과 구분되어 있는 상태이다. 이런 물체 부분에 해당하는 각각의 화소들을 하나의 영역으로 묶어 각 영역에 레이블 값을 할당하는 과정을 레이블링이라 한다. 레이블링 방법은 영상의 첫 번째 라인부터 마지막 라인까지 스캔하면서 레이블을 할당하는 래스터 스캔(raster scan) 방법과 레이블이 할당된 초기 화소로부터 레이블을 전파하는

방법으로 나눌 수 있다. 그러나 후자의 방법은 병렬 처리가 지원되지 않는 컴퓨터에서는 수행 시간이 매우 길어지게 되므로 주로 전자의 방법을 사용한다[8].

본 논문에서는 기존의 2-패스 레이블링 알고리즘보다 수행 속도가 빠르고 기억 장소의 낭비가 적은 경계 추적에 의한 1-패스 레이블링 알고리즘을 사용한다. 이 알고리즘은 레이블링 윈도우를 사용하여 영상을 스캔하면서 레이블을 할당한다. 레이블링 윈도우는 2×2 윈도우로써 이미 스캔한 다른 화소들의 레이블 값을 확인하여 적절한 레이블을 할당하는 역할을 수행한다. 레이블링 윈도우는 영상의 좌에서 우로 위에서 아래로 스캔을 한다. 레이블링 윈도우를 사용하여 레이블을 할당하기 위한 두 가지 조건이 있다. 이런 경우에는 새로운 레이블을 레이블링 윈도우의 기준 화소에 할당한다. 두 번째 조건은 레이블링 윈도우의 다른 화소에 이미 레이블이 할당되어 있는 경우이다. 이런 경우에는 이미 할당되어 있는 레이블을 기준 화소에 할당한다.

그림 3은 경계 추적 레이블링 알고리즘의 적용예를 보인다. S 는 외부 경계 시작점을 S_i 는 내부 경계 시작점을 나타낸다. S 에서 레이블을 할당하는 첫 번째 조건을 만족하기 때문에 새로운 레이블을 할당하고 S 를 시작점으로 하여 외부 경계 추적을 시작한다. 외부 경계 추적은 외부 경계 시작점과 같은 레이블을 할당한다. 추적 알고리즘이 다시 시작점으로 돌아오면 추적을 정상적으로 종료하고 추적 알고리즘을 적용하기 전의 방향으로 스캔을 다시 시작한다.

그러나 그림 3의 예와 같이 영역 안에 구멍이 존재하는 경우에, 외부 경계 추적은 외부 경계에 이미 시작점과 다른 레이블이 할당되어 있어서 추적 알고리즘이 시작점으로 돌아오지 못하고 다른 레이블과 충돌하게 된다. 이 화소를 내부 경계의 시작점인 S_i 로 설정하고 외부 경계 추적 방향과 반대 방향으로 내부

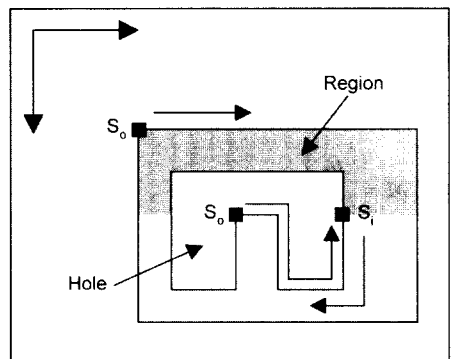


그림 3. 경계 추적 레이블링 알고리즘의 적용 예.

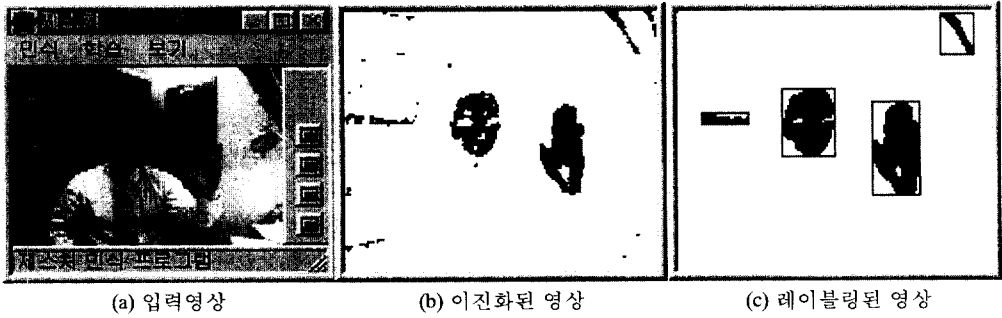


그림 4. 레이블링의 예.

경계 추적을 시작하며 시작점 S_i 와 같은 레이블을 할당한다. 내부 경계 추적이 다시 S_i 로 돌아오면 추적을 종료하고, 내부 경계 추적을 하기 전 단계인 외부 경계 시작점에서 원래의 진행 방향으로 다시 스캔을 시작한다. 이러한 단계를 반복하여 수행함으로써 레이블링 작업을 수행한다.

그림 4는 이진화와 레이블링의 예를 보인다. 그림 4(c)는 하나의 레이블을 사각형으로 표시하여 레이블링된 결과를 보인다.

3.2 손영역 추적

전 시점으로부터의 손영역 정보를 이용하여 현 시점의 손영역을 추적하기 위해 초기 손영역을 추출하여야 한다. 본 논문에서는 시작영역을 설정하고 사용자와의 상호작용을 통하여 시작영역의 범위 내에서 초기 손영역을 추출한다. 시작영역을 설정함으로써 시스템은 시작영역 내에서 찾아지는 영역이 두 가지의 초기 손영역 추출 조건을 만족하는가 확인하고 만족한다면 초기 손영역으로 인식한다. 초기 손영역 추출 조건은 2절에서 기술한 학습 모델 추출 조건과 같다. 인식 준비 단계에서 획득된 시작영역의 영상과 현 시점의 시작영역의 영상의 유사성을 비교하고 현 시점과 전 시점의 시작영역내의 영상의 유사성을 비교하여, 전자의 유사성이 낮고 후자의 유사성이 높은 경우에 획득된 현 시점의 손영역을 초기 손영역으로 추출한다. 그림 5는 응용 시스템에 설정된 시작영역과 초기 손영역을 추출하는 예이다.

칼만필터는 이동 물체의 움직임에 대한 정보가 칼만 필터의 상태 인수들에 누적되어 저장될 수 있기 때문에 이동 물체의 추적에 많이 이용된다[12]. 본 논문에서는 칼만필터를 손색상 모델의 갱신과 영역 분리를 위한 탐색영역의 제한에 사용한다. 칼만필터는 시스템의 상태의 최적 예측인 선형 최소 오차(LMV: Linear Minimum Variance of error) 예측을 위한 순차적이면서 재귀

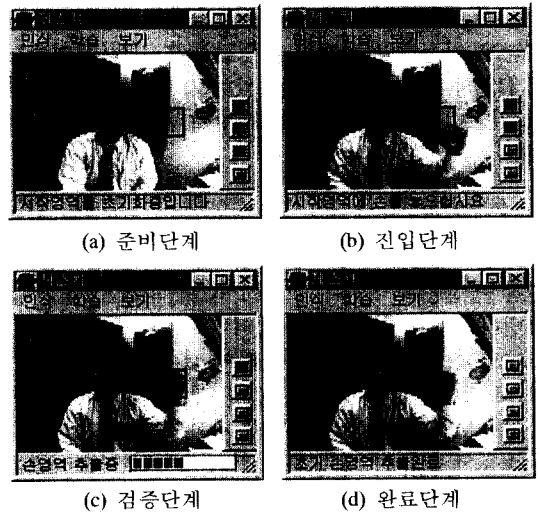


그림 5. 초기 영역의 추출 예.

적인 알고리즘을 제공한다. 먼저 상태 모델이 선형이라고 가정하고 식 (10)과 같이 정의한다[13].

$$x(t) = \Phi(\Delta t)x(t - \Delta t) + w(t - \Delta t) \quad (10)$$

여기서 $x(t)$ 는 시점 t 에서의 시스템 상태를 나타내고 $\Phi(\Delta t)$ 는 상태 전이 행렬을 나타낸다.

본 논문에서는 시스템의 상태를 16차원의 벡터로 표현하고 단위 시간 동안의 손영역의 위치 변화와 크기 변화, 그리고 손색상 모델의 변화를 표현한다. 즉, 식 (10)의 Δx 와 Δy 는 각각 손영역의 x 축과 y 축 상의 중점의 변위를 나타내고 $\mu_h, \sigma_h, \mu_b, \sigma_b$ 는 손색상모델을 나타낸다. 여기서 $\Delta x, \Delta y, x_s, y_s, \mu_h, \sigma_h, \mu_b, \sigma_b$ 는 각각 $\Delta x, \Delta y, x_s, y_s, \mu_h, \sigma_h, \mu_b, \sigma_b$ 의 시간 t 에 대한 변화율을 나타낸다. 손영역의 궤적은 등가속도 운동하고 물체의 크기와 손색상 모델은 선형적으로 변화한다고 가정할 때 상태 전이 행렬은 식 (11)과 같이 정의된다.

$$x(t) = \alpha(\Delta x) = \begin{bmatrix} \Delta x(t) \\ \Delta y(t) \\ xs(t) \\ ys(t) \\ \mu_x(t) \\ \sigma_x(t) \\ \mu_y(t) \\ \sigma_y(t) \\ \Delta x(t) \\ \Delta y(t) \\ xs(t) \\ ys(t) \\ \mu_x(t) \\ \sigma_x(t) \\ \mu_y(t) \\ \sigma_y(t) \end{bmatrix} \alpha(\Delta x) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & \Delta x & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & \Delta y & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & \Delta x & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & \Delta y & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & \Delta x & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & \Delta y & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \Delta x & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \Delta y \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix} \quad (11)$$

그리고 시스템의 상태와 측정값들의 집합 사이에는 식 (12)와 같이 선형 관계가 있다고 가정한다. 칼만필터 알고리즘은 측정값들의 집합에 기반하여 시스템의 상태를 예측한다.

$$y(t) = H(t)x(t) + v(t) \quad (12)$$

여기서 $y(t)$ 는 측정값들의 집합을 나타내고, $H(t)$ 는 관찰 행렬을 나타낸다. 그리고 $v(t)$ 는 측정 오류를 나타낸다. 각 시점에서의 손영역의 위치 변화와 크기, 그리고 손색상 모델을 측정하여 $y(t)$ 의 값들을 얻는다. 시스템 모델과 측정 모델이 정의되면, 동작 인수들의 LMV 예측을 얻기 위해 재귀적 칼만필터 알고리즘을 적용할 수 있다. 재귀적 칼만필터 알고리즘은 초기화, 상태 예측, 측정 갱신의 세 단계의 작업으로 구성된다[14].

초기화 단계에서는 초기 상태 예측 $\hat{x}(0)$, 실제 초기 상태인 $x(0)$ 와 $\hat{x}(0)$ 의 오차를 나타내는 초기 오류 공분산 행렬인 $P(0)$, 예측 오류에 대한 상관 관계 행렬 $Q(t)=E(w(t)w(t)T)$, 측정 오류에 대한 상관 관계 행렬 $R(t)=E(v(t)v(t)T)$ 등을 결정한다.

3.3 손영역의 획득

칼만필터 알고리즘은 손영역의 예측된 중심의 변위와 크기를 제공한다. 이들 예측된 중심의 변위인 $\Delta(t)$ 와 $\Delta(t)$ 는 중심을 구하는데 사용되고, 예측된 크기 $xs(t)$ 와 $ys(t)$ 를 이용하여 탐색영역을 설정한다. 탐색영역의 중심은 예측된 중심으로 하고 길이와 너비는 예측된 크기 $xs(t)$ 와 $ys(t)$ 의 1.5배로 설정한다. 손영역 분리는 탐색영역에서만 수행되어 후보 영역들을 추출한다. 전시점에서 획득된 손영역 정보를 이용하여 전 시점의 손영역과 현 시점의 영역들 사이의 유사도를 계산함으로써 후보 영역들 중의 하나를 손영역으로 결정한다. 즉, 전 시점의 손영역을 추적하여 현 시

점의 손영역을 결정한다.

유사도(Likelihood)는 유사성과 이동성의 가중치의 합으로써, 유사성은 전 시점의 손영역과 현 시점의 영역이 겹치는 정도를 나타내고 이동성은 현 시점의 영역의 이동 정도를 나타낸다. 유사성과 이동성에 대한 전제 조건으로는 손의 움직임이 너무 빠르지 않고 손의 형태가 빠르게 변하지 않는다는 조건, 그리고 손영역은 지속적으로 움직인다는 조건이 있다. 두 가지의 특징은 서로 상반되는 개념으로서 유사성이 높아지면 이동성이 낮아지고 이동성이 높아지면 유사성이 낮아지는 특성을 갖고 있지만 두 특징 값에 가중치의 합을 함으로써 서로의 단점을 보완하는 역할을 수행할 수 있게 된다. 유사성과 이동성에 대한 정의는 식 (13)과 같다.

유사성은 전 시점의 손영역 $R_h(t-1)$ 를 참조하여 계산한다. 후보 영역 $R_i(t)$ 가 전 시점에서 획득된 손영역 $R_h(t-1)$ 와 완전히 겹쳐지면 1의 값을 갖고 $R_i(t)$ 가 $R_h(t-1)$ 과 전혀 겹쳐지는 부분이 없다면 0의 값을 갖는다. 반면에, 이동성은 전 시점의 모든 후보영역들을 참조하며 계산한다. $R_i(t)$ 가 전 시점으로부터 완전히 움직였으면 1의 값을 갖고, 전혀 움직이지 않았으면 0의 값을 갖는다. 유사도는 유사성과 이동성에 적당한 α 와 β 를 곱하여 더함으로써 계산한다. 현 시점의 모든 영역들에 대하여 유사도를 계산하여 그 값들 중에서 가장 큰 값을 갖는 영역을 현 시점에서 획득된 손영역으로 결정한다. 유사도는 식 (14)와 같다.

$$Simil(R_i(t), R_h(t-1)) = 2 \times \frac{Area(R_h(t-1) \cap R_i(t))}{Area(R_h(t-1) + Area(R_i(t)))} \quad (13)$$

$$Mobil(R_i(t)) = \frac{Area(R_i(t)) - Area(R_i(t) \cap (\cup_j R_j(t-1)))}{Area(R_i(t))}$$

$$Likelihood(R_i(t)) = \alpha \times Simil(R_i(t), R_h(t-1)) + \beta \times Mobil(R_i(t)) \quad (14)$$

where, $\alpha, \beta \geq 1, \alpha > \beta$

손영역 획득에 실패하는 경우는 사용자의 손이 카메라의 시야 밖으로 이동하는 경우와 손의 이동 속도가 빨라서 예측된 탐색영역을 벗어나는 경우가 있다. 첫 번째 경우는 손이 카메라의 시야를 벗어난 지점에서 일정 시간동안 손이 재 진입하기를 기다린다. 그러나 일정시간이 경과하여도 손이 재 진입하지 않는다면 손영역 획득 모듈은 초기 상태로 돌아가서 새로운 초기 손영역을 추출하기 위해 사용자에게 시작 영역을 제시한다.

두 번째의 경우에는 입력 영상 전체를 탐색영역으로 설정하여 손영역 추출을 시도한다. 손이 빠르게 움

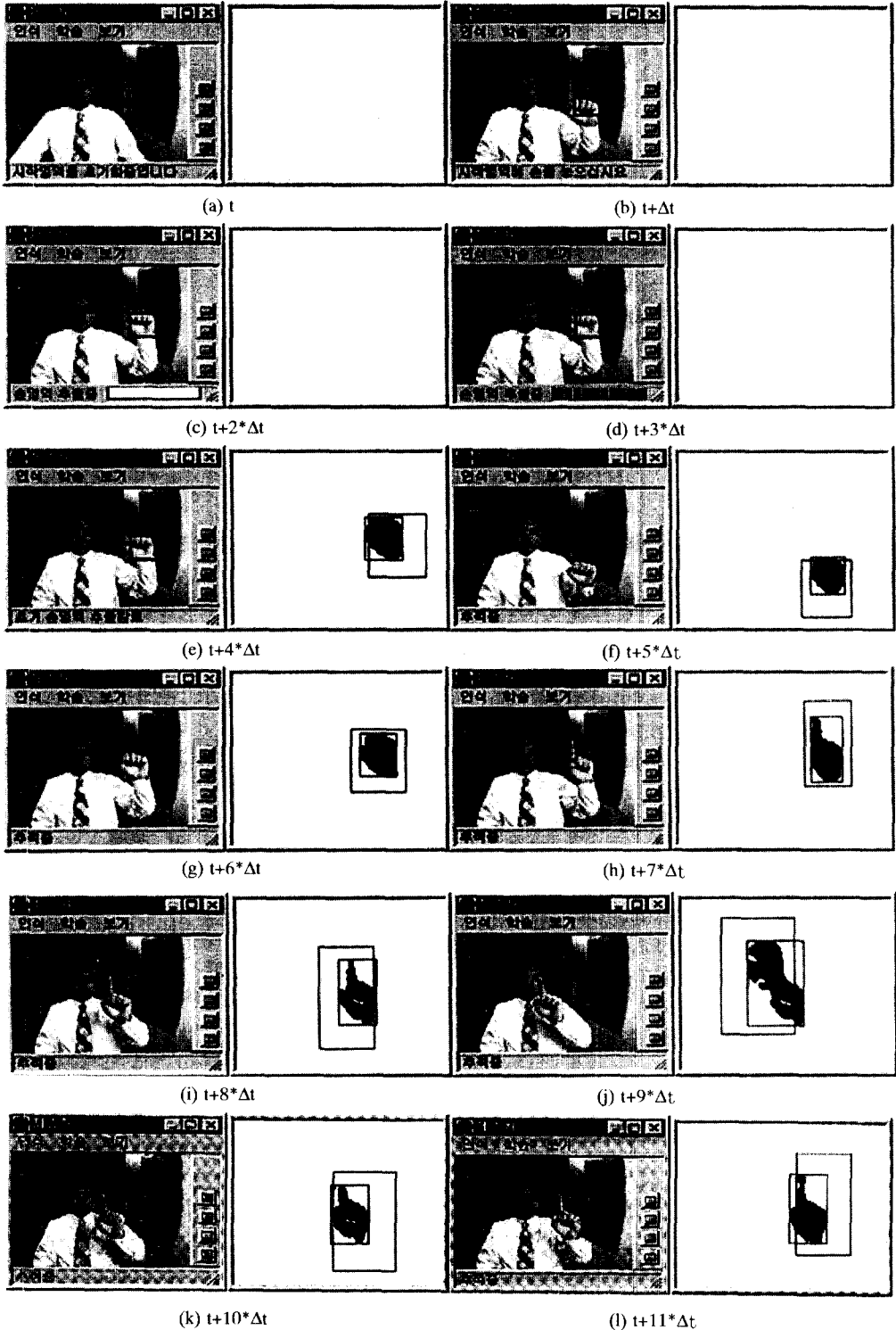


그림 6. 손영역 획득 결과.

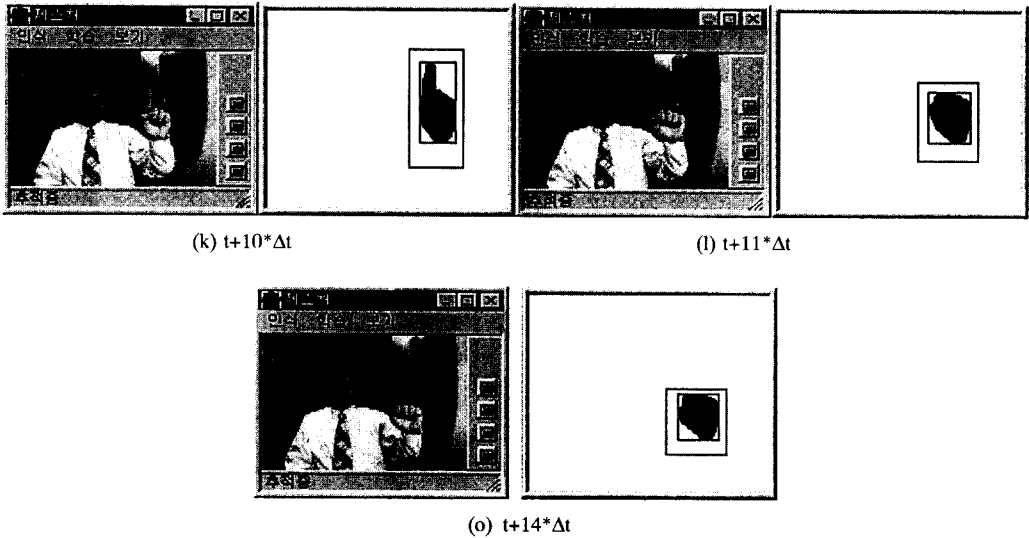


그림 6. 손영역 획득 결과.

직이는 때에는 손모양의 변화가 많지 않다는 전제조건을 사용하여 전 시점과 비교하여 영역의 형태가 매우 유사하고 매우 많이 이동한 영역을 손영역으로 획득한다. 형태의 유사성과 영역의 이동성을 구하기 위한 계산 식은 식 (13)과 같다. 최종적으로 현 시점의 모든 영역들에서 형태 유사성과 영역 이동성의 결과 값이 임계값 이상인 영역을 손영역으로 획득한다.

획득된 손영역으로부터 구분력이 좋은 특징 값을 얻기 위해서는 먼저 손가락의 끝이 위를 향하도록 위치를 정규화 하여야 한다. 손영역의 방향을 구하기 위해 기본축의 기울기를 계산한다. 기본축은 그림과 같이 영역의 무게중심을 지나는 장축과 단축을 말한다. 기본축의 기울기 θ 는 2차 모멘트(moment)를 이용하여 모멘트의 고유벡터를 계산함으로써 구한다. 고유벡터는 두 개의 기본축인 주 기본축과 부 기본축을 표현하고, 고유 값인 K 는 2차 모멘트로 정의되어 있다. K 는 식 (15)와 같으며, 기본축의 기울기는 식 (16)과 같이 정의된다.

$$K = \frac{(m_{20} + m_{02}) + \sqrt{(m_{20} - m_{02})^2 + 4m_{11}^2}}{2} \quad (15)$$

$$\theta = \tan^{-1} \left(\frac{K - m_{02}}{m_{11}} \right) \quad (16)$$

주기본축의 단일 방향을 결정하기 위해서 다음의 두 가지 제약 사항을 고려한다. 첫 번째 조건은 $m_{20} \geq m_{02}$ 이다. 만일, 첫 번째 조건이 만족되지 않으면 θ 는

90도씩 증가된다. 두 번째 조건은 $m_{03} < 0$ 이다. 만일, 두 번째 조건이 만족되지 않으면 θ 는 180도씩 증가된다. 최종적으로 θ 는 두 가지 제약 사항을 고려한 값을 갖는다. 식 (16)에 의해 계산된 기본축의 기울기는 손영역을 손가락의 끝이 위를 향하도록 회전시킨다. 식 (17)은 손영역을 회전시키는 수식을 보인다.

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \quad (17)$$

4. 실험 및 결론

손영역 획득 모듈이 입력 영상으로부터 손영역을 획득하는 결과를 그림 6에서 보인다. 그림 6의 손영역 획득 실험은 1분 동안 수행되었고, 초당 3-4 프레임 임을 처리하므로 1분 동안에 획득되어 처리되어진 입력 영상은 약 200 프레임이다. 그림 6은 200 프레임 중에서 14 프레임을 적당한 시간 간격으로 획득하여 순서대로 보인 것이다.

왼쪽 영상은 입력 영상을 나타내고 오른쪽 영상은 획득된 손영역을 보인다. 오른쪽 영상에서 손영역을 가장 근접하게 둘러싸고 있는 사각형은 손영역의 최소인접사각형을 나타낸 것이고 조금 큰 형태의 사각형은 칼만필터에 의해 예측된 탐색영역을 나타낸다.

그림 6(a)부터 (e)까지는 초기 손영역 추출 단계를 보이며 전 시점으로부터 손영역을 추적하여 획득하는 단계는 그림 6(e)부터이다. 그림 6(e)부터 (f)까지는 손을 위에서 그리고 다시 아래로 위로 움직이고

있다. 이때 칼만필터의 예측에 의한 탐색영역이 손이 움직이던 방향으로 치우쳐 있다. 또한 그림 6(g)와 같이 움직임이 크지 않을 경우에는 탐색영역의 중심과 손영역의 중심이 유사하다. 이러한 결과는 칼만필터에 의한 예측이 잘 수행되고 있음을 보인다. 만약 그림 6(j)와 같이 다른 영역과 손영역이 겹쳐진 경우에는 다른 영역과 손영역이 하나의 영역으로 합쳐진다. 그러나 다시 손영역이 분리되면 그림 6(k)에서 보는 것과 같이 원래의 손영역을 계속 추적하게 된다. 이 결과에 의하여 손영역 획득 모듈은 손영역이 다른 영역과 겹치는 경우에도 안정적으로 다음 시점의 손영역을 계속 추적할 수 있다는 것을 알 수 있다.

본 논문에서는 컴퓨터 시각을 이용한 제스처 기반 인터페이스 시스템의 구현에 필요한 손영역 획득 및 추적 방법을 제안하였다. 칼라 입력 영상의 영역 분리를 쉽게 하기 위해 정의한 손색상 모델과 손색상 결정함수는 손색상에 해당하는 영역을 배경 영상으로부터 잘 분리하였고, 손색상 모델의 갱신과 탐색영역의 제한을 위해 사용한 칼만필터도 좋은 결과를 보여 주었다. 손영역 획득은 초기 손영역을 사용자와의 상호작용에 의해 획득한다. 그리고 다음 시점부터는 칼만필터에 의해서 탐색영역을 제한하고, 탐색영역 내에서만 손색상 모델을 이용한 영역 분리를 수행하여 추출된 후보 손영역들 중에서 전 시점의 손영역과 유사도가 가장 높은 후보 손영역을 현 시점의 손영역으로 획득하였다. 결론적으로 손색상 모델과 칼만필터를 이용한 영역 추출과 추적에 의한 손영역 획득 방법은 복잡한 배경 영상으로부터 안정적으로 손영역을 획득할 수 있는 가능성을 제시하였다.

참고문헌

[1] Francis K.H. Quek, "Toward a Vision-Based Hand Gesture Interface," Conference on Virtual Reality Software and Technology, pp.23-26, 1994.
 [2] Thomas S. Huang, Vladimir I. Pavlovic, "Hand Gesture Modeling, Analysis, and Synthesis," International Workshop on Automatic Face- and Gesture-Recognition, pp.73-79, 1995.
 [3] Ying Dai, Yasuaki Nakano, "Extraction of Facial

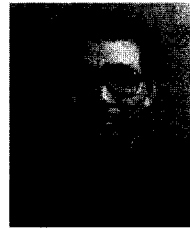
Images from Complex Background using Color Information and SGLD Matrices," International Workshop on Automatic Face- and Gesture-Recognition, pp.238-242, 1995.
 [4] Yu-Ich Ohta, Takep Kanade, Toshiyuki Sakai, "Color Information for Region Segmentation," Computer Vision, Graphics, and Image Processing, **13**, pp.222-241, 1980.
 [5] Christopher C. Yang, Jeffrey J. Rodriguez, "Efficient Luminance and Saturation Processing Techniques for Bypassing Color Coordinate Transformations," IEEE Systems, Man and Cybernetics, pp.667-672, 1995.
 [6] 양선옥, 최형일, 윤석환, "사용자 상호작용에 의한 적응적 분할 정보 획득," 정보과학논문지(B), **23**(12), pp. 1289-1297, 1995.
 [7] Haiyuan Wu, Qian Chen, Masahiko Yachida, "An Application of Fuzzy Theory : Face Detection," International Workshop on Automatic Face- and Gesture-Recognition, pp.314-319, 1995.
 [8] Dana H. Ballard, Christopher M. Brown, Computer Vision, Prentice Hall, 1982.
 [9] Ronald Lumia, Linda Shapiro, Oscar Zuniga, "A New Connected Components Algorithm for Virtual Memory Computers," Computer Vision, Graphics, and Image Processing, **22**, pp.287-300, 1983.
 [10] Hironobu Takahashi, Fumiaki Tomita, "Fast Region Labeling with Boundary Tracking," IEEE International Conference on Image Processing, pp.369-373, 1989.
 [11] Francois G. Meyer, Patrick Bouthemy, "Region-Based Tracking using Affine Motion Models in Long Image Sequences," Computer Vision, Graphics, and Image Processing : Image Understanding, **60**(2), pp. 119-140, 1994.
 [12] Nuria Oliver, Alex Pentland, "LAFTER : Lips and Face Real Time Tracker," M.I.T Media Laboratory Perceptual Computing Section TR-396, 1996.
 [13] Mohinder S. Grewal, Angus P. Andrews, Kalman Filtering Theory and Practice, Prentice Hall, 1993.
 [14] G. Minkler, J. Minkler, Theory and Application of Kalman Filtering, Magellan, 1994.
 [15] 장대식, 김계영, 최형일, "모델에 기반한 이동물체의 추적," 정보과학논문지(B), **23**(12), pp.1298-1308, 1995.



양선옥(Seon-Ok Yang)

1991년 : 숭실대학교 전자계산학과 졸업(공학사)
1993년 : 숭실대학교 대학원 전산학과 졸업(공학석사)
1998년 : 숭실대학교 대학원 전산학과 박사과정 수료
1997년~현재 : 숭실대학교 전자계산원 전임강사 재직 중

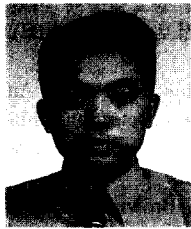
주관심분야 : 제스처 인식, 동작이해, 지식기반 시스템, 화상기반 원격교육시스템 등



최형일(Hyung-II Choi)

1979년 : 연세대학교 전자공학과 졸업(공학사)
1982년 : 미시간대학교 전산공학과 졸업(공학)
1987년 : 미시간대학교 전산공학과 졸업(공학)
현재 : 숭실대학교 컴퓨터학부 교수 재직 중

주관심분야 : 컴퓨터 비전, 퍼지 및 신경망 이론, 패턴 인식, 지식기반 시스템 등



고일주(Il-Ju Ko)

1992년 : 숭실대학교 전자계산학과 졸업(공학사)
1994년 : 숭실대학교 대학원 전산학과 졸업(공학석사)
1997년 : 숭실대학교 대학원 전자계산학과 졸업(공학박사)
1997년~현재 : 숭실대학교 생산기술연구소 연구원으로 근무중

관심분야 : 제스처 인식, 동작이해, 지식기반 시스템, 컴퓨터 비전, 패턴 인식 등