

멀티미디어 시스템을 위한 영상내의 손 인식에 관한 연구

A Study on Hand Recognition in Image for Multimedia System

양환석*, 정혜원**

조선대학교 자연과학대학 전산통계학과*, 조선이공대학 시각디자인 애니메이션과**

Hwan-Seok Yang(badhack@chosun.ac.kr)*, Hye-Won Jung(hwjung@mail.chosun-c.ac.kr)**

요약

본 논문에서는 별도의 센서 없이 영상만을 이용하여 실시간으로 손 영상을 인식하는 알고리즘을 제안한다. 손은 모양이 매우 복잡하기 때문에 2차원 모양의 불변량에 해당하는 에지의 방향성 히스토그램을 이용하여 인식을 행한다. 이 방법은 복잡한 배경에서 색상정보를 이용하여 손 영역이 정확히 추출되며 계산량이 적고 조명변화에 덜 민감하기 때문에 실시간 손 영상 인식에 적합하다. 본 논문에서는 손의 모양제시 방향이 변하는 경우에도 인식을 가능하게 하기 위해 주성분 분석법을 사용하여 인식오차를 줄이는 방법을 기술한다. 이 방법을 사용함으로써 손 영상이 3차원적으로 회전에 의해 변하는 경우도 인식가능하게 되었다. 또한 에지방향성 데이터를 이용하기에 주성분 공간 생성 시간을 현저히 줄이게 되었다.

■ 중심어 : | 영상 인식 | 주성분 분석 | 에지 방향성 히스토그램 |

Abstract

In this paper, we proposed an algorithm which cognize hand pose in real time using only image. Hand recognizes using edge orientation histogram which comes under a constant quantity of 2D appearance because hand pose is intricate. This method suit hand pose recognition in real time because it extracts hand space accurately, has little computation quantity, and is less sensitive to lighting change using color information in complicated background. Method which reduces recognition error using principal component analysis method to can recognize through hand shape presentation direction change is explained. A case that hand shape changes by turning 3D also by using this method is possible to recognize. Besides, principal component space creation time is reduced remarkably because edge directional data is used.

■ keyword : | Image Recognition | Principal Component Analysis | Edge Orientation Histogram |

I. 서론

물체 인식은 컴퓨터 비전 분야에서 가장 활발히 연구되고 있는 분야중의 하나이다[1][2]. 물체 인식을 정의하는 데에도 여러 분류가 있을 수 있지만 간단하게 정

의하자면 미리 데이터베이스를 가지고 있는 물체에 대해서 영상 안에 그 물체가 있는지 없는지를 밝혀내는 것이라 할 수 있다. 인식을 위한 영상들 중에 손은 인간에게 있어서 가장 중요한 도구의 하나로 그 동작과 모양이 의미하는 바를 인식하는 것은 정보 시스템과 인간

사이의 자연스러운 인터페이스를 실현할 수 있다는 점에서 오랜 기간 동안 많은 사람들의 연구테마로 다루어져 왔다[3]. 부가적인 하드웨어 없이 사람의 손 영상을 컴퓨터와 인간 사이의 상호작용을 위한 의사 전달 수단으로 사용한다면 기존의 하드웨어 의존적인 인터페이스를 사용자 중심의 인간 친화적인 인터페이스로 재구성할 수 있다. 그러나 복잡한 환경으로부터 사람의 손 영상만을 정확히 추출하고 인식하여 의미있는 입력정보로 사용하는 것은 어려운 일이다. 기본적으로 정확한 손 영역 분할을 위한 알고리즘이 필수이며 손 영상의 분석과 인식을 위한 빠른 처리 속도 또한 고려해야 할 요소이다 [4]. 기존의 연구들은 크게 손 영상을 분석하기 위해 Data Glove를 사용하는 방법[5]과 3차원 모델을 사용하거나 스테레오 비전을 이용하여 추출된 손의 3차원 기하학적인 정보를 분석하는 방법[6][7], 손의 2차원 패턴 정보를 분석하는 방법[8][9]으로 분류할 수 있다. Data glove를 사용하는 방법은 정밀도가 낮고 전체 시스템과 접속하기 위해 반드시 연결선이 필요하다는 점에서 자연스러운 인터페이스 구축에는 많은 문제점을 안고 있다[3]. 3차원 기하 정보를 이용하는 방법은 스테레오 영상을 이용한 3차원 정보추출과 복잡한 손의 3차원 운동 모델을 구축하기 위한 많은 계산량이 필요하므로 고성능의 하드웨어가 필요하다는 단점이 있다[10]. 2차원 정보를 이용하는 방법들은 사람의 피부색의 컬러 분포를 사전 학습과정에서 계산하여 입력 영상 내 해당 피부 컬러 분포에 해당되는 영역만을 추출하는 방법을 적용하였으나 대부분 단일 배경이거나 조명 변화가 거의 없다고 가정하였다.

본 논문에서는 손의 모양 변화로 인한 손 영역의 국부적인 조명변화와 복잡한 배경으로부터 손 영역을 추출한 후 화소간의 휘도 경도를 구해 에지 방향성 히스토그램을 만들고 이 히스토그램을 평활화하여 얻어진 특징 벡터를 모양 구분의 척도로 사용하였다. 또한 에지 방향성 히스토그램의 데이터(36개)를 파라메트릭 손 포즈 공간이라는 저차원적 공간으로 표현하여 3차원적으로 손의 방향이 틀어지는 현상을 포즈 단위로 인식가능하게 되었다.

본 논문의 구성은 다음과 같다. 2절에서는 입력 영상

에서 손 영상의 인식을 위해 피부색을 이용한 손 영역 추출 방법에 대하여 기술한다. 3절에서는 에지 방향성 히스토그램을 이용한 손 영상 인식 방법에 대하여 기술한다. 4절에서는 본 논문에서 제안한 방법에 대해 성능 평가를 하고 마지막으로 5절에서는 결론을 맺는다.

II. 피부색을 이용한 손 영역 추출

1. 손 영역 추출

칼라 공간에서 사람의 살색 분포는 자연계의 다른 물체와 구분할 수 있는 뚜렷한 특징을 가지고 있다. 비록 사람과 인종마다 피부색은 다르지만 색차 평면상에서의 그 분포는 매우 비슷하면서도 협소한 영역을 차지하고 있고, 단지 밝기(intensity)의 차이만이 있어, 색차 성분을 이용하면 쉽게 살색 영역을 추출할 수 있다. 칼라 정보만을 이용하여 손 영역을 검출할 경우에는 살색 계통의 물체가 배경영역에 존재하여 함께 검출될 수 있다. 따라서 보다 정확하고 빠른 검출을 위하여 본 논문에서는 다음과 같은 방법을 제안했다. 이 방법은 간단하고 빠르게 RGB 공간 안에서 피부색을 추출 할 수 있다.

RGB 데이터로부터 살색 영역을 검출하기 위해 YUV 색상 모델을 이용하였다. YUV 색상 모델은 일반적인 피부색상은 전체 색상 영역 중 매우 좁은 영역에 분포하므로 다른 색상 모델에 더욱 효율적으로 살색을 추출할 수 있다.

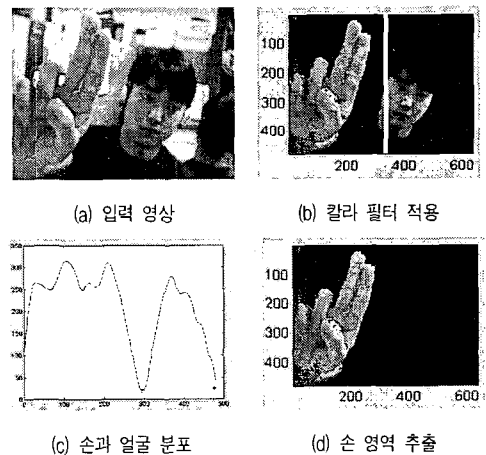


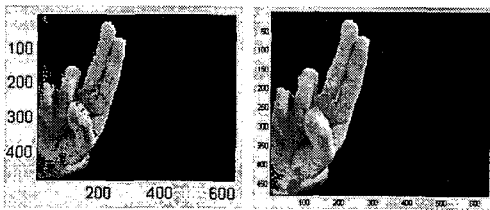
그림 1. 손 영역 추출 과정

그러나 매우 정확한 칼라 필터라도 영상에서 손만을 추출하기는 쉽지 않다. [그림 1]에서 보는 바와 같이 얼굴과 손에서 손만을 추출하기 위해서는 본 논문에서는 다음과 같은 방법을 사용했다.

YUV 색상 모델을 적용하여 얻어진 영역은 대개 얼굴과 손에 해당한다고 가정하면 이를 손으로 분류하는 방법을 찾으면 된다. 손이 머리보다 카메라에 가깝다고 가정하면 손 영역이 더 크게 보이고 면적정보를 이용하면 손을 구별해 낼 수 있다. 이 알고리즘은 다이내믹하게 조명이 변하는 상황에도 매우 좋은 결과를 나타내었다.

그러나 손을 추출한 영상에서 노이즈로 인해 손 모양이 매끄럽지 않은 현상이 생긴다. 이를 해결하기 위해 식 (1)을 이용하여 [그림 2]처럼 손의 노이즈를 제거하여 개선된 손 모양을 얻을 수 있었다.

$$g_c(m, n) = \bigcap_{(i,j) \in B} g_d(m+i, n+j) = \bigcap_{(i,j) \in B} \left(\bigcup_{(i',j') \in B} f(m+i-i', n+j-j') \right) \quad (1)$$



(a) 손 영역 추출 (b) 잡음 제거 및 보정
그림 2. 손 영역의 잡음 제거 및 보정 결과

III. 영상내의 손 인식

영상이 지닌 밝기의 산술적 차이만을 비교하여 손과 같이 복잡한 물체의 모양을 인식한다는 것은 매우 어려운 일이다. 본 논문에서는 실시간으로 손 모양을 인식하기 위해 입력 영상이 들어오면 에지 방향성 히스토그램을 생성한 후 이미 만들어진 데이터와 실시간으로 비교하여 손 모양을 인식한다. 2차원 회전에 불변 특성을 지닌 에지 방향성 히스토그램 정보를 이용하는 알고리

즘을 제안한다. [그림 3]은 본 논문에서 제안하는 인식 시스템의 구성도를 보여주고 있다.

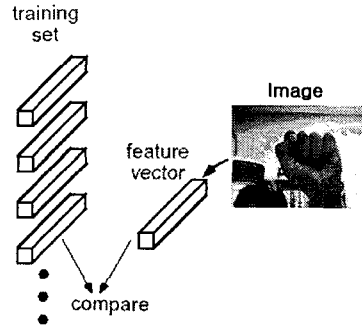


그림 3. 인식 시스템

1. 에지 추출

에지 방향성 히스토그램을 생성하기 위해서는 먼저 에지를 추출해야 하는데 본 논문에서는 식 (2)를 이용하여 에지를 구한다. 영상으로부터 추출한 에지를 가지고 에지의 방향성을 이용하여 에지 방향성 히스토그램을 생성하면 이것을 영상의 특징 정보로 사용할 수 있다.

$$magnitude = \sqrt{s_x^2(i, j) + s_y^2(i, j)} \quad (2)$$

위에 식으로 얻어진 에지 값은 모든 픽셀에 대하여 구해진다. 따라서 실제로 에지가 아닌 부분에서도 에지 값을 갖게 되므로 특정 픽셀이 에지인지 아닌지를 판별하는 과정이 필요하다. 특정 픽셀이 에지인지 아닌지를 판별하기 위해서 임계값에 따라 에지로 판정된 픽셀만을 대상으로 에지 방향성 히스토그램을 생성한다. 각 임계값별 에지 영상은 [그림 4]와 같다.

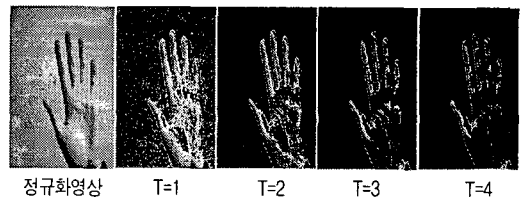


그림 4. 임계값(T)에 따른 에지영상

2. 에지 방향성 히스토그램 생성

임계값으로 얻어진 에지 영상을 16 단계의 방향으로 식 (3)을 이용하여 에지 방향성 히스토그램을 생성한다.

$$direction = \tan^{-1} \frac{s_y(i, j)}{s_x(i, j)} * \frac{36}{\pi} \quad (3)$$

식 (3)으로부터 얻어진 픽셀들을 360°로 세분된 공간에 매핑하고 픽셀들의 수로부터 얻어진 방향성 히스토그램을 모델 영상과 비교하여 실시간으로 손 영상을 인식하게 된다. [그림 5]는 방향성 히스토그램을 생성한 모습이다.

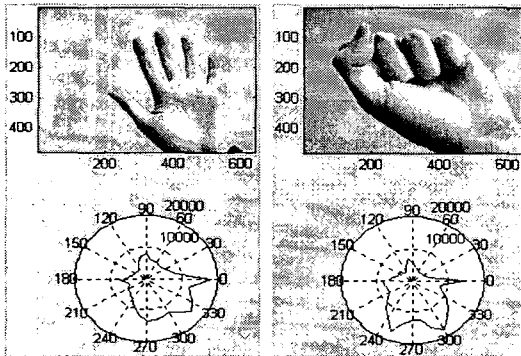


그림 5. 방향성 히스토그램

3. 주성분 분석을 이용한 손 모양 인식

단일 카메라를 이용하여 손의 방향이 틀어지는 경우를 같은 모양으로 인식하기에는 많은 어려움이 있다. 본 논문에서는 [그림 5]에서와 같은 손의 방향성이 틀어지는 경우에도 같은 모양으로 인식하기 위해 주성분 분석을 이용한 손 방향성 문제의 해결 방법을 제시한다. 앞에서 설명한 방법을 통하여 얻어진 36개의 에지 방향성 히스토그램의 데이터들을 이용하여 손 영상의 전체적인 외관 특징을 표현할 수 있는 저차원 벡터 공간, 즉 파라메트릭 공간을 생성한다. 이 공간은 주성분 분석법이라는 통계적 방법에 의해 만들어진다. 주성분 분석이란 각 손 모양 영상의 36개의 에지 방향성 히스토그램의 데이터들의 공간적 위치 값이 주로 각 영상에서 어디에 분포하는가를 계산하여 확률 빈도가 높은 벡터 값들을 고

유치 값에 비례하여 재구성하는 방법이다. 따라서 이 방법은 고유벡터(eigenvector)와 고유치(eigenvalue)를 계산하여, 손 모양의 36개의 평균 모델을 구하여 이용한다. [그림 6]은 6개의 손 모양을 10°씩 틀어서 획득한 영상이다.(1개의 포즈 = 40 image)

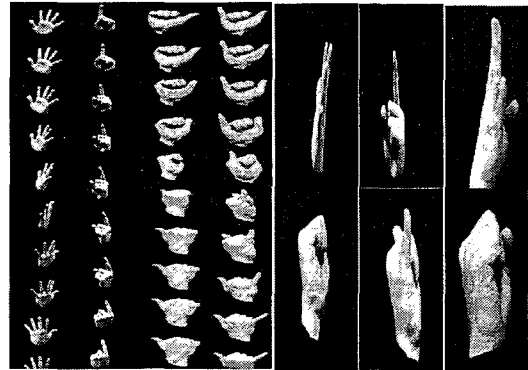


그림 6. 6개의 손 모양 및 한 포즈 당 10°씩 획득한 영상

고유벡터를 계산하기 위해서는 먼저 [그림 7]에서 보는 바와 같이 에지 방향성 히스토그램의 데이터를 정규화 한 후 모든 데이터의 평균 데이터를 구하여 각 영상들의 데이터의 차를 구한다. 평균 데이터 C와 새로운 영상 데이터 집합 X를 식 (4)와 식 (5)와 같이 나타낸다.

$$C = (1/N) \sum_{i=1}^N x_i \quad (4)$$

$$X = \{x_{1,1}^{(1)} - c, x_{2,1}^{(2)} - c, \dots, x_{R,1}^{(b-1)} - c, x_{R,L}^{(b)} - c\} \quad (5)$$

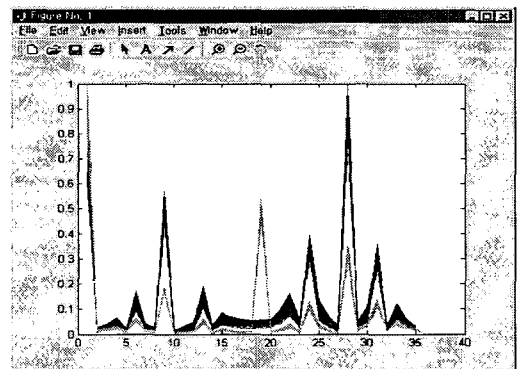


그림 7. 에지 방향성 히스토그램의 정규화 데이터

[그림 7]은 40개의 이미지에 대한 36개의 데이터와 3개의 포즈의 값을 예시 방향성 히스토그램의 정규화한 모습을 보여주고 있다.

고유공간을 구하기 위해서는 $M * N$ 의 크기를 지닌 영상 집합 X 를 식 (6)과 같이 계산하고 식 (7)을 만족하는 고유벡터를 구하면 된다.

즉, 공분산 행렬 Q 에 대한 고유치 λ 와 고유벡터 e 를 구한다.

$$Q = XX^T \quad (6)$$

$$\lambda_i e_i = Q e_i \quad (7)$$

본 논문에서는 특이치 분해(Singular Value Decomposition)를 이용하여 고유 공간을 구하였다. 특이치 분해를 이용하여 식 (8)에서처럼 행렬이 분해되는데, 영상집합 X 의 공분산 행렬에 대한 고유 벡터를 쉽게 얻을 수 있다.

$$[X] = [V] \cdot \begin{bmatrix} \sigma^1 & & \\ & \sigma^{21} & \\ & & \sigma^{31} \end{bmatrix} \cdot [U^T] \quad (8)$$

행렬은 식 (9)와 같이 분해가 가능하며 원 식은 식 (13)과 같이 나타낼 수 있다.

$$X = \sum_{i=1}^r \sigma_i V_i u_i^T \quad (9)$$

$$V = [v_1, \dots, v_r] \quad (10)$$

$$U = [u_1, \dots, u_r] \quad (11)$$

$$\Sigma = \text{diag}(\sigma_1, \dots, \sigma_r) \quad (12)$$

$$X = V \Sigma U^T \quad (13)$$

V 와 U 행렬은 각각 직교 행렬이기 때문에 식(14)에서 나타낸 것처럼 표현된다.

$$[V^T] \cdot [V] = [U^T] \cdot [U] = [I] \quad (14)$$

이런 과정으로 얻어진 행렬 중 공분산의 고유 벡터로 이용되는 행렬은 공분산 행렬 X 와 크기가 일치하는 U 이다. 특이치 분해 과정에서 나온 고유벡터를 고유치가 큰 순서대로 재구성한다. 각 고유벡터가 지닌 고유치의 크기는 그 고유벡터의 중요도를 의미하므로 그 고유공간을 규정하는 중요 고유벡터를 식 (15)를 이용하여 선택한다. 따라서 모든 고유벡터를 고유 공간 구성에 사용하지 않고 많은 영상을 대표할 수 있는 주성분의 벡터만을 이용할 수 있다.

$$\frac{\sum_{i=1}^k \lambda_i}{\sum_{i=1}^N \lambda_i} \geq T_1 \quad (15)$$

여기서 T_1 은 고유벡터의 개수를 조정하는 문턱치이며, 인식과 포즈 평가시 이용되는 고유벡터는 저차원 공간을 구성하기 위해 실험에서는 $K=3$ 을 이용했다. 이렇게 구성된 공간을 손 포즈 공간으로 표현한다.

위와 같이 얻어진 손 포즈 공간에서 평균 영상 C 에서 뺀 영상 집합 x 를 모두 식(16)을 이용하여 투영시킨다.

$$f_j = [e_1, e_2, e_3, \dots, e_k]^T (x_n - c) \quad (16)$$

구해진 f_j 손 포즈 공간상에서 점들로 표현되는데 이러한 점들을 손 포즈 특징 심벌을 구한 것과 같이 입력 특징 심벌을 구성하게 된다. 투영시킨 결과는 이산적인 점들로 표현되며, 이들 각 점은 입력 동작 하나 하나를 의미한다. 고유공간의 특성상 비슷한 특징 값을 가지는 벡터는 고유공간에서도 가까운 곳에 투영된다. 따라서 같은 동작의 경우 서로 비슷한 특징 벡터 값을 가지게 되고, 고유공간에서 비슷한 영역에 분포하게 된다. 6개의 손동작을 본 고유공간에 투영시킴으로서 모델 동작

을 모델링 한다. 각 모델동작에 대한 고유공간내의 분포는 [그림 8]과 같다.

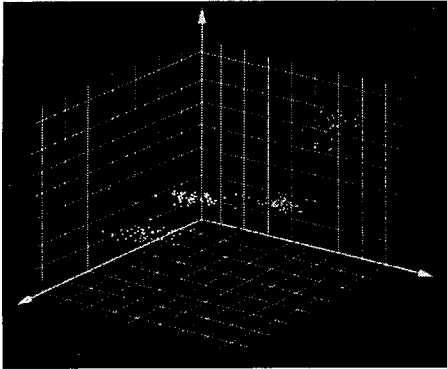


그림 8. 고유공간에서의 모델 손동작 분포

고유 공간에 투영된 점이 가까울수록 영상들의 높은 상관관계를 가지기 때문에 손의 모양 이미지가 입력이 순차적으로 따라가면서 인식을 수행하게 된다. 즉 거리가 가까울수록 두 영상은 닮은 영상이 된다.

IV. 실험 및 결과

1. 손 영상 인식

실험은 범용 USB 카메라(초당30프레임)를 장착한 펜티엄 4 2.4GHz, 메모리 512MB를 가진 PC에서 수행하였으며, 프로그램은 Visual C++을 사용하여 구현하였다. 손 영상을 인식하기 위해서 사용한 영상은 15가지의 의미를 부여한 손 모양을 사용하였다. 각 영상의 크기는 640*480이며 실시간으로 인식했다. [그림 9]는 실험에 사용한 손 영상을 보여주고 있다.

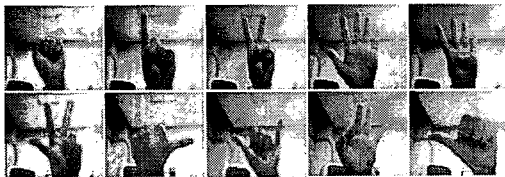


그림 9. 실험에 사용한 손 영상

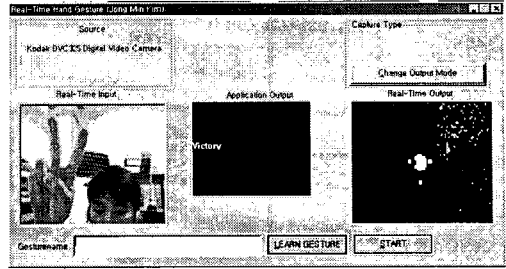


그림 10. 손 인식 시스템

[그림 10]은 손 모양을 인식하는 시스템이다. 왼쪽 창은 실시간 입력 영상을 보여주고 있으며 가운데 창은 입력 영상의 인식 결과를 보이고 오른쪽 창은 방향성 히스토그램 생성결과 및 간단한 손 모양 추적을 보여준다. [표 1]에서는 실험에 사용된 영상과 인식 결과를 수치로 나타내었다.

표 1. 실험 결과

영상의 종류	영상의 개수	성공률(%)	실패율(%)
단순한 영상	15	94.5	5.5
복잡한 영상	15	90.5	9.5

2. 실험 결과

실험은 손 모양의 한 이미지가 아닌 한 포즈 단위로 인식을 행하게 된다. 여기서 한 포즈는 40(10씩 회전해서 얻은 이미지수) × 36(한 이미지의 에지방향성 히스토그램)로 나타내어진다. 포즈는 6개의 포즈로 공간을 구성한 후에 실험을 하였고 인식률은 평균 85%를 얻을 수 있었다.

표 2. 실험 결과

매칭방법	포즈의 개수	매칭 실패	잘못된 매칭	매칭 성공
거리계산	6	6%	9%	85%

V. 결론

본 논문에서 제안한 컬러 정보를 이용한 실시간 손 영

상 인식은 수행속도가 매우 빠르며, 어느 정도의 손 방향의 변화를 모델을 구성함으로써 같은 모양으로 인식하게 되었고, 조명 변화 및 복잡한 배경에서도 안정적인 인식률을 보였다. 본 연구의 실용성을 위해 다른 응용 프로그램(윈도우 미디어 플레이어) 제어에 적용해 본 결과 좋은 인식률과 빠른 수행 속도를 보여 주었다. 그리고 장갑이나 다른 센서를 달지 않기 때문에 저 비용 인식 시스템이라 할 수 있다. 또한 손의 방향이 틀어지는 경우 인식이 불가능한 문제를 주성분 분석법을 이용하여 손의 포즈단위로 인식을 함으로 View Direction 문제를 해결하였다.

그러나 단지 영상의 컬러 정보만을 이용하기 때문에 배경에 살색 영역이 많이 포함되거나 얼굴과 손이 겹쳐서 나타날 경우 손을 잘 찾지 못하는 경우가 발생하였고, 손 모양과 포즈 인식 시 두 시스템간의 약간의 충돌이 발생하였다. 향후 이런 문제점 보완과 3차원 데이터를 이용하는 연구와 수행되어야 한다.

참고 문헌

[1] J. Ohya, and Y. Kitamura, et al, "Real-Time Reproduction of 3D Human Images in Virtual Space Teleconferencing," in Proc. of '93 IEEE Virtual Reality Annual Int. Symp. pp.408~414, 1993.

[2] J. Segen, and S. Kumar, "Shadow Gestures: 3D Hand Pose Estimation Using a Single Camera," CVPR99, Vol.1, pp.479~485, 1999.

[3] K. Imagawa, S. Lu, and S. Igi, "Color-Based Hand Tracking System for Sign Language Recognition," Proceedings of International Conference on Automatic Face and Gesture Recognition, Killington, 1996.

[4] S. Ahmad, "A Usable Real-Time 3D Hand Tracker," Conference Record of the Asilomar Conference on Signals, System and Computers, pp.1257~1261, 1994.

[5] V.Pavlovic, R. Sharma, and T.Huang,

"Visual Interpretation of Hand Gestures for Human-Computer Interaction: A Review," IEEE Trans. on PAMI, Vol.19, No.7, pp.677~695, 1997.

[6] W. T. Freeman, and M. Roth, "Orientation histograms for hand gesture recognition," Intl. Workshop on Automatic Face and Gesture Recognition, IEEE Computer Society, pp.296~301, June 1995,

[7] Hiroshi Murase and Shree K, Nayar, "Visual Learning and Recognition 3-D object from appearance," International Journal of Computer Vision, Vol.14, 1995.

[8] D. G. Lowe, "Three-dimensional object recognition from single twodimensional images," Artificial Intelligence, Vol.31, pp.355~395, 1987.

[9] R. Ciplolla, Y. Okamoto, and Y. Kuno, "Qualitative visual interpretation of 3D hand gestures using motion parallax," in Proc. of MVA'92, pp.477~482, 1992.

[10] V.Pavlovic, R. Sharama, and T. Huang, "Visual Interpretation of Hand Gestures for Human-Computer Interaction: A Review," IEEE Trans. on PAMI, Vol.19, No.7, pp.677~695, 1997.

저자 소개

양 환 석(Hwan-Seok Yang)

정회원



- 1996년 2월 : 호원대학교 전자계산학과(이학사)
- 1998년 2월 : 조선대학교 전산통계학과(이학석사)
- 2005년 2월 : 조선대학교 전산통계학과(이학박사)

• 현재 : 조선대학교, 서강정보대학 시간강사

<관심분야> : 영상 처리, 컴퓨터 비전, 모바일 콘텐츠

정혜원(Hye-Won Jung)

정회원



- 1992년 2월 : 전남대학교 시각디자인전공(미술학사)
 - 1998년 8월 : 조선대학교 산업디자인과(미술학 석사)
 - 2005년 2월 : 조선대학교 전산멀티미디어전공(이학박사)
 - 2001년 3월~현재 : 조선이공대학 시각디자인·애니메이션과 조교수
- <관심분야> : 콘텐츠디자인, 애니메이션콘텐츠, 문화콘텐츠 등