

파라메트릭 손 포즈 공간에서 방향성 히스토그램 데이터를 이용한 손 포즈 인식

김종민, 위승정, 양환석, 이용기*
조선대학교 대학원 수학·전산통계학부

mrjjoung@hanmail.net, badhack@image.chosun.ac.kr, lee@image.chosun.ac.kr

Hand Pose Recognition Using Orientation Histogram Data In Hand Pose Space

Jong-Min Kim, Seung-Jung Wi, Hwan-Seok Yang, Woong-Ki Lee*

Division Mathematics and Computer Science and Statistic Graduate School, Chosun. Univ.

* Division Mathematics and Computer Science and Statistic College of Natural Sciences, Chosun. Univ.

요 약

본 논문에서는 별도의 센서를 부착하지 않고 영상만을 이용하여 실시간으로 손 형상을 인식하는 알고리즘에 대해 기술한다. 손은 형상이 매우 복잡하기 때문에 2차원 형상의 불변량에 해당하는 에지 방향성 히스토그램을 이용하여 인식을 행한다. 이 방법은 복잡한 배경에서 색상정보를 이용하여 손 영역이 정확히 추출되면 계산 량이 적고 조명변화에 덜 민감하기 때문에 실시간 손 형상 인식에 적합하다. 본 논문에서는 손의 형상제시 방향이 변하는 경우에도 인식을 가능하게 하기 위해 주성분 분석법을 사용하여 인식오차를 줄이는 방법을 기술한다. 이 방법을 사용함으로써 손 형상이 3차원적으로 회전에 의해 변하는 경우도 인식이 가능하게 되었다. 실험부분에서 제안하는 방법을 이용하여 가정용 가전제품이나 게임을 제어하는 실시간 휴먼 인터페이스 시스템 제작기술로 활용될 수 있음을 보인다.

1. 서 론

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 기간동안 많은 사람들의 연구테마로 다루어져 왔다. 부가적인 하드웨어 없이 사람의 손 형상을 컴퓨터와 인간 사이의 상호작용을 위한 의사 전달 수단으로 사용한다면 기존의 하드웨어 의존적인 인터페이스를 사용자 중심의 인간 친화적인 인터페이스로 재구성할 수 있다.

기존의 연구들은 크게 손 형상을 분석하기 위해 Data Glove를 사용하는 방법[1]과 3차원 모델을 사용하거나 스테레오 비전을 이용하여 추출된 손의 3차원 기하학적인 정보를 분석하는 방법[2], 손의 2차원 패턴 정보를 분석하는 방법[3-4]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. 3차원 기하 정보를 이용하는 방법은 스테레오 영상을 이용한 3차원 정보추출과 복잡한 손의 3차원 운동 모델을 구축하기 위한 많은 계산 량이 필요하므로 고성능의 하드웨어가 필요하다 는 단점이 있다. 2차원 정보를 이용하는 방법들은 사람의 피부

색의 컬러분포를 사전 학습과정에서 계산하여 입력 영상 내 해당 피부 컬러 분포에 해당되는 영역만을 추출하는 방법을 적용 하였으나 대부분 단일 배경이거나 조명 변화가 거의 없다고 가정하였다. 본 논문에서는 손의 형상 변화로 인한 손 영역의 국부적인 조명변화와 복잡한 배경으로부터 손 영역을 추출한 후 화소간의 휘도 경도를 구해 에지 방향성 히스토그램을 만들고 이 히스토그램을 평활화 하여 얻어진 특징 벡터를 형상 구분의 척도로 사용하였다. 또한 에지 방향성 히스토그램의 데이터(36개)를 주성분 분석하여 3차원적으로 손의 방향이 바뀌는 경우도 인식이 가능하게 되었다.

2. 전체 시스템 개요

본 논문에서는 피부색의 컬러 분포를 이용하여 손 영역을 추출한 후 손 영역의 에지 방향성 히스토그램을 구하고 이 정보에 주성분 분석법을 가하여 손 형상을 인식한다. 제안된 알고리즘의 전체 구성 도를 그림 1에 나타내었다. 카메라를 통하여 입력 영상이 들어오면 각 단계에서 제안된 블록단위의 알고리즘이 수행되어, 손 영역이 추출되고 이로부터 에지에 대한 방향성 히스토그램을 얻는다. 추출된 영상을 모델 영상과 비교하여 손 형상에 대한 어휘를 출력한다.

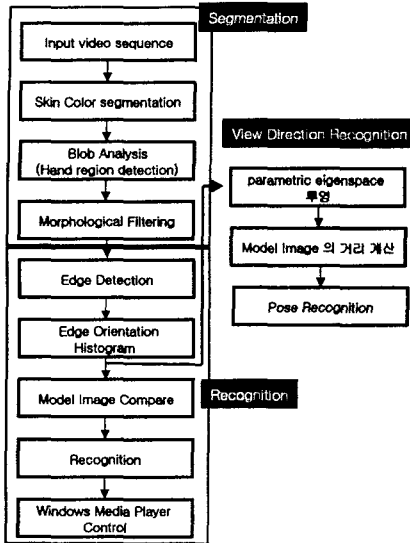


그림 1 시스템 전체 구성도

View-Direction 부분에서는 에지 방향성 히스토그램 데이터를 이용하여 주성분 분석한 후 고유공간에서 모델 영상과 비교하여 손의 형상그룹 단위로 인식을 하게 된다. 각 단계에서의 결과 영상은 순차적으로 다음 입력으로 사용되게 된다.

3. 손 영역 추출 및 인식

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.

3.1 피부색을 이용한 손 영역 추출

칼라 정보만을 이용하여 손 영역을 검출할 경우에는 살색 계통의 물체가 배경영역에 존재하여 함께 검출될 수 있다. 따라서 보다 정확하고 빠른 검출을 위하여 본 논문에서는 YUV 색상 모델 방법을 제안했다. YUV 색상 모델은 색상 정보로부터 광도를 분리하는 또 다른 색상 모델이다. 영상의 광도는 Y로 표시되고, 영상의 푸른 정도와 붉은 정도는 UV로 기호화된다. RGB 정보로부터 YUV 모델로의 변환은 식 1과 같이 구할 수 있다.

$$\begin{cases} Y = 0.29R + 0.59G + 0.11B \\ U = -0.147R - 0.289G + 0.436B = 0.492(B - Y) \\ V = 0.615R - 0.515G - 0.1B = 0.877(B - Y) \end{cases} \quad \text{식(1)}$$

특히, YUV 색상 모델에서 일반적인 피부색상은 전체 색상 영역 중 매우 좁은 영역에 분포하므로 다른 색상 모델에 더욱 효율적으로 피부색을 추출할 수 있다. 또한 YUV 색상 모델의 특성상 RGB 색상 모델에 비해 조명 변화에 덜 민감하기 때문에 조명 조건의 변화에도 안정적인 결과를 보인다.

3.2 손 형상 인식

영상이 지닌 밝기의 산술적 차이만을 비교하여 손과 같이 복잡한 물체의 형상을 인식한다는 것은 매우 어려운 일이다.

본 논문에서는 실시간으로 손 형상을 인식하기 위해 우리는 간단하고 빠른 알고리즘(에지 방향성 히스토그램)을 선택했다.

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

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

임계 처리로 얻어진 에지영상을 16 단계의 방향으로 에지 방향성 히스토그램을 생성한다.(식 3)

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

식(3)으로부터 얻어진 픽셀들을 36저장소에 매핑하고 픽셀들의 수로부터 얻어진 방향성 히스토그램을 모델영상과 비교하여 실시간으로 손 형상을 인식하게 된다. 그림 2는 방향성 히스토그램을 생성한 모습이다.

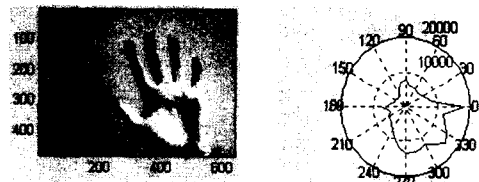


그림 2 방향성 히스토그램

4. 주성분 분석법을 이용한 손 포즈 인식

단일 카메라를 이용하여 손의 방향이 틀어지는 경우를 같은 형상으로 인식하기에는 많은 어려움이 있다. 본 논문에서는 손의 방향성이 틀어지는 경우(그림 3)에도 같은 형상으로 인식하기 위해 주성분 분석법을 이용한 손 방향성 문제의 해결 방법을 제시한다. 3.2에서 설명한 방법을 통하여 얻어진 36개의 에지 방향성 히스토그램의 데이터들을 이용하여 손 형상의 전체적인 외관 특징을 표현할 수 있는 저차원 벡터 공간, 즉 파라메트릭 공간을 생성한다. 이 공간은 주성분 분석법이라는 통계적 방법에 의해 만들어진다.

4.1 주성분 분석법을 이용한 고유 공간 구성

고유벡터를 계산하기 위해서는 먼저 그림 4에서 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)과 같이 나타낸다.

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

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

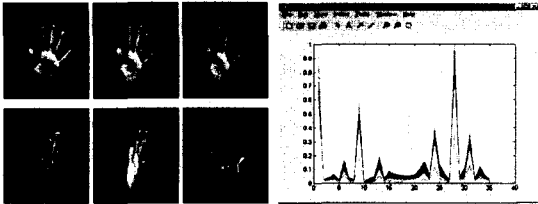


그림 3 손의 방향이 틀어지는 경우의 일부 영상
그림 4 에지 방향성 히스토그램의 정규화 한 데이터(40이미지*36데이터*3개의포즈)

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.

$$Q = XX^T \quad \text{식(6)}$$

$$\lambda, e_i = Qe_i \quad \text{식(7)}$$

여기서 M 은 한 영상의 에지방향성 히스토그램의 데이터이고 N 은 전체 영상의 개수를 나타내는 정수이다. 본 논문에서는 특이치 분해(Singular Value Decomposition)를 이용하여 고유 벡터를 구하였다. 특이치 분해를 이용하여 식(8)에서처럼 행렬이 분해되는데, 영상집합 X 의 공분산 행렬에 대한 고유 벡터를 쉽게 얻을 수 있다.

$$X = V \sum U^T \quad \text{식(8)}$$

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.

4.2 고유 공간에서 손 포즈 인식

앞 절에서 얻어진 손 포즈 공간에서 평균 영상 C 에서 n 영상 집합 x 를 모두 식(9)을 이용하여 투영시킨다.

$$f_j = [e_1, e_2, e_3, \dots, e_k]^T (x_n - c) \quad \text{식(9)}$$

구해진 f_j 손 포즈 공간상에서 점들로 표현되는데 이러한 점들을 손 포즈 특징 심볼을 구한 것과 같이 입력 특징 심볼을 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 동작을 모델링 한다. 각 모델동작에 대한 고유공간내의 분포는 그림 5와

같다.



그림 5 고유공간에서의 모델 손동작 분포

고유 공간에 투영된 점이 가까울수록 영상들의 높은 상관관계를 가지기 때문에 손의 형상 이미지가 입력이 순차적으로 따라가면서 인식을 수행하게 된다.

5. 손 형상 인식 실험 결과 및 결론

실험은 손 형상의 한 영상이 아닌 그룹단위로 인식을 행하게 된다. 여기서 한 포즈는 $40(5^{\circ} \text{씩 회전해서 얻은 이미지수}) * 36$ (한 이미지의 에지방향성 히스토그램)로 나타내어진다. 포즈는 6개의 포즈를 공간을 구성한 후에 실험을 하였다. 인식률은 평균 85%이나 약간의 제약을 가하면 100% 가까운 인식이 가능하다.

표 1. 실험 결과

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

5. 결론

본 논문에서는 실시간 고속 처리를 목적으로 에지방향성 히스토그램을 이용한 손 형상 알고리즘에 대해 기술하였다. 손 영역이 정확히 세그멘테이션이 되면 인식률이 매우 높으나 세그멘테이션 과정에서 영상의 컬러 정보만을 이용하기 때문에 배경에 살색 영역이 많이 포함되거나 얼굴과 손이 겹쳐서 나타날 경우 인식에 어려움이 있다. 향후 이 문제점을 보완하고 3차원 데이터를 이용하는 방법과 결합하면 보다 인식률이 개선될 것으로 기대된다.

6. 참고문헌

[1] J. Ohya and Y. Kitamura, etc, "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, Fort Collins, Colorado, June, 23-25, 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] Hiroshi Murase and Shree K, Nayar, "Visual Learning and Recognition 3-D object from appearance", international journal of Computer Vision, Vol,14,1995.