

논문 2013-50-10-19

이중 링 마스크 저니키 모멘트를 이용한 손동작 인식

(Gesture Recognition Using Zernike Moments Masked By Duel Ring)

박 정 수*, 김 태 용**

(Jung-Su Park and Tae-Yong Kim[Ⓢ])

요 약

일반적으로 저니키 모멘트 값을 이용한 매칭 시에는 외접원 안에 속하는 이미지 정보를 Zernike 기저함수로 투영시켜 얻은 모멘트 값을 이용하여 매칭에 사용한다. 하지만 손 이미지의 특성상 무게중심 부근에 중복되는 정보가 많이 포함되는데 이로 인해 변별력이 떨어질 수 있는 문제점이 있다. 또한 중복되는 정보를 제외한 이미지 정보들 중에서 특정 영역에 있는 정보들은 손의 모양정보를 구분하는데 변별력을 높여줄 수 있다. 본 논문에서는 이러한 문제점을 해결하기 위해 중복되는 정보를 제거하는 R1 링 마스크를 통해 얻은 이미지와 변별력을 높여 줄 수 있는 정보에 가중치를 부여하는 R2 링 마스크를 통해 얻은 이미지를 결합하여 R3 링 마스크를 설계한다. R3 링 마스크를 이용하여 얻은 모멘트들을 주성분분석(PCA)을 통해 차원을 축소함으로써 매칭량을 감소시킨다. 제안한 방법의 우수성을 확인하기 위해 7가지 손 모양을 다른 방법과 비교실험 하였으며, 그 결과 우수성을 확인 할 수 있었다.

Abstract

Generally, when we apply zernike moments value for matching, we can use those moments value obtained from projecting image information under circumscribed circle to zernike basis function. However, the problem is that the power of discrimination can be reduced because hand images include lots of overlapped information due to its special characteristic. On the other hand, when distinguishing hand poses, information in specific area of image information except for overlapped information can increase the power of discrimination. In this paper, in order to solve problems like those, we design R3 ring mask by combining image obtained from R2 ring mask, which can weight information of the power of discrimination and image obtained from R1 ring mask, which eliminate the overlapped information. The moments which are obtained by R3 ring mask decrease operational time by reducing dimension through principle component analysis. In order to confirm the superiority of the suggested method, we conducted some experiments by comparing our method to other method using seven different hand poses.

Keywords : Hand gesture recognition, Hand pose recognition, Zernike moments, PCA

I. 서 론

* 학생회원, ** 평생회원, 중앙대학교 첨단영상대학원
(Graduate School of Advanced Image Science,
Multimedia, and Film, Chung-Ang University)

Ⓢ Corresponding Author(E-mail: kimty@cau.ac.kr)

※ 이 논문은 2013년도 정부(교육부)의 재원으로 한국
연구재단의 지원을 받아 수행된 기초연구사업임
(2013-009166).

접수일자: 2013년6월25일, 수정완료일: 2013년10월3일

현대 사회에서 많은 작업들이 컴퓨터를 이용하여 수
행됨에 따라 인간과 컴퓨터의 상호작용(HCI)에 대한 연
구가 활발히 진행되고 있다. 그 중에서도 시각에 기반
을 둔 동작 인식 기술은 다양한 환경에서 쉽고 편리하
게 사용할 수 있어 그 필요성이 증가하고 있다. 그 예로

기존의 입력 장치인 키보드와 마우스를 사용하기 힘든 장애인을 위한 인터페이스를 제공 할 수 있고 게임 및 가상현실 등에도 활용 될 수 있다. 다양한 시각 기반의 동작 인식 중에서도 손동작은 특히 표현력이 뛰어나기 때문에 널리 사용된다^[1].

시각 기반 인터페이스에서 손동작 인식을 위해서는 다음과 같은 문제점들을 해결하여야 한다. 첫째, 입력 영상으로부터 배경과 손 영역을 정확하게 분리할 수 있어야 한다. 하지만 실제 환경에서의 입력 영상은 조명의 밝기와 색상에 민감하여 배경과 손 영역을 정확하게 분리하기 어렵다. 그리고 기존의 많은 관련 연구들이 컬러공간 기반에서 손 영역을 분리하는 연구가 진행되었다^{[2][3]}. 둘째, 손의 모양 정보를 정확히 인식하여야 한다. 하지만 많은 관련연구들이 손의 모양 정보보다 손가락의 개수에 따른 손동작 인식에 초점을 맞췄다^[4]. 셋째, 손동작의 인식과정에서도 기타 환경이나 노이즈, 회전, 크기 등에 영향을 받지 않고 강건하게 인식할 수 있어야 한다. 기존의 방법들은 회전에 강인한 손 모양 인식은 하였지만 불변은 아니다^[5].

본 논문에서는 이러한 문제점을 개선하기 위하여 MS사의 Kinect^[6] 카메라로부터 받아온 depth 정보를 이용하여 손 영역을 분리 한다. 손 모양 인식에서는 회전에 불변하는 서술자인 저니키 모멘트를 이용한다^[7].

저니키 모멘트는 잡음 특성, 정보 중복도, 영상 표현 능력 등의 특성을 비교 분석한 결과에서 그 우수성이 증명되었다^[8]. 일반적으로 손 모양 정보에 대한 저니키 모멘트 값을 얻기 위해서는 외접원 안에 속하는 손 이미지 정보를 저니키 기저함수로 투영시켜 그 값을 얻는다. 하지만 손 이미지는 무게중심 부근으로부터 일정 영역에 속하는 정보가 중복된다. 따라서 이로 인해 변별력이 떨어질 수 있는 문제점이 있다. 이와 반대로 특정 영역에 있는 이미지 정보들은 손의 변별력을 높여 줄 수 있다. 본 논문에서는 이러한 중복되는 이미지 정보는 제거하고, 중요한 이미지 정보는 강조하는 마스크 링을 설계하였다.

본 논문의 구성은 다음과 같다. II장에서는 저니키 모멘트와 주성분분석에 대해 소개하고 III장에서는 depth 정보를 이용하여 손 영역을 분리하는 방법에 대해 소개한다. IV장에서는 제안한 이중 링 마스크 저니키 모멘트에 대하여 소개하고 V장에서는 실험 및 결과에 대해 소개한다. VI장에서는 끝으로 결론을 맺는다.

II. 저니키 모멘트와 주성분분석

1. 저니키 모멘트(Zernike Moments)

(1) 저니키 모멘트 정의

저니키 모멘트는 회전(rotation)에 불변하는 특징을 가지고 있으며, 정규화를 통하여 크기(scale)와 이동(translation)에 불변하게 만들 수 있다. 또한 영상의 노이즈에 강인하며 적은 개수의 모멘트로 영상을 효율적으로 표현할 수 있기 때문에 패턴인식 분야에서 널리 사용되어 왔다. 저니키 모멘트는 입력 영상을 기저함수로 투영하는 것으로 간주될 수 있으며, 차수(order)가 n 이고 반복수(repetition)가 m 인 저니키 모멘트의 기저함수는 다음 식 (1)과 같이 정의된다.

$$V_{nm}(x, y) = V_{nm}(\rho, \theta) = R_{nm}(\rho) \exp(jm\theta) \quad (1)$$

여기서 V_{nm} 는 극좌표(polar coordinates) 공간에서 단위 원(unit circle)안에 직교하는 다항식의 집합이며 정보의 중복이 없다. 따라서 각각의 차수와 반복수는 영상의 고유특성을 나타내는 것을 의미한다. 또한 다음의 규칙을 가지는데 n 은 0또는 양의 정수이고, m 은 $n - |m|$ 이 짝수이며 $|m| \leq n$ 을 동시에 만족하는 음수가 아닌 정수이다. ρ 는 원점에서 점 (x, y) 까지의 거리이며, $0 \leq \rho \leq 1$ 의 범위에서 유효하다. θ 는 점 (x, y) 가 x 축과 이루는 각의 크기를 나타내며, $0 \leq \theta \leq 2\pi$ 의 범위에서 유효하다.

$R_{nm}(\rho)$ 는 저니키 실수 방사 다항식(radial polynomial)이며 다음 식 (2)와 같이 정의 된다.

$$R_{nm}(\rho) = \sum_{s=0}^{(n-|m|)/2} (-1)^s \frac{(n-s)!}{s! \left(\frac{n+|m|}{2} - s\right)! \left(\frac{n-|m|}{2} - s\right)!} \rho^{n-2s} \quad (2)$$

차수가 n 이고 반복수가 m 인 저니키 모멘트 Z_{nm} 은 다음 식 (3)과 같이 정의된다.

$$Z_{nm} = \frac{n+1}{\pi} \iint_{x^2+y^2 \leq 1} f(x, y) V_{nm}^*(x, y) dx dy \quad (3)$$

여기서 V^* 는 복소 공액(complex conjugate)을 나타낸다. 실제 입력 영상에서 저니키 모멘트를 구하려면 구하

고자 하는 영상의 무게 중심으로부터 영상의 모든 정보를 포함하는 외접원을 구하고, 구해진 외접원안의 이미지를 기저함수의 반경과 동일하도록 크기와 이동에 대해 정규화(normalization) 시켜야 한다. 이산 영상에 대하여 저니키 모멘트를 구하기 위해서, 식 (3)을 식 (4)와 같이 저니키 모멘트 식을 근사화 할 수 있다.

$$Z_{nm} = \frac{n+1}{\pi} \sum_x \sum_y f(x,y) V_{nm}^*(x,y), x^2 + y^2 \leq 1 \quad (4)$$

(2) 저니키 모멘트 회전 불변 특성

저니키 모멘트의 가장 큰 장점은 모멘트들의 절대 값은 회전에 불변하다는 사실이다^[7]. 이러한 회전에 불변하는 특성은 다음과 같이 유도될 수 있다. 영상 $f(x,y)$ 를 극좌표계 $f(\rho,\theta)$ 로 변환 하였을 때, 영상이 α 만큼 회전된 영상은 다음 식 (5)와 같이 정의 된다.

$$f^r(\rho,\theta) = f(\rho,\theta + \alpha) \quad (5)$$

위의 수식 (5)를 수식 (3)에 적용하면 식(6)과 같다.

$$Z_{nm}^r = Z_{nm} \exp(jm\alpha) \quad (6)$$

위의 수식 (6)이 성립하므로 회전된 영상은 위상 값에만 영향을 주고 그 절댓값은 변하지 않는다. 이를 수식 (7)로 나타내면 아래와 같이 쓸 수 있다.

$$|Z_{nm}^r| = |Z_{nm}| \quad (7)$$

따라서 저니키 모멘트의 절댓값을 특징 값으로 사용하는 것은 회전에 불변하다는 것을 의미한다.

2. 주성분분석(Principle Component Analysis)

주성분분석(PCA)은 고차원 공간상에 분포하는 데이터들의 주축을 찾아주는 역할을 한다. 이는 적은 수의 파라미터를 이용하여 데이터의 정보를 표현하는데 유용한 방법이다^[9].

주성분분석을 통해 차원을 축소하는 알고리즘은 다음과 같다. 먼저 데이터들의 평균을 구하고 평균값을 이용하여 공분산 행렬을 계산한다. 그 후 공분산행렬의 고유값(eigenvalue)과 고유벡터(eigenvector)를 구한 후

고유값이 큰 순서대로 재 정렬한다. 이때 각각의 고유값은 고유값이 해당되는 고유벡터 방향으로 축을 변화 하였을 때 분산 값을 의미한다. 그러므로 데이터의 전체 분산은 식 (8)과 같이 구할 수 있다.

$$V_N = \sum_{i=1}^n \lambda_i \quad (8)$$

즉 전체 분산 V_N 는 전체 n 개의 고유값을 모두 더한 값이다. 앞에서 구한 고유값, 고유벡터, 전체 분산 값을 가지고 차원을 축소할 수 있으며 식 (9)과 같다.

$$\sum_{i=1}^p \lambda_i \geq f_v V_N \quad (9)$$

여기서 p 는 사용하고자 하는 특징 벡터의 개수이며, f_v 는 표현하고 싶은 데이터의 양을 결정하는 비례상수이다. 만약 $f_v = 0.95$ 라면 원래의 정보의 95%를 표현 할 수 있으며, 원래의 정보를 잘 표현 한다고 할 수 있다^[10].

III. Depth 정보를 이용한 손 영역 추출

기존 Color-Based 방식의 손동작 인식 방법은 복잡한 배경이나 조명변화에 의한 왜곡된 정보를 가져오는 문제점이 있다. 또한 대부분의 기존 손동작 인식 방식은 근거리에서는 높은 인식률을 보이지만 카메라와의 거리가 멀어질수록 인식률이 현저히 떨어지는 문제점이 있다. 이러한 문제점을 본 논문에서는 Depth 정보를 이용하여 해결한다.

1. 손 영역 추출

손 영역 추출 단계에서는 MS사의 Kinect[6] 카메라로부터 받아들인 depth 정보를 이용하여 손 영역 추출, 노이즈 제거, 무게중심 추출, 크기와 이동에 대해 정규화(normalization) 작업이 진행된다.

그림 1의 왼쪽은 Kinect 카메라로부터 받아들인 입력 영상의 depth 정보를 나타낸다. 그림 1의 오른쪽은 받아들인 depth 정보를 이용하여 밝기 값과 누적 빈도수에 따른 히스토그램으로 변환한 정보를 나타낸다.

히스토그램 정보에서 화살표로 나타낸 부분은 최빈 값을 나타내며 사람의 신체부분에 해당한다. 여기서 손

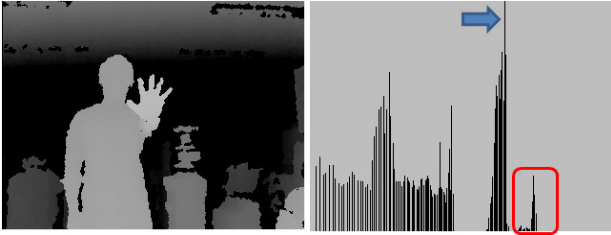


그림 1. dpeth 카메라로부터 받아온 입력 영상과 거리 정보에 대한 히스토그램
Fig. 1. Input Image from depth camera and Histogram about depth information.



그림 2. 히스토그램 최빈값의 임계비례상수에 따른 손 영역 추출
Fig. 2. Hand region extraction according to threshold of histogram mode.

영역에 해당하는 부분은 최빈값을 기준으로 좀 더 밝은 값, 즉 좀 더 가까운 영역에 있는 부분이다. 그러므로 그림 1에서 손 영역에 해당하는 부분은 빨간색 네모 영역이 된다. 따라서 손 영역을 효과적으로 추출하는 방법은 식 (10)과 같다.

$$P(x,y) = \begin{cases} 1 & D(x,y) > D_{mode} * T \\ 0 & otherwise \end{cases} \quad (10)$$

여기서 $D(x,y)$ 는 거리정보에 대한 값, D_{mode} 는 히스토그램의 최빈값을 나타내며, T 는 임계비례상수를 나타낸다. 따라서 식 (10)의 위의 식을 만족하는 $P(x,y)$ 값이 1인 부분이 손 영역에 해당하는 부분이 된다. 여기서 두 가지 제약사항이 있다. 첫째, 사람의 손은 사람의 신체보다 앞에 위치해 있어야 한다. 둘째, 사람의 신체보다 큰 다른 객체는 없다고 가정한다.

그림 2는 임계비례상수에 따른 손 영역 추출결과를 보여주며 0.1의 임계비례상수를 사용할 때 가장 깔끔하게 손 영역을 분리하는 것을 알 수 있었다. 따라서 본 논문에서는 0.1의 임계비례상수를 사용한다. 히스토그램을 이용한 손 영역 추출은 거리정보를 이용하기 때문에 동적으로 손 영역 추출이 가능하다.

손 영역을 추출한 다음 영상에 포함되어 있는 불필



그림 3. 정규화 전(좌)과 정규화 후(우)의 손 이미지
Fig. 3. Hand image before and after normalization.

요한 노이즈 성분을 제거한다. depth 값이 많이 튀고 노이즈 성분이 다소 포함되어 있는데, 이는 kinect 카메라가 정밀성을 위해 제작된 카메라가 아니기 때문이다. 노이즈 성분 제거를 위해 중간 값 필터링을 사용하였다.

노이즈를 제거한 다음 크기와 이동에 불변하는 정규화를 위한 무게중심을 추출한다. 무게중심을 구한 후 모든 이미지에 대해 공정한 분류를 위해 정규화(normalization) 과정이 진행된다. 이미지를 저니키 모멘트의 단위원 안의 기저함수로 투영시키므로 모든 이미지 정보를 포함하는 동일한 외접원 크기로 정규화시켜야 한다. 외접원을 구하는 방법은 무게 중심으로부터 가장 멀리 떨어진 픽셀까지의 거리 값을 정해진 크기로 동일하게 정규화 시켜주면 된다. 무게중심에서 가장 멀리 떨어진 픽셀의 거리 값을 구하는 수식은 식 (11)과 같다.

$$D_{max} = \arg \min_D \left\{ D = \sqrt{(C_x - i_x)^2 + (C_y - i_y)^2} \right\} \quad (11)$$

여기서 C_x 와 C_y 는 입력된 영상의 무게중심 좌표이며, i_x 와 i_y 는 외곽선 픽셀의 x, y 좌표 값이다. 모든 입력 이미지를 동일한 크기의 단위원(Unit Circle)으로 정규화할 수 있으며 수식은 식 (12)와 같다.

$$U = \frac{S_I / 2}{D_{max}} * \frac{S_N}{2} \quad (12)$$

여기서 S_I 는 입력된 이미지의 크기이며 S_N 은 정규화하고자 하는 이미지의 크기이다.

그림 3은 정규화 전과 정규화 후의 입력된 손의 일부 이미지를 나타내며, 정규화 된 손 이미지는 저니키 기저함수로 투영하기 위한 입력 값이 된다. 정규화 후의 이미지를 저니키 모멘트로 표현하면 크기(scale), 이동(translation), 회전(rotation)에 불변이다.

2. 손 추적

손 영역 추출 과정에서 획득한 정규화 된 손 이미지와 손의 무게중심을 이용하여 제스처를 실시간으로 인식하기 위해서는 손 추적 과정이 필요하다. 본 논문에서는 손의 동선 정보를 이용한 제스처 인식보다는 손의 포즈 정보를 이용한 제스처 인식에 초점을 맞췄기 때문에 제스처 인식을 위해 손의 무게중심 정보의 추적만 필요하다. 따라서 기존의 추적 알고리즘들을 사용하는 대신, 실시간으로 입력되는 정규화 된 이미지의 무게중심 좌표 값을 단위원 안에서 원점으로 하는 이미지를 저니키 모멘트의 기저함수로 투영하는 방법을 사용했다.

IV. 이중 링 마스크 저니키 모멘트

1. R1 링 마스크를 통한 중복 부분제거

기존의 저니키 모멘트는 입력된 전체 모양정보에 대해 얻은 모멘트 값을 척도로 하는 분류방법을 사용했다. 하지만 손 이미지의 특성상 무게중심을 기준으로 일정 범위 내에 속하는 많은 이미지 정보들이 중복된다. 따라서 본 논문에서는 손의 모양정보를 구별하는 변별력을 높이기 위해 중복되는 정보를 제거하는 R1 링 마스크 방법을 사용한다. 따라서 R1 링 마스크의 크기는 이미지 내접원의 크기보다 작아야 하며 내접원의 반지름 크기를 구하는 수식은 식(13)과 같다.

$$D_{\min} = \arg \min_D \left\{ D = \sqrt{(C_x - i_x)^2 + (C_y - i_y)^2} \right\} \quad (13)$$

여기서 입력이미지는 정규화 과정에서와 마찬가지로 외곽선 이미지를 사용하며 식 (11)을 계산할 때 같이 계산한다. C_x 와 C_y 는 입력된 영상의 무게중심 좌표이며, i_x 와 i_y 는 외곽선 픽셀의 x, y 좌표 값이다. 앞의 식 (13)에서 얻은 값이 R1 링 마스크의 외부 크기가 된다. 내부 크기의 시작점은 외부 크기를 기준으로 단위원 U 크기의 0.25 비율 작은 값이 된다. 여기에서 원을 사용하지 않고 링을 사용한 이유는 같은 포즈를 취하는 사람에 따라 내접원의 크기가 달라지기 때문이다. 원을 사용하게 되면 내접원이 커질수록 제거되는 정보량 또한 커지게 되어 공정성이 없어지지만 링을 사용하면 좀 더 공정한 정보량을 제거할 수 있다. R1 링 마스크가 적용된 손 이미지



그림 4. R1 링 마스크가 적용된 손 이미지
Fig. 4. Hand image applied to R1 ring mask.



그림 5. 예외 처리된 R1 링 마스크가 적용된 손 이미지
Fig. 5. Exceptional Hand image pose applied to R1 ring mask.

의 샘플을 그림 4에 나타내었다.

R1 링 마스크가 일정비율 보다 클 경우 R2 링 마스크의 범위에 포함되게 되므로 예외 처리가 필요하며 수식은 식 (14)와 같다.

$$R1(x, y) = \begin{cases} 1 & D_{\min} < U * 0.5 \\ 0 & otherwise \end{cases} \quad (14)$$

여기서 단위원 U 의 0.5비율보다 작은 값이 R1 링 마스크의 영역이 되는데 이후에 설명할 R2 링 마스크와 범위가 중복되기 때문이다.

주먹을 쥔 포즈 같은 경우 내접원의 크기가 단위원 U 의 0.5 비율을 넘어 버리게 된다. 이런 경우 R1 링 마스크와 R2 링 마스크의 경계부분인 단위원 U 의 0.375 - 0.625비율에 속하는 정보를 사용한다. 예외 처리된 이미지의 샘플을 그림 5에 나타내었다.

2. R2 링 마스크를 통한 가중치 부여

손의 특성상 많은 정보가 중복되는 부분인 R1 링 마스크에 해당하는 정보를 이전 절에서 제거하였다. R2 링 마스크는 손의 포즈별 변별력을 높여줄 수 있는 부분에 대해 가중치를 부여하는 방법이다. 본 논문에서 R2 링 마스크의 작은 경계의 크기를 단위원 U 의 0.5비율에 해당하는 기준으로 정하였는데, 그 이유는 손의 기하학적 특성상 단위원 U 의 약 0.44비율 부근에서 포즈의 공정성이 교차하는 것을 발견하였기 때문이다.

이는 그림 6에 나타내었으며 가시성을 높이기 위해 손 이미지를 그레이 색상으로 나타내었다.

그림 6의 좌측 위의 이미지와 두 번째, 세 번째 영상

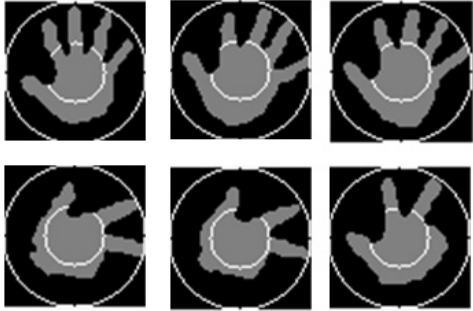


그림 6. R2 링 마스크의 내부 경계부분
Fig. 6. Inner boundary area for R2 ring mask.



그림 7. R2 링 마스크의 외부 경계부분
Fig. 7. Outer boundary area for R2 ring mask.



그림 8. R2 링 마스크가 적용된 손 이미지
Fig. 8. Hand image applied to R2 ring mask.

과 비교해 볼 때 손가락 마디 부분에서 경계선이 발생 하는 것을 볼 수 있다. 이는 사람마다 손가락의 길이 가 다르기 때문인데, 따라서 본 논문에서는 R2 링 마스크 의 내부 경계부분을 단위원 U 의 0.5 비율로 정하였다.

내부 경계와 비슷한 이유로 R2 링 마스크의 외부 경 계부분은 단위원 U 의 0.75 비율로 정하였다. 이는 [11] 에서 엄지손가락의 경우 나머지 4개의 손가락보다 가동 범위가 크기 때문이다. R2 링 마스크의 경계는 단위원 U 의 약 0.84 비율의 부근에서 발견하였으며 이에 대한 샘플영상을 그림 7에 나타내었다.

따라서 R2 링 마스크는 단위원 U 의 0.5 - 0.75 비율 범위에 해당하는 정보 값을 통과 시키는 링으로 설계된 다. R2 링 마스크가 적용된 손 이미지 샘플을 그림 8에 나타내었다.

R2 링 마스크에서 또한 예외가 존재하는데 수식(14) 에서 예외처리가 적용된 R1 링 마스크는 R2 링 마스크

도 동일한 이미지 정보를 사용한다. 이는 R1 링 마스크 와 R2 링 마스크의 경계부분이기 때문이다.

3. 이중 링 마스크 저니키 모멘트

앞에서 얻은 R1 링 마스크와 R2 링 마스크가 적용된 이미지를 매칭에 사용하게 되면 기존 저니키 모멘트 보 다 적은 차수의 사용만으로도 인식률이 우수하다. 하지 만 동일한 차수의 저니키 모멘트와 비교해 볼 때 2배의 연산량이 필요하다. 따라서 R1 링 마스크와 R2 링 마 스크를 결합하여 R3 링 마스크를 설계한다.

R3 링 마스크에 대한 이미지는 R1 링 마스크가 적용 된 R1 이미지와 R2 링 마스크가 적용된 R2 이미지를 더한 값이며 수식은 식 (15)와 같다.

$$R3 = R1 * w1 + R2 * w2 \tag{15}$$

여기서 $w1$ 은 R1 링 마스크가 적용된 이미지에 대한 가 중치 변수, $w2$ 는 R2 링 마스크가 적용된 이미지에 대한 가중치 변수이다. 본 논문에서는 각 0.5의 가중치를 사용 하였다. R3 링 마스크가 적용된 손 이미지의 샘플을 그림 9에 나타내었다.

R3 링 마스크를 통해 얻은 이미지를 저니키 기저함 수로 투영시켜 얻은 모멘트 값을 매칭에 사용한 결과와 R1 링 마스크를 통해 얻은 이미지와 R2 링 마스크를 통해 얻은 이미지를 기저함수로 투영시켜 얻은 모멘트 값의 합을 매칭에 사용한 결과는 동일하다.



그림 9. R3 링 마스크가 적용된 손 이미지
Fig. 9. Hand image applied to R3 ring mask.

V. 실험 및 분석

본 논문에서는 제안한 이중 링 마스크 저니키 모멘트 를 이용한 제스처 인식 방법의 성능을 분석하기 위해 7 개 포즈에 대해 비교 실험을 하였다. 또한 기본 저니키 모멘트와 기존의 Inner, Outer 방법[12]과 제안한 이중 링 마스크 저니키 모멘트 방법과 비교 실험을 하였다.

데이터 셋은 7개 포즈에 대해 각 포즈별 50장씩 총



그림 10. 정규화 되지 않은 7개의 포즈별 샘플
Fig. 10. Unnormalized seven different poses sample.



그림 11. 정규화된 7개의 포즈별 샘플(실험에서는 왼쪽
위부터 오른쪽으로 포즈 1 - 4, 왼쪽 아래부터
오른쪽으로 포즈 5 - 7번으로 지정)
Fig. 11. Normalized seven different poses sample(from
upper left to right pose 1 - 4, from lower left to
right pose 5 - 7).

350장의 이미지를 실험에 사용하였다. 그림 10은 실험에 사용한 정규화 되지 않은 각 포즈별 샘플을 보여준다. 앞 장에서 언급하였듯이 저니키 모멘트는 회전에만 불변이기 때문에 이동과 크기에 정규화 된 데이터 셋을 훈련 이미지와 테스트 이미지로 사용한다. 따라서 그림 10의 샘플들을 매칭에 이용하기 위해서는 정규화 과정을 거쳐야 한다. 그림 11은 65*65 크기로 정규화 과정을 거친 7개의 포즈에 대한 샘플을 보여준다.

매칭 방법은 유클리디언 거리값을 사용하였으며, 입력된 이미지와 각 포즈별 7개의 참조이미지들과의 매칭을 통해 거리값이 가장 작은 포즈로 분류된다. 입력된 이미지의 거리값을 구하는 수식은 식 (16)과 같다.

$$dist(R, I) = \sqrt{\left(\sum_{i=1}^k (R_i - I_i)^2 \right)} \quad (16)$$

여기에서 R 은 포즈별 참조 이미지, I 는 입력된 이미지이며 k 는 매칭에 사용된 전체 모멘트 개수이며 i 는 i 번째 모멘트 값이다.

1. 기본 저니키 모멘트와의 비교 실험

그림 12는 기본 저니키 모멘트의 인식 성능을 5 - 15 차 까지 나타낸 그림이다. 그림 12를 보면 차수가 증가할수록 절대적이지는 않지만 인식률이 점차 높아지는 것을 볼 수 있다. 하지만 저니키 모멘트의 기저함수 정의식이 매우 복잡하기 때문에 차수가 증가할수록 필요한 연산량 또한 기하급수적으로 증가한다. 따라서 기존에 많은 논문에서 연산시간을 줄이기 위한 많은 연구들이 있었고, 본 논문에서는 [13]에서 사용한 q-recursive 방법을 고속 연산을 위해 실험에 사용하였다.

기본 저니키 모멘트와 q-recursive 방법의 5 - 10차 까지 평균 연산시간을 표 1에 나타내었다.

본 논문에서는 그림 12의 모멘트 차수별 인식정확도와 표 3의 평균연산시간을 고려하여 실시간 손동작 인식을 위해 8차 저니키 모멘트를 사용하여 기존의 방법

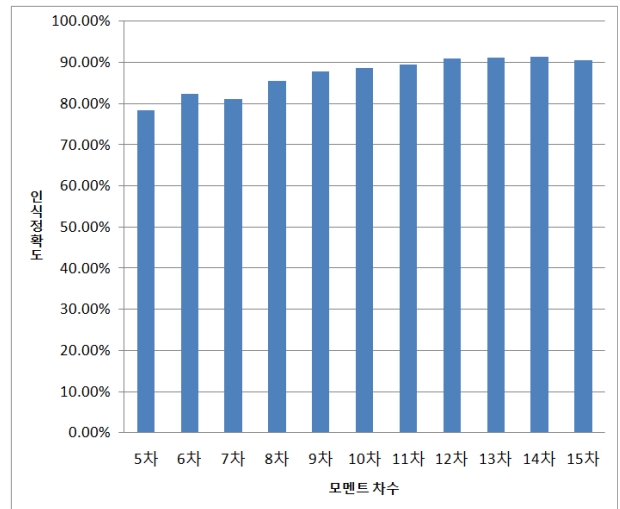


그림 12. 저니키 모멘트 차수별 평균 인식정확도
Fig. 12. Zernike moments average recognition accuracy rate per order.

표 1. 저니키 모멘트 차수별 평균연산시간
Table 1. Zernike moments average execution time per order.

차수(n)	연산시간(ms)	
	Direct method	q-recursive[13]
5	101.36	16.18
6	157.2	22.14
7	199.34	28.38
8	253.44	36.4
9	339.08	45.14
10	421.52	54.18

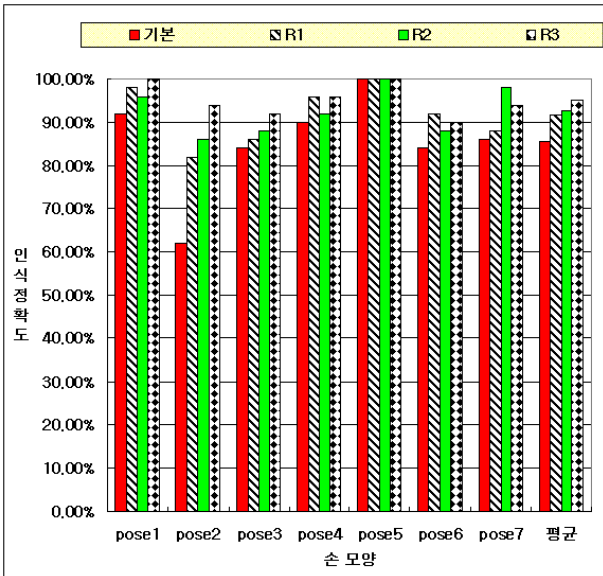


그림 13. 제안된 방법과의 인식정확도 비교
Fig. 13. Recognition accuracy rate compare to proposed method.

과 제안된 방법을 비교 실험하였다.

그림 13의 도표가 나타내는 실험 결과는 다음과 같다. 먼저 전체 평균 인식률을 보면 각각 기본 이미지 85.43%, R1 링 마스크를 적용한 이미지 91.71%, R2 링 마스크를 적용한 이미지 92.57%, R3 링 마스크를 적용한 이미지 95.14% 이다. 손의 기하학적 특성상 무게 중심 부근에 중복되는 정보가 많이 있었는데, 이를 보완하기 위한 R1 링 마스크를 통하여 얻은 이미지의 경우 기본 이미지를 저니키 기저함수로 투영한 결과보다 전체 평균 약 6.29%의 인식률의 향상이 있었다. 특히 포즈 2번의 경우 약 20%의 인식률 향상이 있었다. 이는 포즈 2번을 1번으로 오인식 하는 경우가 많았는데 중복되는 정보를 제거하여 변별력을 높인 결과이다.

R2 링 마스크는 변별력을 높여줄 수 있는 영역에 가중치는 부여하는 방법이다. R2 링 마스크를 통하여 얻은 이미지의 경우 기본 이미지를 저니키 기저함수로 투영한 이미지와 비교하였을 때 전체 평균 약 7.14%의 인식률 향상이 있었다. R2 링 마스크도 R1 링 마스크와 비슷하게 포즈 2번에서 약 24%의 인식률 향상이 있었다. 이는 제안한 R2 링 마스크의 영역에 있는 정보가 포즈 1번과의 변별력을 높여주기 때문이다. R3 링 마스크의 경우 기본 저니키 모멘트와 비교하였을 때 전체 평균 약 10%의 인식률 향상이 있었다. 이는 R1 링 마스크를 통해 중복되는 정보를 제거하고, R2

링 마스크를 통해 변별력이 높은 정보에 가중치를 부여한 결과이다.

2. 기존 Inner, Outer 방법과의 비교 실험

본 논문에서 제안한 이중 링 마스크 저니키 모멘트의 우수성을 검증하기 위하여 기존의 Inner, Outer Circle 방법^[12]과의 비교 실험을 하였다.

그림 14의 도표는 본 논문에서 제안한 방법과 비교 실험에 사용된 기존 방법과의 인식 성능 비교 결과를 나타낸다. Inner, Outer Circle 방법은 이미지 정보를 내부정보와 외부정보로 영역을 분할하여 각각의 영역에 가중치를 적절히 적용하여 인식성능을 향상시키는 방법이다^[12]. 본 논문에서는 내부와 외부 영역의 경계를 0.5 비율로 정하여 실험하였다. 전체 평균의 인식률을 볼 때 Inner Circle 영역의 인식률이 상대적으로 낮은 것을 볼 수 있다. 이것은 손의 기하학적 특성상 무게중심 부근에 중복되는 정보가 많이 물려있는 것을 증명할 수 있다.

Outer Circle 방법의 경우 Inner Circle 방법과 비교해 볼 때 인식률이 약 21% 더 우수한 것을 볼 수 있다. 이는 사람의 지각이 모양의 외곽정보를 더 중요하게 생각하는 것과 상관관계가 있다^[12]. R2 링 마스크 방법의 경우 단독으로 사용하였을 때 Outer Circle 방법과 인식률의 차이가 미비하지만 R1 링 마스크와 조합하여 R3

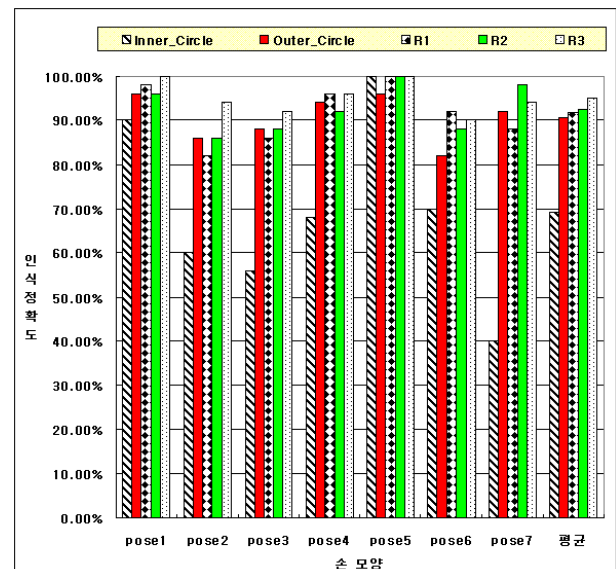


그림 14. 기존 방법과의 인식정확도 비교
Fig. 14. Recognition accuracy rate compare to existing method.

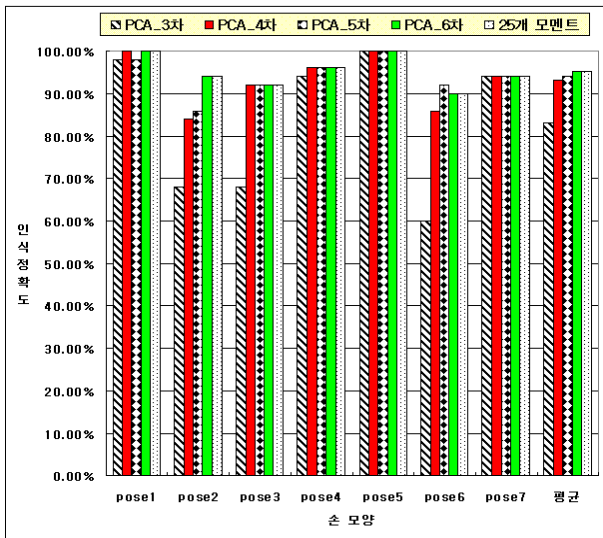


그림 15. 전체 모멘트와 차원 축소된 모멘트의 인식정확도 비교

Fig. 15. Recognition accuracy rate compare to total moments and moments which are reduced dimension.

링 마스크를 만들어 사용할 경우 인식률의 폭이 약 5%로 조금 더 커진 것을 볼 수 있다.

주성분분석(Principle Component Analysis)을 통해 차원이 축소된 모멘트 정보의 인식정확도를 비교 실험하였다. 식 (9)의 고유값의 비를 이용하여 특징벡터의 수를 결정할 수 있다. 본 논문에서 사용한 이중 링 마스크 이미지의 8차 저니키 모멘트에 대한 모멘트 개수는 25개이다. 이 값들을 주성분분석을 통해 6차원부터 3차원까지 축소할 경우 각 100%, 96.9%, 92.4%, 87.1%의 정보를 표현할 수 있다.

그림 15는 R3 링 마스크를 통해 얻은 이미지를 8차원 저니키 모멘트로 투영하여 손 모양을 분류한 인식정확도와, 주성분 분석을 통해 6차원에서 3차원까지 축소한 인식정확도를 나타낸다. 25개의 모멘트 정보를 손 모양 분류에 사용한 것과 주성분 분석을 통해 6차원으로 축소한 후 손 모양 분류에 사용한 인식정확도는 동일하다. 또한 차원의 수를 5차원과 4차원으로 축소할 경우에도, 각각 94%, 93.14%의 높은 인식률을 보장할 수 있다.

따라서 25개의 모멘트 정보를 모두 사용하지 않고 주성분분석을 통해 인식률과의 상관관계를 고려하여 적절한 차원으로 축소하여 사용할 경우, 손 모양을 분류할 때 드는 매칭 횟수를 감소시킬 수 있다.

VI. 결 론

본 논문에서는 이중 링 마스크 저니키 모멘트를 이용한 손동작 인식 방법에 대해 제안하였다. 제안한 방법은 손의 기하학적 특성상 무게 중심 부근의 중복된 정보가 많은 부분을 제거하는 R1 링 마스크와 손의 변별력을 높여줄 수 있는 대역폭에 가중치를 부가하는 R2 링 마스크의 장점을 조합하여 R3 링 마스크를 설계하는 것이었다.

제안한 방법의 우수성을 확인하기 위해 7가지 손 모양을 기본 저니키 모멘트 방법과 기존 논문과의 비교 실험 하였다. 그 결과 기본 저니키 모멘트보다 약 10%의 인식정확도 향상이 있었다. 또한 기존의 Outer Circle 방법과의 비교에서도 약 5%의 인식정확도 향상이 있었다.

제안한 이중 링 마스크를 통해 얻은 이미지의 모멘트 값들을 주성분분석(PCA)을 통해 차원을 축소하여 매칭 횟수를 감소시킬 수 있었다. R1 링 마스크와 R2 링 마스크 중 중요한 정보에 더 큰 가중치를 부여할 수 있다면 향후 연구에서 더 높은 인식률을 가져올 것으로 보인다.

REFERENCES

- [1] R. Watson, "A survey of Gesture Recognition Techniques," Technical Report, TCD-CS-1993-11, pp. 1-31, 1993.
- [2] K. W. Kim, W. J. Lee, C. H. Jeon, "Dynamic Hand Gesture Recognition using Guide Lines," The Institute of Electronics Engineers of Korea, vol. 47, no. 5, pp. 469-477, Sep 2010.
- [3] J. H. Yun, C. H. Lee, "Design of Computer Vision Interface by Recognizing Hand Motion," The Institute of Electronics Engineers of Korea, vol. 47, no. 3, pp. 256-265, May 2010.
- [4] N. R. Shin, B. G. Heo, K. C. Hong, "Development of Hand Gesture Recognition Technic based on Image's Skin Color Information," Korea Computer Congress, vol. 37, no. 1, pp. 484-489, 2010.
- [5] S. I. Lee, H. S. Min, H. C. Shin, E. G. Lim, D. H. Hwang, Y. M. No, "Hand posture recognition robust to rotation using temporal correlation between adjacent frames," Journal of Korea Multimedia Society, vol. 13, no. 11, pp.

1630-1642, Nov 2010.

- [6] <http://en.wikipedia.org/wiki/Kinect>
- [7] A. Khotanzad, Y. H. Hong, "Invariant image recognition by zernike moments," IEEE Transactions on Pattern analysis and Machine Intelligence, vol. 12, no. 5, pp. 489-497, May 1990.
- [8] C. H. The, R T. Chin, "On Image Analysis by the methods of moments," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 10, no.4, pp. 496-513, Jul 1988.
- [9] M. Kirby, L. Sirovich, "Application of the Karhunen-Loeve procedure for the characterization of human faces," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 12, no. 1, pp. 103-108, Jan 1990.
- [10] D. Swets, "Using discriminant eigenfeatures for image retrieval," IEEE Transaction on Pattern Analysis and Machine Intelligence vol. 18, no. 8, pp. 831-836, Aug 1996.
- [11] H. I. Suk, J. H. Lee, S. W. Lee, "Real-Time Hand Pose Tracking and Finger Action Recognition Based on 3D Hand Modeling," Journal of KIISE, vol. 35, no. 12, pp. 780-788, Dec 2008.
- [12] J. D. Kim, H. G. Kim, "Zernike Moments Shape Descriptor with Region Partitioning," The Korean Society of Broadcast Engineers, pp. 53-57, Nov 1999.
- [13] C. W. Chong, P. Raveendran, R. Mukundan, "A comparative analysis of algorithms for fast computation of Zernike moments," Pattern Recognition, vol. 36, no. 3, pp. 731-742, March 2003.

저 자 소 개



박 정 수(학생회원)
2011년 경일대학교 컴퓨터공학과
학사 졸업.
2013년 중앙대학교 첨단영상대학
원 영상학과 석사 졸업.

<주관심분야 : 컴퓨터 비전, 패턴인식, 제스처 인식>



김 태 용(평생회원)
1986년 한양대학교 전기공학과
학사 졸업.
1988년 한양대학교 전자통신공학
과 석사 졸업.
1998년 포항공과대학교 전자계
산학과 박사 졸업.

<주관심분야 : 컴퓨터 비전, 영상처리, 컴퓨터 게임>