논문 2010-47CI-1-11

다양한 환경에 강건한 DSTW 기반의 동적 손동작 인식

(A Robust Method for the Recognition of Dynamic Hand Gestures based on DSTW)

지 재 영*, 장 경 현*, 이 정 호*, 문 영 식**

(Jae Young Ji, Kyung Hyun Jang, Jeong Ho Lee, and Young Shik Moon)

요 약

본 논문에서는 Dynamic Space Time Warping(DSTW) 알고리즘을 이용하여 손동작을 다양한 배경에서도 정확하게 인식할 수 있는 방법을 제안한다. DSTW 알고리즘을 이용한 기존의 손동작 인식 방법은 질의 영상의 매 프레임마다 검출된 다수의 손 후보 영역과 모델 영상을 시간 축 상으로 비교하는 방법이다. 그러나 DSTW 알고리즘을 이용한 기존의 손동작 인식 방법은 손을 포함하지 않은 후보 영역들(배경, 팔꿈치 등)에 의해 오 인식될 수 있는 경로를 생성하며, 그 결과로 사용자가 의도하지 않은 손동작으로 인식된다. 이러한 단점을 해결하기 위해서, 본 논문에서는 손 후보 영역의 불변 모멘트를 이용하여 질감 정보를 추출한 후 후보 영역들 사이의 유사도를 비교한다. 제안한 방법을 통해 계산된 유사도는 모델 영상과 질의 영상의 매칭 비용에 가중치로 적용된다. 실험 결과를 통해 제안한 방법은 다양한 배경에서도 사용자의 손동작을 정확하게 인식하였으며 기존의 방법에 비해 약 13%의 인식률이 향상한 것을 확인하였다.

Abstract

In this paper, a method for the recognition of dynamic hand gestures in various backgrounds using Dynamic Space Time Warping(DSTW) algorithm is proposed. The existing method using DSTW algorithm compares multiple candidate hand regions detected from every frame of the query sequence with the model sequences in terms of the time. However the existing method can not exactly recognize the models because a false path can be generated from the candidates including not-hand regions such as background, elbow, and so on. In order to solve this problem, in this paper, we use the invariant moments extracted from the candidate regions of hand and compare the similarity of invariant moments among candidate regions. The similarity is utilized as a weight and the corresponding value is applied to the matching cost between the model sequence and the query sequence. Experimental results have shown that the proposed method can recognize the dynamic hand gestures in the various backgrounds. Moreover, the recognition rate has been improved by 13%, compared with the existing method.

Keywords : Hand Gesture Recognition, Invariant Moment, DSTW Algorithm

I. 서 론

인간은 일상생활에서 말이나 문자와 같은 언어적 수단 뿐 만 아니라 표정, 제스처와 같은 비 언어적 수단을 이용하여 상대방과 의사소통을 한다. 그러나 서로 간의 인터페이스가 다를 경우, 문제가 발생한다. 따라서 사람과 컴퓨터 간의 보다 효과적인 상호작용을 하기 위해서는 두 개체 간의 의사를 잘 이해할 수 있는 편리하고 자연스러운 인터페이스가 요구된다[1].

제스처 기반 사용자 인터페이스는 사용자의 움직임을 통해서 보다 편안하고 자연스러운 상호작용을 제공한다[2~4]. 자연스러운 상호작용에 있어서 제스처 기반 사용자 인터페이스는 다른 인터페이스(음성, 촉감, 시점

* 학생회원, ** 평생회원-교신저자, 한양대학교 컴퓨터공학과
(Dept. of Computer Science and Engineering, Hanyang University)

※ 이 논문은 2009년 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (2009-0077434)

접수일자: 2009년12월16일, 수정완료일: 2010년1월11일

등)에 비해 비교적 직관적이며 간단하다. 그러나 단점으로는 데이터 글로브, 움직임 추적 장치 등과 같은 부가적인 장치를 사용해야 하는 불편함이 있다. 따라서 부가적인 장치의 사용에 따른 불편을 해소하고 사용자의 자연스러운 움직임을 보장하기 위해 컴퓨터 비전(Computer Vision) 기술을 이용한 제스처 기반 사용자 인터페이스에 대해 연구가 활발히 진행되고 있다[1]. 일반적으로 컴퓨터 비전 기술을 이용한 제스처 기반 사용자 인터페이스는 초기화(Initialization), 추적(Tracking), 포즈 예측(Pose Estimation), 그리고 인식(Recognition) 과정을 거쳐 획득한 영상으로부터 제스처를 인식한다. 그러나 추적 기반의 제스처 인식 방법은 객체의 갑작스런 움직임, 가려짐, 그리고 복잡한 배경이 존재하는 경우 객체를 정확하게 추적하기 어렵다. 특히, 색상 기반 손동작 인식 방법의 대부분은 피부색과 유사한 색상의 객체가 존재하거나 또는 배경이 피부색과 유사한 경우 손동작을 정확하게 인식하기 어렵다. 따라서 추적 기반 제스처 인식의 단점을 보완하고 다양한 배경에서 손동작을 정확하게 인식할 수 있는 방법에 대한 연구가 필요하다.

본 논문의 구성은 다음과 같다. Ⅱ장에서는 제안한 방법의 전체 시스템 구성, 손 후보 영역 검출, 특징 추출, 유사도 비교, 그리고 제안된 DSTW 알고리즘에 대해 설명한다. Ⅲ장에서는 기존의 방법과 제안된 방법에 대하여 각각 성능을 평가한다. 마지막으로 Ⅳ장에서는 결론 및 향후 연구 방향을 제시한다.

Ⅱ. 본 론

본 논문에서 제안한 손동작 인식 방법은 손 후보 영역 검출 과정, 특징 추출 과정, 유사도 비교 과정, 그리고 모델·질의 영상 매칭 과정으로 구성된다. 손 후보 영역 검출 과정은 색상(Color)과 동작(Motion) 정보를 이용하여 K개의 손 후보 영역을 검출하는 과정이다. 특징 추출 과정은 검출된 후보 영역의 2가지 특징(위치, 속도)을 추출하는 과정이다. 유사도 비교 과정은 첫 프레임의 각 손 후보 영역을 기준으로 다른 프레임의 손 후보 영역들과 유사도를 비교하는 과정이다. 마지막으로 모델· 질의 영상 매칭 과정은 제안된 DSTW 알고리즘을 사용하여 최적의 매칭 경로와 비용을 계산하는 과정이다. 그림 1은 제안하는 방법의 전체 시스템의 구성도를 보여준다.

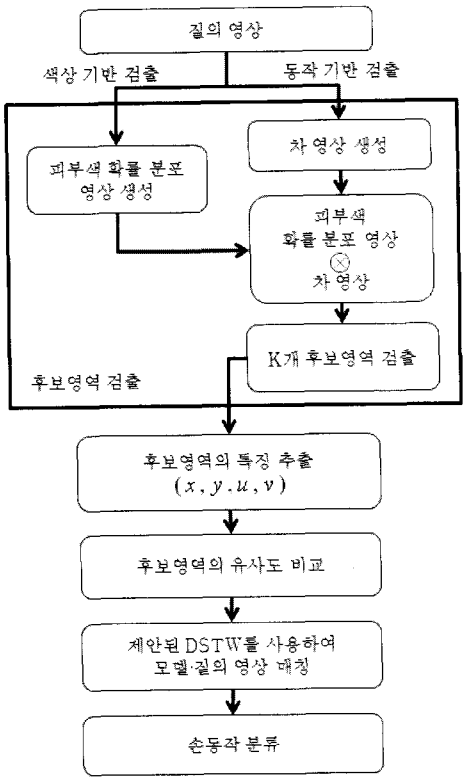


그림 1. 제안한 손동작 인식 방법의 시스템 구성도
Fig. 1. Framework of the proposed method.

1. 손 후보영역 검출

본 절에서는 질의 영상에서 매 프레임마다 K개의 손 후보 영역을 검출하는 방법에 대해 설명한다. 본 논문에서는 비교적 간단하면서도 효과적인 결과를 보여주는 색상과 동작 정보를 이용하여 손 후보 영역을 검출한다.

가. 피부색 확률 분포 영상 생성

본 논문에서는 피부색 히스토그램(Skin and non-skin RGB color histogram)[5]을 사용하여 질의 영상의 매 프레임마다 피부색 확률 분포 영상을 생성한다.

영상 내에서 한 픽셀에 대한 피부색 확률 값은 식(1)의 베이즈 규칙(Bayes rule)을 이용하여 계산된다.

$$P(skin|rgb)=\frac{P(rgb|skin)}{P(rgb|skin)+P(rgb|\sim skin)} \quad (1)$$

위 식(1)에서 사전확률 $P(skin)$과 $P(\sim skin)$은 모두 0.5를 사용한다.

나. 차 영상 생성

본 논문에서는 구현이 간단하면서 연산 속도가 빨라 실시간 실현이 용이한 프레임 차 방법을 사용하여 차

영상을 생성한다. 프레임 차 방법은 식 (2)와 같이 현재 프레임과 이전 프레임 간의 픽셀 명도 값의 차이를 이용한다.

$$\triangle I_{abs}(x,y)=|(I_j(x,y)-I_{j-1}(x,y))| \quad (2)$$

이때 $I(x,y)$와 j는 각각 영상 내 x,y 좌표의 픽셀 밝기 값과 프레임 번호이다.

다. 손 후보 영역 검출 과정

그림 2는 손 후보 영역 검출 과정을 보여준다. 제안된 방법은 (a)를 입력 영상으로 사용하여 생성된 피부색 확률 분포 영상 (b)와 차 영상 (c)를 곱한다. (b)와 (c)를 곱한 결과 영상은 (d)와 같다. 그 후 결과 영상 (d)에 대해 손 크기의 블록(30×40)을 사용하여 회선(Convolution)을 수행한다. 회선 연산은 속도를 향상시키기 위해 적분 영상(Integral Image)[6]을 사용한다. 회선 연산을 수행한 후에 (e)와 같이 영상 내에서 피부색 확률 분포 합이 가장 큰 블록 영역을 검출한다. (f),(g)와 같이 검출된 블록 영역을 제외한 나머지 피부색 확률 분포에 대해 같은 방법으로 피부색 확률 분포 합이 가장 큰 상위 K개의 손 후보 영역을 검출한다. (h)는 8개의 손 후보 영역을 검출한 영상이다.

2. 특징 추출

본 논문에서는 손 후보 영역의 특징으로 위치(Position)와 속도(Velocity)를 사용한다. 질의 영상의 j번째 프레임에서 k번째 손 후보 영역의 특징 벡터 Q_{jk}는 식(3)과 같다.

$$Q_{jk}=(x_{jk},y_{jk},u_{jk},v_{jk}) \quad (3)$$

(x,y)는 손 후보 영역의 중심 위치이고, (u,v)는 옵티컬 플로우[7]를 사용하여 계산한 속도이다. 중심 위치는 손 후보 영역에서 간단하게 추출할 수 있는 특징 값으로 손 후보 영역들의 중심 위치는 손동작의 전체 궤적을 의미한다. 속도는 위치에 불변한 특징 값으로 손 후보 영역이 움직인 방향과 크기를 의미한다.

3. 유사도 비교

가. 불변 모멘트

후보 영역들의 유사도를 비교하기 위해서는 후보 영역에 대한 질감 정보를 알아야 한다. 본 논문에서는 후보 영역들 사이의 유사도를 비교하기 위해 이동, 회전, 크기 변화에 강인하고 후보 영역의 질감을 잘 표현할 수 있는 불변 모멘트(Invariant Moment)를 사용

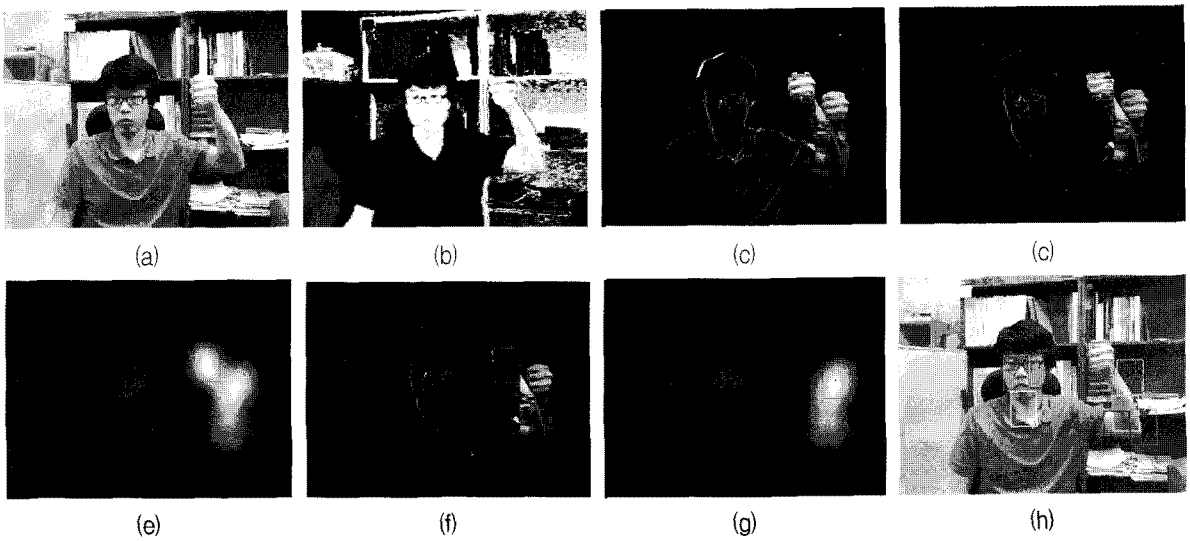


그림 2. 손 후보 영역 검출. (K = 8) (a) 입력 영상, (b) 피부색 확률 분포 영상 , (c) 차 영상, (d) 피부색 확률 분포 영상 ⊗ 차 영상, (e) 피부색 확률 분포 합이 가장 큰 블록 영역, (f) 검출된 블록 영역을 제외한 영상, (g) 피부색 확률 분포 합이 두 번째로 큰 블록 영역, (h) 검출된 손 후보 영역

Fig. 2. Detection of the candidate regions of hand (a) Input image, (b) Probability distribution of skin color, (c) Frame difference, (d) Result image multiplying two images (b) and (c), (e) The block with the highest sum of pixel intensity (f) Result image removing the detected block, (g) The block with the second highest sum of pixel intensity, (h) The detected candidate hand region.

한다[8~13].

$$\phi_1 = \eta_{20} + \eta_{02}$$
$$\phi_2 = (\eta_{20} + \eta_{02})^2 + 4\eta_{11}^2$$
$$\phi_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2$$
$$\phi_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2$$
$$\phi_5 = (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03}) \times [3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \quad (4)$$
$$\phi_6 = (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03})$$
$$\phi_7 = (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] - (\eta_{30} - 3\eta_{12})(\eta_{21} + \eta_{03}) \times [3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2]$$

본 논문에서는 중심 모멘트와 정규 모멘트를 이용하여 불변 모멘트를 추출한 후 후보 영역의 유사도 비교에 사용한다. 불변 모멘트는 식 (4)와 같이 2차, 3차 정규 모멘트로 구성된다.

식 (4)에서 정의한 불변 모멘트의 의미는 다음과 같다.

ϕ_1 : 수평, 수직 방향 분산의 합. 수직과 수평축으로 많이 퍼져 있을수록 값이 커짐.

ϕ_2 : 수직, 수평축의 분산 값이 비슷할 경우 수직, 수평축에 대한 상호 분산 값.

ϕ_3 : 좌우, 상하로 치우친 값을 강조하는 결과.

ϕ_4 : 좌우, 상하로 치우친 값을 상쇄하는 결과.

ϕ_5, ϕ_6, ϕ_7 : 크기, 회전, 위치에 불변한 특징 값을 추출.

나. 유사도 비교 과정

본 논문에서는 7개의 불변 모멘트 중 유사도 비교에 사용할 불변 모멘트를 선별하기 위해 후보 영역을 크게 5개의 그룹(손, 부분 손, 팔뚝, 얼굴, 배경)으로 구분한다. 5개의 그룹은 실험을 통해 검출 비율이 높은 후보 영역들을 선별하였고, 각 그룹은 200개의 샘플 영상을 사용하였다. 그림 3은 후보 영역의 예를 보여 준다.

표 1은 5개의 그룹에서 각각 200개의 샘플 영상을 사용하여 계산한 불변 모멘트의 평균과 분산이다.

불변 모멘트는 후보 영역의 질감 정보를 의미한다. 본 논문에서는 각 후보 영역을 보다 정확하게 구분하기 위해 식 (5)와 같이 가중치가 부여된 불변 모멘트를 사용한다.

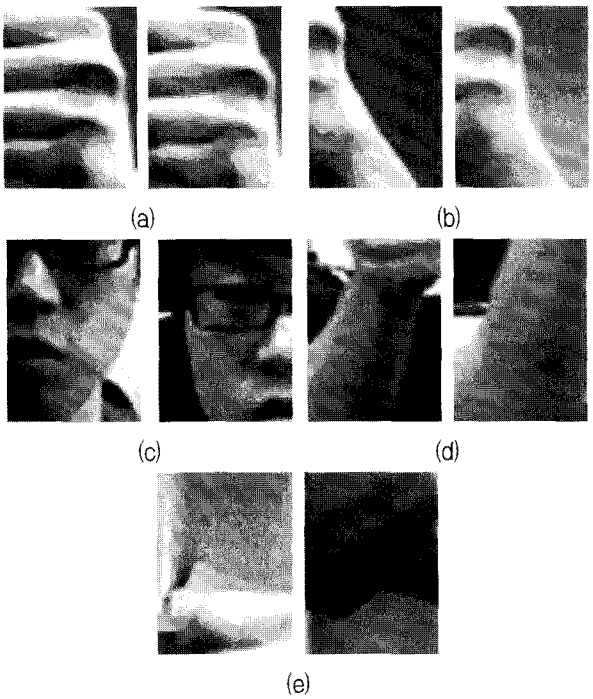


그림 3. 후보 영역 예.
(a) 손 (b) 부분 손 (c) 얼굴 (d) 팔뚝 (e) 배경

Fig. 3. Examples of candidates (a) Hand, (b) A part of hand, (c) Face, (d) Forearm, (e) Background.

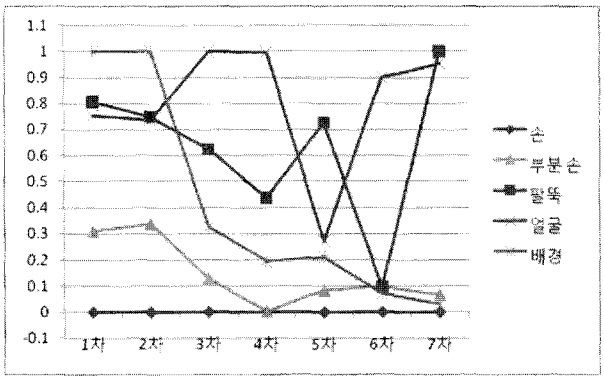


그림 4. 각 후보 영역의 정규화 된 평균 불변 모멘트

Fig. 4. Normalized average of invariant moments in terms of each candidate region.

$$\phi_n = \omega_n \times \phi'_n \quad n = 1,...,7 \quad (5)$$

n은 불변 모멘트의 차수이고, ϕ'_n은 표 1의 불변 모멘트의 평균이다. ω_n은 불변 모멘트의 평균을 사용하여 구한 가중치이며, 1~7차수의 각 가중치는 표 2와 같다.

그림 4는 각 불변 모멘트의 분별력을 비교하기 위하여 그룹별 각 차수의 불변 모멘트의 평균값을 나타낸 그래프이다. 그림 4에서 확인할 수 있듯이 6, 7차 모멘

표 1. 후보 영역의 불변 모멘트의 평균과 분산
Table 1. Average and variance of invariant moments in candidate regions.

후보영역		손	부분 손	팔뚝	얼굴	배경
1차	평균	1.05317E-03	1.19711E-03	1.42746E-03	1.40294E-03	1.51706E-03
	분산	5.36854E-05	1.81526E-04	2.30864E-04	2.91068E-04	3.12575E-04
2차	평균	8.48621E-08	1.17266E-07	1.56482E-07	1.55135E-07	1.80474E-07
	분산	1.59513E-08	6.97223E-08	1.19356E-07	1.00236E-07	1.27780E-07
3차	평균	2.45870E-12	9.79459E-12	3.78405E-11	5.92930E-11	2.09592E-11
	분산	5.73867E-12	2.27735E-11	5.85482E-11	3.20026E-10	7.01170E-11
4차	평균	1.05041E-11	1.03591E-11	2.97341E-11	5.42901E-11	1.90117E-11
	분산	8.78430E-12	2.71268E-11	5.05376E-11	2.01377E-10	6.82121E-11
5차	평균	4.81229E-24	3.05540E-23	2.29208E-22	7.93267E-23	6.99297E-23
	분산	1.84256E-22	2.67803E-21	7.42712E-21	2.55281E-19	1.03368E-20
6차	평균	2.31365E-15	1.20731E-15	3.37295E-15	1.22777E-14	3.10992E-15
	분산	2.52322E-15	1.37375E-14	2.72019E-14	9.37132E-14	3.16761E-14
7차	평균	9.35803E-24	6.79959E-27	1.31420E-22	1.24728E-22	5.12487E-24
	분산	1.83280E-22	1.20561E-21	6.78491E-21	1.98380E-19	1.37865E-20

표 2. 불변 모멘트의 가중치
Table 2. Weights of invariant moments.

차수	1차	2차	3차	4차	5차	6차	7차
가중치	1.0E+4	1.0E+8	1.0E+12	1.0E+12	1.0E+23	1.0E+15	1.0E+23

트는 손 영역(손, 부분 손)과 나머지 후보 영역들 사이의 값의 차이가 작기 때문에 후보 영역들을 정확하게 구분할 수 없다. 5차 불변 모멘트는 분산이 크기 때문에 후보 영역들을 정확하게 구분하기 어렵다. 불변 모멘트의 분별력 비교 결과 7개의 불변 모멘트 중 1~4차 불변 모멘트만 각 후보 영역들을 정확하게 구분할 수 있는 것을 확인하였다. 따라서 본 논문에서는 후보 영역 간의 유사도를 비교하기 위해 1~4차 불변 모멘트를 사용한다.

4. 제안된 Dynamic Space Time Warping 알고리즘

DSTW는 DTW를 공간영역으로 확장한 알고리즘으로써 질의 영상에서 매 프레임마다 검출된 K개 손 후보 영역의 특징을 이용한다[14].

전체 m개의 프레임으로 구성된 모델 영상의 시퀀스는 $M=(M_1,...,M_m)$이며, M_i는 i번째 프레임에서 추출된 특징 벡터를 나타낸다. 전체 n개의 프레임으로 구성된 질의 영상의 시퀀스는 $Q=(Q_1,...,Q_n)$이다. j번째 프레임 Q_j는 총 K개의 손 후보 영역을 포함하며 $Q_j=\{Q_{j1},...,Q_{jK}\}$로 나타낸다. 이때 Q_{jk}는 j번째 프레임의 k번째 손 후보 영역에서 추출된 특징 벡터이다. 와핑 경로(Warping Path) W는 M과 Q 사이의 최적의 매칭 경로(Matching Path)이며 $W=(w_1,...,w_T)$로 나타낸다. 이때 W의 길이 T는 식 (6)과 같다.

$$\max(m,n) \le T \le m+n-1 \tag{6}$$

W의 한 요소는 $w_t=(i,j,k)$이며 모델 영상의 특징벡터(M_i)가 질의 영상의 특징벡터(Q_{jk})와 매칭 됨을 의미한다. W의 제약 사항은 다음과 같다.

◆ 경계 조건 : $w_1=(1,1,k)$, $w_T=(m,n,k')$
W는 모델 영상의 첫 번째 프레임과 질의 영상의 첫 번째 프레임이 매칭 됨으로써 시작하고 모델 영상의 마지막 프레임과 질의 영상의 마지막 프레임이 매칭 됨으로써 끝난다.

◆ 시간의 연속성 : $w_t=(a,b,k)$ 일 경우
$w_{t-1}=(a',b',k')$ 이며
W는 M, Q 의 두 시간 축에 대해 인접한 셀로 이동한다.

◆ 시간의 단조성 : $w_t=(a,b,k)$ 일 경우
$w_{t-1}=(a',b',k')$ 이며
$a-a' \ge 0$, $b-b' \ge 0$ 이다.
W는 M, Q 의 두 시간 축에 대해 단조 증가한다.

w_t가 (i,j,k) 일 때, $N(i,j,k)$은 제약 사항을 따르는

w_{t-1}의 집합이며, 식(7)로 정의한다.

$$N(w_t)=\{(i-1,j),(i,j-1),(i-1,j-1)\}\times\{1,...,K\} \quad (7)$$

DSTW 알고리즘의 전체 구성은 그림 5와 같다. 손 후보 영역의 중심 위치 (x,y)는 간단하면서 다루기 쉬운 특징 벡터이다. 그러나 중심 위치는 절대 좌표이기 때문에 모델 영상의 손동작 위치가 질의 영상의 손동작 위치에 불변(Translation Invariance)하기 위해 중심 위치를 상대 좌표로 변환할 필요가 있다. 즉, 첫 번째 프레임의 k번째 손 후보 영역을 기준으로 나머지 프레임에 속한 모든 손 후보 영역의 중심 위치를 상대 좌표로 변환한다. 변환된 상대 좌표를 특징으로써 가지는 손 후보 영역들을 포함하는 질의 영상은 한 개의 독립적인 질의 영상으로 간주되며 DSTW 알고리즘을 사용하여 모델 영상과 비교한다. 이와 같이 전체 K개의 독립적인 질의 영상을 사용하여 총 K번의 DSTW 알고리즘을 실행한다. 산출된 K개의 매칭 비용 중 가장 작은 값이 질의 영상과 모델 영상 사이의 최적의 매칭 비용이다.

DSTW 알고리즘을 이용한 기존의 손동작 인식 방법은 색상 정보와 동작 정보를 이용하여 질의 영상의 매 프레임 마다 손 일 확률이 높은 다수의 후보 영역들을 검출한다. 검출된 손 후보 영역은 비교적 단순한 특징인 중심 위치와 속도가 추출되어 모델 영상과 질의 영상 사이의 매칭 경로와 비용을 계산하는데 사용된다. 그러나 기존의 방법은 모델 영상과 손 후보 영역에서 추출된 특징의 거리 차만을 사용하여 매칭 경로를 구성하기 때문에 손이 아닌 후보 영역들도 매칭 경로를 구성할 수 있다. 따라서 기존의 방법을 이용하여 손동작 인식을 수행하였을 경우 사용자가 의도하지 않은 손 후보 영역들로 인해 잘못된 모델 영상으로 인식될 확률이 높다.

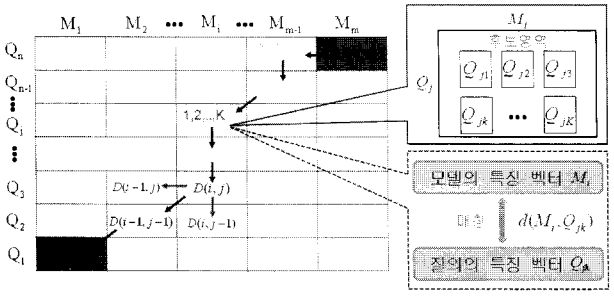


그림 5. DSTW 알고리즘의 전체 구성
Fig. 5. The composition of DSTW algorithm.

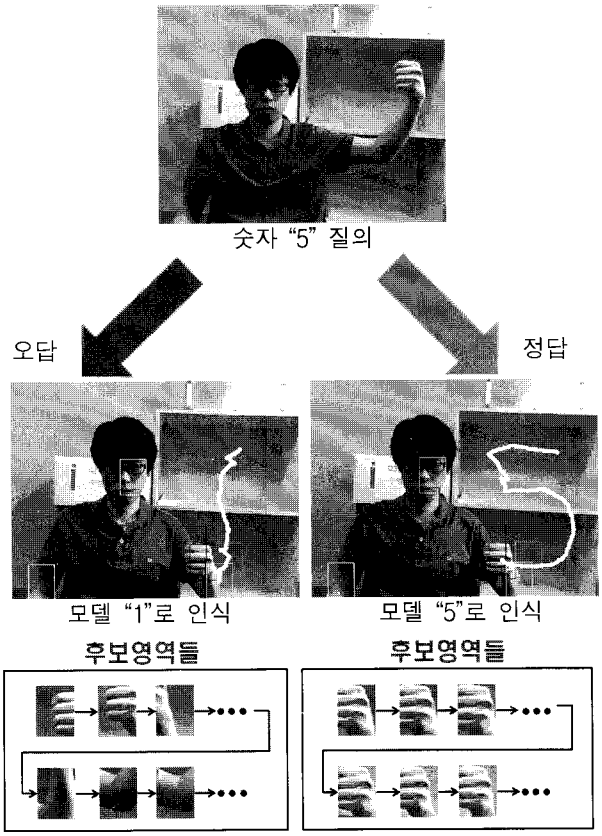


그림 6. 오답과 정답의 매칭 된 후보 영역들 예
Fig. 6. Examples of true/false matching.

DSTW 알고리즘을 이용한 기존의 손동작 인식 방법은 첫 번째 프레임에서 기준이 되는 손 후보 영역과 나머지 프레임의 손 후보 영역들 사이의 관계를 고려하지 않고 질의 영상의 손 후보 영역 중에서 모델 영상과 거리 차가 가장 작은 손 후보 영역을 선택하여 매칭 경로를 구성한다. 따라서 만일 질의 영상 내에서 사용자가 의도하지 않은 모델 영상과 매칭 비용이 가장 작은 매칭 경로를 생성한다면 사용자가 질의한 손동작은 정확하게 인식될 수 없다. 그림 6은 기존의 방법으로 사용자가 숫자 5를 질의할 경우 오답과 정답의 매칭 된 손 후보 영역들을 보여준다.

따라서 첫 번째 프레임의 기준이 되는 손 후보 영역과 나머지 프레임의 손 후보 영역들 사이의 관계를 고려한 새로운 DSTW 알고리즘의 정의가 필요하다.

본 논문에서는 이를 해결하기 위해 손 후보 영역의 질감 정보를 비교한 후 유사도를 매칭 비용에 가중치로 적용하는 방법을 제안한다. 제안된 방법은 불변 모멘트를 사용해 첫 번째 프레임에서 기준이 되는 손 후보 영역과 나머지 프레임에 속한 모든 손 후보 영역들 사이의 유사도를 계산한다. 유사도는 식 (8)과 같이 1~4차

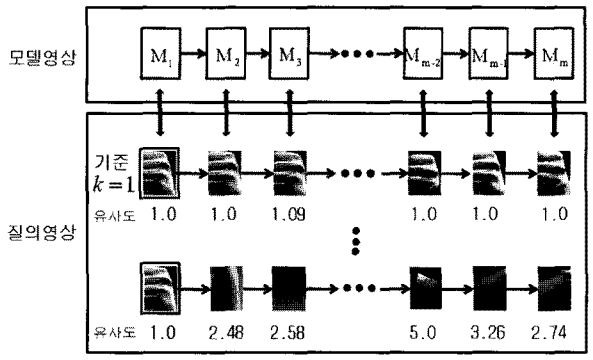


그림 7. 정규화 된 유사도 가중치를 이용한 모델· 질의 영상 매칭 예 ($C=4$)

Fig. 7. Example of matching result between model and query by using the normalized similarity weights. ($C=4$).

불변 모멘트의 차의 누적 합으로 계산한다.

$$l(A,B)=\sum_{i=1}^{4}|\phi_i^A-\phi_i^B| \tag{8}$$

A는 첫 번째 프레임에서 기준이 되는 손 후보 영역이고 B는 나머지 프레임에 속한 손 후보 영역(Q_{jk})이다. $l(A,B)$는 A와 B의 유사도를 나타내며 M_i와 Q_{jk}의 거리 차를 계산할 때 정규화 된 가중치로 적용된다.

정규화 된 유사도 가중치를 계산하는 방법은 다음과 같다.

1. 질의 영상의 첫 번째 프레임에서 기준이 되는 손 후보 영역을 선정한 후 나머지 프레임에 속한 모든 손 후보 영역(Q_{jk})과의 유사도를 계산한다.

2. 질의 영상의 j번째 프레임 내 K개의 유사도 값 중에서 가장 큰 값(기준 손 후보 영역과 가장 유사하지 않은 손 후보 영역의 유사도 값, Max)과 가장 작은 값(기준 손 후보 영역과 가장 유사한 손 후보 영역의 유사도 값, Min)을 추출한다.

3. Max와 Min을 사용하여 정규화 된 유사도 가중치를 계산한다.

$$L_{rjk}''=\frac{abs(l(A,B)-Min)}{abs(Max-Min)}$$

r : 첫 번째 프레임의 기준 후보 영역

$1\le r\le K$

4. 정규화 된 유사도 가중치의 분별력을 향상시키기 위해 임의의 상수 값을 곱한다.

$$L_{rjk}'=L_{rjk}''\times C$$

5. 마지막으로 가장 유사한 손 후보 영역의 거리 차 값이 무조건 0으로 계산되는 것을 방지하기 위해 0 ~1 사이의 정규화 된 유사도 가중치의 범위를 1~ $L_{rjk}'+1$ 사이의 범위로 변경한다.

$$L_{rjk}=L_{rjk}'+1$$

입력 : 모델 영상의 특징 벡터 $M_i,\ 1\le i\le m$

질의 영상의 특징 벡터 $Q_j=\{Q_{j1},\ldots,Q_{jK}\},\ 1\le j\le n$

기준 후보 영역 $S_r,\ 1\le r\le K$

분별력 향상을 위한 상수 C

S_r과 Q_{jk}의 정규화 된 유사도 가중치 L_{rjk}

출력 : 매칭 비용 D^*,

최적의 와핑 경로 $W^*=(w_1^*,\ldots,w_T^*)$

```
for r = 1 : K do
    j = 0
    for i = 0 : m do
        for k = 1 : K do
            D(i,j,k) = ∞
        end
    end
    D(0,0,1) = 0

    for j = 1 : n do
        for i = 0 : m do
            for k = 1 : K do
                if i = 0 then
                    D(i,j,k) = ∞
                else
                    w = (i,j,k)
                    L_rjk = (L''_rjk × C) + 1
                    for w' ∈ N(w) do
                        C(w',w) = D(w') + (L_rjk × d(w))
                    end
                    D(w) = min_{w'∈N(w)} C(w',w)
                    b(w) = argmin_{w'∈N(w)} C(w',w)
                end
            end
        end
    end

    k* = argmin_k {D(m,n,k)}
    D* = D(m,n,k*)
    w_T* = (m,n,k*)

    w_{t-1}* = b(w_t*)

end
```

그림 8. 제안된 DSTW 알고리즘

Fig. 8. The proposed DSTW algorithm.

그림 7은 정규화 된 유사도를 가중치로 사용하여 모델 영상과 질의 영상을 매칭 한 예를 보여준다.

정규화 된 유사도를 가중치로 적용하여 모델 영상 M과 질의 영상 Q 사이의 와핑 경로 W를 계산하는 제안된 DSTW 알고리즘은 그림 8과 같다. $d(i,j,k)$는 M_i와 Q_{jk}의 거리 차이며, L_{rjk}는 첫 번째 프레임에서 기준이 되는 손 후보 영역 S_r과 질의 영상 내 다른 프레임의 손 후보 영역 Q_{jk} 사이의 정규화 된 유사도 가중치이다. 제안된 DSTW 알고리즘은 최종적으로 최적의 와핑 경로 W^*와 매칭 비용 D^*를 산출한다.

Ⅲ. 실 험 결 과

1. 실험 환경

실험은 Visual Studio C++ 6.0과 OpenCV 1.0을 사용하여 구현하였고, 삼성 SPC-A130M 웹캠을 사용하여 초당 30프레임으로 획득된 영상(320×240)을 사용하였다. 인식에 사용된 제스처는 필기체 입력을 위해 Palm 사에서 고안한 Graffiti 숫자[15]를 사용하며 그림 9는 화면상으로 보이는 숫자의 방향을 나타낸다. 실험은 총 5명의 사용자를 대상으로 수행하였다. 모델 영상은 사용자 한 명당 각 숫자를 3회씩 생성하여 한 개의 평균 모델 영상을 만든다. 생성된 평균 모델 영상은 총 50개이다. 질의 영상은 사용자 한 명당 배경별로(단순한 배경, 복잡한 배경, 피부색 배경), 복장에 따라(반팔, 긴팔) 각 숫자를 3회씩 생성하여 총 900개이다.

그림 9. Palm's Graffiti 숫자
Fig. 9. Palm's Graffiti digits.

2. 실험 결과

본 논문에서 제안하는 방법의 성능을 평가하기 위해서 질의 영상에 대한 손동작 인식 실험을 수행하였다. 실험은 두 가지 방법으로 진행하였다. 첫 번째 실험은 사용자 종속 실험으로써 한 사용자가 생성한 30개의 질의 영상을 동일한 사용자가 생성한 10개의 평균 모델 영상과 비교한다. 두 번째 실험은 사용자 비종속 실험으로써 한 사용자가 생성한 30개의 질의 영상을 다른 4명의 사용자들이 생성한 40개의 평균 모델 영상과 비교한다. 성능을 평가하기 위한 인식률은 사용자 5명의 평균값이다.

다음 절에서는 사용자 종속 실험으로써 복잡한 배경에서 반팔 복장의 사용자가 기존의 방법과 제안된 방법을 사용하여 숫자를 질의한 예를 살펴보고 마지막 절에서는 사용자 종속 실험과 사용자 비종속 실험으로써 기존의 방법과 제안된 방법의 인식률에 대한 성능평가를

그림 10. 기존의 방법의 매칭 경로 (숫자 8)
Fig. 10. The matching path of the existing method (Number 8).

살펴본다.

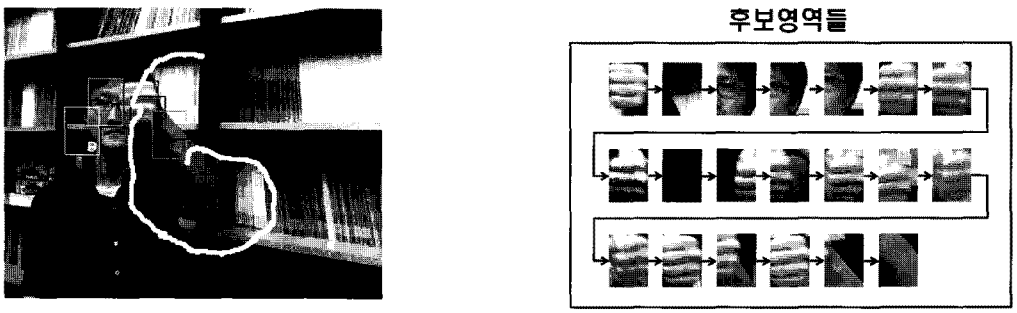
가. 사용자 종속 실험 : 복잡한 배경

(1) 기존의 방법의 실험 결과

그림 11의 (a)는 DSTW 알고리즘을 이용한 기존의 손동작 인식 방법으로 사용자 종속 실험을 위해 복잡한 배경에서 반팔 복장의 사용자가 숫자 8을 질의한 결과이다. 그림 11의 (b)는 결과로써 숫자 6의 모델 영상과 매칭 된 후보 영역들을 보여 준다. 그림 10은 기존의 방법으로 숫자 8을 질의한 경우에 숫자 6의 모델 영상과 매칭 된 전체 경로를 보여 준다. 기존의 방법은 손 후보 영역 사이의 관계를 고려하지 않기 때문에 그림 10과 같이 질의 영상 내에서 사용자의 얼굴이나 배경에 의해 잘못된 경로가 생성된 경우 의도하지 않은 모델 영상과 매칭 되는 단점이 있다.

(2) 제안된 방법의 실험 결과

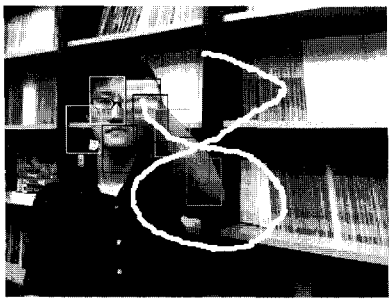
그림 13의 (a)는 제안된 DSTW 알고리즘을 이용한 손동작 인식 방법으로 사용자 종속 실험을 위해 복잡한 배경에서 반팔 복장의 사용자가 숫자 8을 질의한 결과이다. 그림 13의 (b)는 결과로써 숫자 8의 모델 영상과 매칭 된 후보 영역들을 보여 준다. 그림 12는 제안된 방법으로 숫자 8을 질의한 경우에 숫자 8의 모델 영상과 매칭 된 전체 경로를 보여 준다. 제안된 방법은 손 후보 영역 사이의 관계를 고려해 매칭 시 가중치로 사용하기



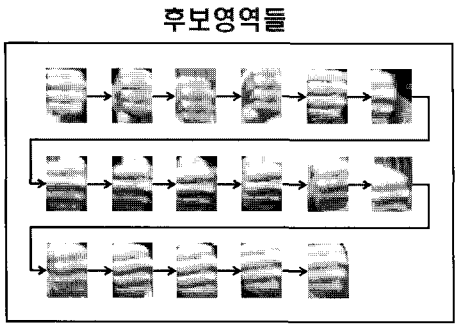
(a) (b)

그림 11. 기존의 방법을 이용한 질의 (숫자 8) (a) 매칭 결과 (b) 매칭 된 후보 영역들
Fig. 11. The query of the existing method (Number 8) (a) Matching result, (b) The matched candidate regions.

그림 12. 제안된 방법의 매칭 경로 (숫자 8)
Fig. 12. The matching path of the proposed method (Number 8).



(a)



(b)

그림 13. 제안된 방법을 이용한 질의 (숫자 8)
(a) 매칭 결과 (b) 매칭 된 후보 영역들

Fig. 13. The query of the proposed method (Number 8)
(a) Matching result, (b) The matched candidate regions.

때문에 그림 12와 같이 첫 번째 프레임에서 기준이 되는 손 후보 영역과 유사도가 높은 후보 영역들로 생성된 경로를 정확하게 매칭 한다.

나. 사용자 종속 실험에 대한 성능 평가

표 3은 사용자 종속 실험으로써 다양한 배경에서 사용자의 복장에 따른 인식률의 실험 결과를 보여 준다. 실험 결과로써 기존의 방법은 0, 1, 4, 7과 같이 비교적 간단한 궤적으로 구성된 숫자는 정확하게 인식하는 반면 2, 3, 5, 6, 8, 9와 같이 복잡한 궤적으로 구성된 숫자는 정확하게 인식하지 못한다. 특히 배경에 피부색과 유사한 색상이 많이 분포하거나, 사용자가 반팔 복장을 착용하여 손 주위의 신체 부위가 많이 노출될수록 오인식 되는 결과가 두드러지게 나타난다. 이는 궤적이 복잡한 숫자일수록 궤적 주위의 배경에서 잘못된 경로를 생성하거나, 손뿐만 아니라 팔뚝의 피부색 영역이 잘못된 경로를 생성하기 때문이다. 그러나 제안된 방법은 유사도가 낮은 후보 영역들로 구성된 경로를 제외함으로써 인식률이 기존의 방법에 비해 2~12% 향상 된 것을 확인할 수 있다.

표 3. 사용자 종속 실험 : 배경과 복장에 따른 인식률 (K = 8)

Table 3. User-dependant experiments : the recognition rate with respect to various circumstances of background and cloth. (K = 8)

(단위 : %)

환경	단순한 배경		복잡한 배경		피부색 배경	
방법	기존 DSTW	제안된 DSTW	기존 DSTW	제안된 DSTW	기존 DSTW	제안된 DSTW
긴팔	92.0	96.6	90.6	92.6	88.6	92.6
반팔	89.3	96.0	83.3	92.0	82.0	92.0

다. 사용자 비종속 실험에 대한 성능 평가

표 4는 사용자 비종속 실험으로써 다양한 배경에서 사용자의 복장에 따른 인식률의 실험 결과를 보여 준다. 표 3의 사용자 종속 실험 결과와 유사한 실험 결과로써 제안된 방법의 인식률이 기존의 방법에 비해 3~13% 향상된 것을 확인할 수 있다. 그러나 표 4의 사용자 비종속 실험은 전반적으로 표 3의 사용자 종속 실험보다 인식률이 감소한 것을 확인할 수 있다. 이는 비교 대상이 될 모델 영상의 숫자 크기가 사용자마다 서로 다르기 때문이다.

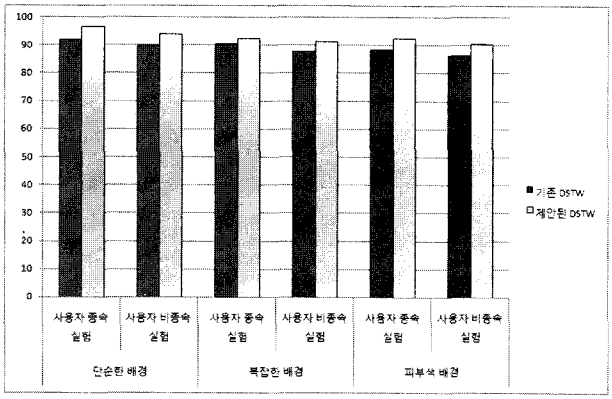
그림 14의 (a)와 (b)는 각각 긴팔, 반팔 복장의 사용자가 기존의 방법과 제안된 방법을 사용하여 수행한 사용자 종속 실험과 사용자 비종속 실험의 인식률을 비교한 결과이다. 그림에서 보는 바와 같이 제안된 방법은 두 가지 실험에서 기존의 방법보다 높은 인식률을 보인다. 특히 기존의 방법은 반팔 복장의 사용자를 대상으로 수행한 실험에서 인식률이 큰 감소를 보이는 반면 제안된 방법은 90% 이상의 인식률을 유지한다.

표 4. 사용자 비종속 실험 : 배경과 복장에 따른 인식률 (K = 8)

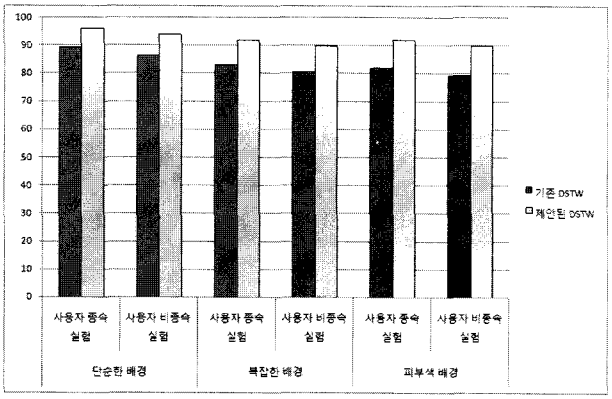
Table 4. User-independant experiments : the recognition rate with respect to background and cloth. (K = 8)

(단위 : %)

환경	단순한 배경		복잡한 배경		피부색 배경	
방법	기존 DSTW	제안된 DSTW	기존 DSTW	제안된 DSTW	기존 DSTW	제안된 DSTW
긴팔	90.0	94.0	88.0	91.3	86.6	90.6
반팔	86.6	94.0	80.6	90.0	79.3	90.0



(a)



(b)

그림 14. 사용자 종속· 비종속 실험 결과 비교
(a) 긴팔 (b) 반팔

Fig. 14. Comparison of user-dependant/independent experimental results.
(a) long sleeve, (b) short sleeve.

IV. 결 론

본 논문에서는 손 후보 영역의 질감 정보를 고려하여 다양한 배경에서도 인식률이 우수한 DSTW 기반의 손 동작 인식 방법을 제안한다. 제안한 방법은 손 후보 영역 사이의 질감 정보를 비교하기 위해 7개의 불변 모멘트 중 분별력이 우수한 4개의 불변 모멘트를 사용한다. 유사도 비교는 질의 영상 내 첫 번째 프레임의 각 손 후보 영역을 기준으로 나머지 손 후보 영역들에 대해 불변 모멘트의 거리 차로 계산되며 모델 영상과 질의영상을 매칭 할 때 정규화 된 가중치로 사용한다. 즉, 모델 영상과 질의 영상의 손 후보 영역에서 추출한 특징 벡터의 거리 차에 정규화 된 유사도 가중치를 적용함으로써 질감이 서로 다른 후보 영역들로 생성된 경로가 다른 모델로 오 인식될 수 있는 것을 제한한다. 실험 결과를 통해 제안된 방법이 기존의 방법보다 다양한 배경에서도 인식률이 우수함을 확인하였다. 그러나 사용자 비종속 실험일 경우 질의 영상과 비교할 모델 영상의 숫자 크기가 사용자마다 달라서 오 인식되는 문제점이 있다. 이를 보완하기 위한 향후 과제로는 모델 영상과 질의 영상의 숫자 크기에 불변한 인식 방법에 대한 연구가 필요하다.

참 고 문 헌

[1] 홍동표, 우운택, "제스처기반 사용자 인터페이스에 대한 연구 동향," Telecommunications Review, Vol. 18, No. 3, pp. 403-413, 2008.

[2] T. B. Moeslund, E. Granum, "A Survey of Computer Vision-Based Human Motion Capture," Computer Vision and Image Understanding, Vol. 81, No. 3, pp. 231-268, 2001.

[3] M. Turk, "Computer Vision in the Interface," Communications of the ACM, Vol. 47, No. 1, pp. 60-67, 2004.

[4] 장효영, 김대진, 김정배, 변중남, "3차원 공간상의 수신호 인식 시스템에 대한 연구," *전자공학회논문지*, 제41권 CI편 제3호, pp. 103-114, 2004.

[5] M. J. Jones, J. M. Rehg. "Statistical color models with application to skin detection," *International Journal of Computer Vision*, Vol. 46, No. 1, pp. 81-96, January 2002.

[6] P. Viola, M. J. Jones, "Rapid object detection using a boosted cascade of simple features," In Proc. IEEE Conference on Computer Vision and Pattern Recognition, Vol. 1, pp. 511-518, 2001.

[7] Q. Yuan, S. Sclaroff, and V. Athistos, "Automatic 2D hand tracking in video sequences," In Proc. WACV, Vol. 1, pp. 250-256, 2005.

[8] M. K. Hu, "Visual pattern recognition by moment invariants," *IEEE Transactions on information Theory*, Vol. 8, No. 2, pp. 179-187, 1962.

[9] Papoulis, *Probability, Random Variables, and Stochastic Processes*, McGraw Hill, 1965.

[10] M. Nadler and E. P. Smith, *Pattern Recognition Engineering*, Wiley-Interscience, pp. 197-1199, 1993.

[11] 신광규, 이강현, "Hu 불변 모멘트를 이용한 장문인식 알고리즘," *대한전자공학회논문지*, 제42권, CI편 제2호, pp. 31-38, 2005.

[12] C. H. Teh and R. T. Chin, "On Digital Approximation of Moment invariants," Computer Vision, Graphics, And Image Processing, Vol. 33,

pp. 318-326, 1986.
[13] M. K. Hu, "Pattern recognition by moment invariants," Proc. IRE Trans. Information Theory, Vol. 8, pp. 179-187, 1962.
[14] J. Alon, V. Athitsos, Q. Yuan, S. Sclaroff, "Simultaneous Localization and Recognition of Dynamic Hand Gestures," IEEE Workshop on Motion and Video Computing (WACV/MOTION '05), Vol. 2, pp. 254-260, 2005.
[15] Palm. Graffiti alphabet. http://www.palmone.com

저 자 소 개

지 재 영(학생회원)
2008년 안양대학교 디지털미디어 공학과 학사 졸업.
2010년 한양대학교 컴퓨터공학과 석사 졸업.
<주관심분야 : 객체추적, 패턴인식, 영상처리>

장 경 현(학생회원)
2005년 한양대학교 전자컴퓨터 공학부 학사 졸업.
2007년 한양대학교 컴퓨터공학과 석사 졸업.
2007년~현재 한양대학교 컴퓨터 공학과 박사 과정.
<주관심분야 : 객체추적, 패턴인식, 영상처리>

이 정 호(학생회원)
2004년 한양대학교 전자컴퓨터 공학부 학사 졸업.
2006년 한양대학교 컴퓨터공학과 석사 졸업.
2006년~현재 한양대학교 컴퓨터 공학과 박사 과정.
<주관심분야 : 영상변형, 영상처리, 머신비전>

문 영 식(평생회원) - 교신저자
1980년 서울대학교 공과대학 전자 공학과 학사 졸업.
1982년 한국과학기술원 전기 및 전자공학과 석사 졸업.
1990년 University of California at Irvine Dept. of Electrical and Computer Engineering. 박사 졸업.
1982년~1985년 한국전자통신연구소 연구원
1989년~1990년 Inno Vision Medical 선임연구원
1990년~1992년 생산기술연구원 선임연구원
1992년~현재 한양대학교 컴퓨터공학과 정교수
<주관심분야: 컴퓨터비전, 영상검색, 패턴인식>