

논문 2010-47CI-3-1

손동작 인식에 의한 컴퓨터 비전 인터페이스 설계

(Design of Computer Vision Interface by Recognizing Hand Motion)

윤진현*, 이종호**

(Jin Hyun Yun and Chong Ho Lee)

요약

손동작을 통한 입력방법은 컴퓨터와 디지털 기기의 발전에 따라 요구되는 새로운 HCI(Human-Computer Interaction) 방법으로 그 가능성을 가지고 있으며 이에 대한 다양한 시도가 있었다. 본 논문에서는 컴퓨터 비전을 기반으로 단일 카메라를 사용하는 손 영역 검출 및 추적방법을 제시하고 이에 의한 컴퓨터 인터페이스를 제안한다. 기존에 많이 쓰이는 피부색 매치 방법에 추가하여 형태 정보를 더함으로써 손 영역 검출능력을 향상 시켰다. 이러한 형태 정보를 추출하는 방법으로 주요 방향 에지 기술자라는 방법을 제안하였고 이는 강력하여 학습 시간 없이 한 가지 손 모델만을 사용하여 손 영역 검출을 할 수 있다. 또한 손 영역 검출과 추적하는 방법을 나누어 추적할 때는 회전에 대한 자유도를 높이도록 설계 하였다. 위 방법을 이용하여 3차원 공간에 그려지는 필기체 숫자 인식에 적용해 보았으며 분류 방법으로 DNAC 알고리즘을 사용하였다. 결과적으로 손 영역 검출은 82%의 검출률을 보였고 필기체 숫자 인식은 90%의 인식률을 보였다.

Abstract

As various interfacing devices for computational machines are being developed, a new HCI method using hand motion input is introduced. This interface method is a vision-based approach using a single camera for detecting and tracking hand movements. In the previous researches, only a skin color is used for detecting and tracking hand location. However, in our design, skin color and shape information are collectively considered. Consequently, detection ability of a hand increased. we proposed primary orientation edge descriptor for getting an edge information. This method uses only one hand model. Therefore, we do not need training processing time. This system consists of a detecting part and a tracking part for efficient processing. In tracking part, the system is quite robust on the orientation of the hand. The system is applied to recognize a hand written number in script style using DNAC algorithm. Performance of the proposed algorithm reaches 82% recognition ratio in detecting hand region and 90% in recognizing a written number in script style.

Keywords : HCI, hand detection, hand tracking, primary orientation edge descriptor, gesture recognition.

I. 서론

HCI에 관한 연구는 디지털 기기라는 것이 생긴 이후로 계속적인 발전을 거듭해 왔다. 현재 키보드, 마우스, 터치비전등 많은 방법들이 HCI로써 개발되어 사용되고 있지만 컴퓨터의 목적이나 그 쓰임새가 다양해짐에 따

라 기존의 방법들은 입력 장치로 불편하거나 불충분한 면을 가지고 있다. 따라서 좀 더 다양하고 사람 중심의 자연스러운 입력을 위해서는 다른 입력 방법을 찾아야 할 것이다.^[1~2] 이 중 매력적인 방법이 손동작을 통한 입력방법이다. 손동작은 이미 사람과 사람 사이의 의사소통에 많이 쓰이는 만큼 사람이 기계에 입력함에 있어서 기존의 입력장치보다 편리함을 줄 것이고 별도의 입력장치가 필요 없다는 장점이 있다.^[3] 이를 이용하여 TV 리모컨이나 휴대폰, 또는 로봇과의 협업에 있어서 새로운 인터페이스로 사용할 수 있다.

손동작에 관한 연구는 접근 방법에 따라 크게 두 가지로 나뉜다.

* 학생회원, ** 정회원, 인하대학교 정보통신공학과
(Department of Information and communication Engineering, Inha University)

※ 이 논문은 인하대학교 연구비 지원에 의하여 연구되었음.

접수일자: 2010년4월5일, 수정완료일: 2010년4월30일

첫 번째로 글러브 데이터 기반 방법은 손과 손가락에 부착된 센서를 사용하여 손의 움직임을 쉽게 파악한다. 하지만 이 장치는 가격이 비싸고 센서 장치를 입어야 하기 때문에 번거롭다는 단점을 가지고 있다.

두 번째 방법으로 영상 기반 방법은 카메라만을 사용하는 방법으로 가격이 비교적 저렴하고 간단한 장치만 있으면 되지만 배경에서 물체를 구분하기 어렵고 조명에 영향을 많이 받기 때문에 손동작을 분석하기 어려운 문제를 가지고 있다.^[4~5]

또한 영상 기반 방법은 3차원 손 모델 인식 기반 방법(3D hand model-based)과 외관 인식 기반 방법(Appearance-based)으로 나뉜다. 3차원 손 모델 기반 방법은 3D 운동학적인 손 모델을 사용하는 방법으로 비교적 정확하게 손 움직임을 파악할 수 있으나 연산량이 많고 많은 각도에서의 영상 데이터를 요구한다. 따라서 실시간 손동작 파악에는 적합하지 않다.^[4, 6]

반면 외관 인식 기반 방법은 2차원 영상의 특징(Feature)으로 손동작을 인식하는 방법으로 연산량이 적으므로 실시간으로 손동작 파악이 가능하다. 하지만 색에 의존도가 강하므로 피부색을 정확하게 구분하지 못하면 배경에 비슷한 색이 있을 경우 파악에 어려움이 생긴다. 이러한 어려움으로 기존 연구들은 배경을 한 가지 색으로 한다든지 색이 있는 장갑이나 마커의 사용 등의 제한적인 요소를 가지고 연구를 하였다.^[4, 7]

이에 저자는 기존의 제한적인 요소에 구애받지 않고 손을 검출하기 위해 피부색과 에지의 방향을 이용하는 방법을 사용하였고 별도의 학습과정 없이 손 모델을 사용하여 검출하였다. 또한 손을 추적하기 위해 무게 중심 추적 방법을 사용하였다. 그리고 그 손을 추적하여 공간에 숫자를 썼을 때 컴퓨터가 숫자를 인식하는 시스템을 구현해 보았다.

본 논문의 구성은 다음과 같다. II장에서는 제안한 방법의 전체 시스템의 구성, 전처리 과정, 손 검출 과정, 손 추적 과정, 필기체 숫자 인식에 대해 설명한다. 또한 III장에서는 실험을 통하여 각 과정에 대한 성능을 비교한다. 제안한 손 검출 방법 성능, 손 추적 방법 성능, 필기체 숫자 인식에 대한 성능을 기존의 방법과 비교 분석한다. IV장은 결론과 향후 연구 방향을 제시한다.

II. 입력 인터페이스 시스템 구성

전체 시스템 구성은 그림 1과 같고 연속화면에서 손

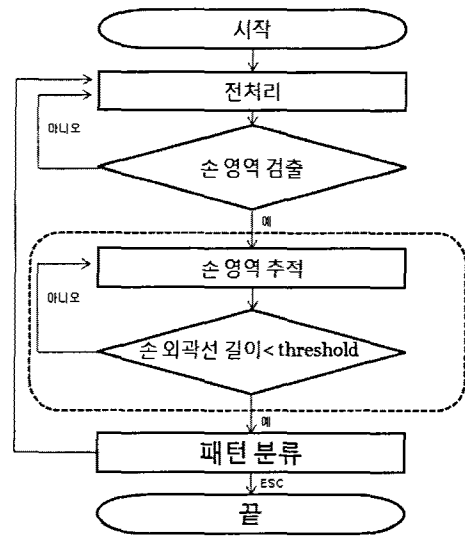


그림 1. 인식 시스템 구성
Fig. 1. Structure of the recognition system.

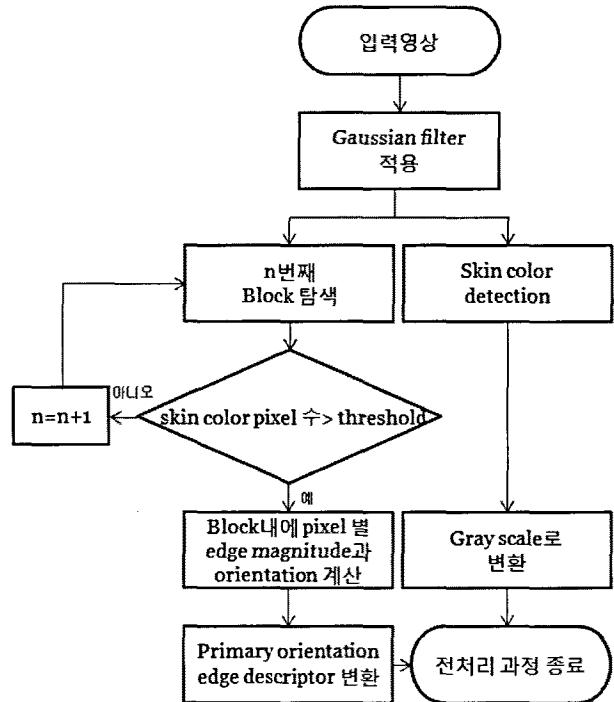


그림 2. 전처리 과정
Fig. 2. Preprocessing.

영역 검출이 완료된 후 손 궤적 추적이 시작되며 손 영역의 검출에는 피부색과 에지를 이용한 세밀한 손 영역의 기준을 적용하고 손 영역 추적은 움직임이 있으므로 손의 회전이 자유로운 방법을 적용하였다. 움직임 궤적의 종료를 나타내는 신호가 감지되면 추적이 끝나고 궤적을 분류하여 인식 결과를 나타내게 되고 다른 입력을 위해 시스템은 다시 대기상태로 돌아가게 된다.

1. 전처리

손 모델 영상과 손 영역을 찾으려는 영상 둘 다 그림 2와 같은 전처리 과정을 거쳐야 한다. 각 영상은 두 가지 영상 공간으로 표현되게 되는데 하나는 피부색 부분만 검출된 그레이 스케일의 영상이고 다른 하나는 주요 방향 에지 기술자(Primary orientation edge descriptor)로 표현된 영상이다.

가. 가우시안 필터링

가우시안 필터링은 영상에서 가우시안 노이즈를 제거하거나 블러링 시키는 효과를 가지고 있다. 즉 화소값이 불규칙하게 있는 경우 그 효과를 볼 수 있다. 따라서 원영상에 가우시안 필터는 화소값의 변화가 작은 부분인 저주파 영역을 제거하고 화소값의 변화가 큰 부분인 고주파 영역을 남기는 방법이므로 세세한 부분은 제거한다.

나. 피부색 검출

피부색을 검출하기 위한 방법으로 YCbCr 색상 공간을 사용하였다. 여러 종류의 색상 공간들은 RGB 색상 공간을 선형 또는 비선형적으로 변형시킴으로써 만들어 지는데 조명에 강하거나 피부색과 피부색이 아닌 부분을 명확히 구분하여 주는 등 여러 목적으로 특성에 맞게 사용된다. 이 중에 특히 YCbCr 색상 공간은 사람의 피부색을 구분하는데 매우 효과적이다.^[8~10] 그 이유는 YCbCr 색상 공간에서 피부색의 분포는 매우 유사하면서도 협소한 영역을 차지하고 있기 때문이고 RGB 정보에 비해 조명에 덜 민감하기 때문이다. RGB에서 변환은 아래와 같은 식을 통해 변환이 가능하다.

$$\begin{aligned} Y &= 0.299R + 0.587G + 0.114B \\ Cr &= R - Y \\ Cb &= B - Y \end{aligned} \tag{1}$$

대표적으로 피부색을 나타내는 값으로 다음과 같은 범위를 지정한다.^[11]

$$\begin{aligned} Y &> 75 \\ 77 &\leq Cb \leq 127 \\ 133 &\leq Cr \leq 173 \end{aligned} \tag{2}$$

식(2) 범위에 해당하는 영상의 픽셀은 남겨두고 나머지 부분은 검은색으로 두면 그림 3(a)와 같은 영상을 얻을 수 있다. 이는 후에 손 영역을 찾을 경우 그 찾는 범위를 줄여주고 명확하게 한다.

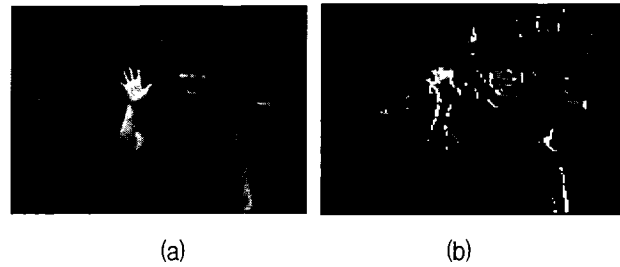


그림 3. 전처리 과정 후 영상 (a) 피부색 검출 영상 (b) 주요 방향 에지 기술자 영상

Fig. 3. After preprocessing, image; (a) detected skin color and (b) primary orientation edge descriptor.

다. 주요 방향 에지 기술자

물체 인식을 위하여 에지의 방향 히스토그램을 이용하여 기술자(Descriptor)로 표현하는 방법은 많이 사용되어 왔다.^[12~13] 그 중에 HOG(Histogram of oriented gradient)기술자는 손 모양의 유사성을 계산하는 데는 유용하다.^[14~15] 본 논문에서는 계산의 편의성을 위해 HOG를 변형하여 각 블록의 방향 히스토그램의 크기가 가장 큰 방향 구간을 선택하여 그 구간값을 블록의 대표치로 하여 기술자를 나타내었다. 기술자를 구하는 과정은 아래와 같다.

- ① Gradient 계산에 있어서 수평 성분을 구하기 위해 $[-1 \ 0 \ 1]$ 마스크를 사용하고 수직 성분을 구하기 위해 $[-1 \ 0 \ 1]$ 마스크를 전치(Transpose)시켜 적용한다.
- ② Gradient의 크기는 각 픽셀에서 수평 성분과 수직 성분을 제공하고 그 합의 제곱근을 구한다.
- ③ Gradient의 방향은 각 픽셀에서 수직 성분을 수평 성분으로 나누어 준 후 역탄젠트값을 구하여 방향을 구하고 일정구간으로 나눈다.
- ④ 일정한 픽셀로 구성된 블록에서 ③에서 구한 방향 구간들의 히스토그램을 구한다. 단, 도수는 각 픽셀 Gradient의 크기로 가중치를 두어 계산한다.
- ⑤ 가장 높은 도수를 가진 방향 구간값이 그 블록의 대표치가 된다.

본 논문에서는 영상의 한 블록에서 피부색을 가진 픽셀이 1/3이상인 부분을 기술자로 표현하고 나머지는 검정색으로 두었다. 이는 피부색이 아닌 부분은 손 영역을 찾는데 노이즈가 될 수 있기 때문이다. 그리고 그림 3(b)은 3X3 픽셀로 된 블록에 위 기술자 구하는

표 1. 방향 구간과 그에 대응하는 그레이 스케일 값
Table 1. Orientation degree range and corresponding gray scale value.

방향 구간 범위(degree)	그레이 스케일 값
0~22.5, 157.5~202.5, 337.5~360	250
22.5~67.5, 202.5~247.5	200
67.5~112.5, 247.5~292.5	150
112.5~157.5, 292.5~337.5	100

방법으로 각 방향구간을 4개의 값의 그레이 스케일로 구분하여 표현 하였다. 각 방향구간과 그와 대응되는 그레이 스케일 값은 표 1과 같다. 이 기술자의 장점은 블럭을 설정하여 계산하기 때문에 빠른 연산이 가능하고 과적합을 방지할 수 있다. 또한 블럭에 방향값을 가지고 표현한 영상은 에지 위치 정보만으로 표현한 영상보다 더 명확하게 적합도를 판단 할 수 있다.

2. 손 검출 과정

손 검출 과정은 그림 4와 같다. 전처리 과정으로 인하여 손 모델 영상과 손 영역을 찾고자 하는 영상은 각각 두 가지 영상으로 나뉜다. 같은 방식으로 표현된 영상들이 짝을 이루어 손 모델 영상을 템플릿으로 하여 슬라이딩 윈도우 방법으로 템플릿 매칭을 한다. 매칭하여 구해진 상관계수 값을 저장하여 상관계수 지도를 만들고 그레이 스케일에서 만들어진 상관계수 지도의 임계값을 넘는 영역 중 주요 방향 에지 기술자에서 만들어진 상관계수 지도의 최대값을 가지는 영역을 손 영역으로 검출한다.

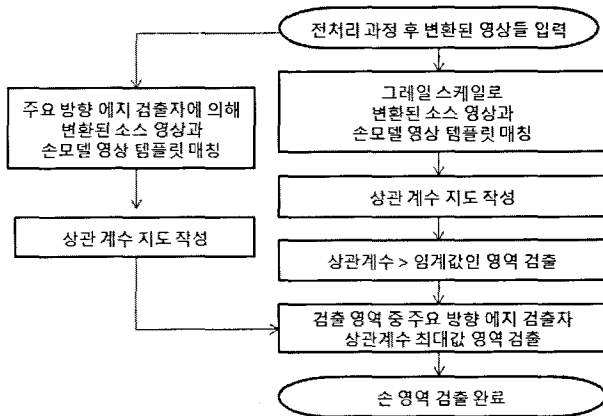


그림 4. 손 검출 과정
Fig. 4. Hand detection processing.

가. 템플릿 매칭

템플릿 매칭은 검사할 영상이 주어졌을 때 미리 주어

진 템플릿 영상을 이용하여 검사할 영상 내부에 있는 유사한 영상패턴을 찾는 것이며, 이 때 템플릿은 일종의 모델 영상이다. 템플릿 매칭은 상관계수를 이용하는 방법과 두 영상의 차를 이용하는 방법이 있다. 본 논문에서는 비교적 정확한 상관계수를 이용하여 템플릿 매칭을 하였다.^[16] 상관계수를 구하는 식 (3)과 같다. 상관계수는 -1에서 1까지 실수로 나오는데 유사할수록 1에 가까운 값이 나온다.

본 논문에서는 원 영상을 두 영상으로 나누어서 템플릿 매칭을 하는데 한 영상에는 피부색을 검출하고 그레이 스케일로 변환한 영상의 템플릿 매칭 결과를 저장한다. 이는 기존의 일반적인 템플릿 매칭과 같은 방법이다. 이 방법은 색에 많이 의존을 하기 때문에 손과 같은 한 가지 색이 영역의 대부분을 차지하는 경우 유사한 색을 가진 얼굴, 팔 등을 검출하는 경우가 있다. 따라서 보완적으로 주요 방향 에지 기술자의 템플릿 매칭으로 형태적인 정보를 결합하는데 이 주요 방향 에지 기술자로 변환한 공간을 템플릿 매칭한 결과를 두 번째 영상에 저장하여 두 영상 정보의 결합을 통해 손 영역을 검출한다.

$$correlation = \frac{\sum_{i=0}^{N-1} (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sqrt{\sum_{i=0}^{N-1} (x_i - \bar{x})^2 \cdot \sum_{i=0}^{N-1} (y_i - \bar{y})^2}} \quad (3)$$

- x는 템플릿 그레이 레벨값 영상
- \bar{x} 는 템플릿 영상의 그레이 레벨 평균값
- y는 소스 영상의 비교 부분
- \bar{y} 는 소스 영상의 비교 부분의 평균값
- N은 부분에서 전체 픽셀 수

나. 상관계수 지도

상관계수 값으로 이루어진 상관계수 지도의 구성은 그림 5와 같다. 템플릿 영상이 한 픽셀씩 또는 한 블럭씩 이동하여 구해진 상관계수를 상관계수 지도 위에서 차례로 한 픽셀에 대응된다고 생각하면 상관계수 지도의 크기는 손을 찾고자 하는 소스 영상 크기에서 비교되는 템플릿 영상 크기를 뺀 넓이와 높이이다. 그리고 그림 5에서 보면 템플릿 영상이 이동하면서 구해진 상관계수 값은 해당 영역의 왼쪽 위 꼭지점 좌표 (x,y)가 상관계수 지도에 같은 좌표(x,y)로 사상 된다.

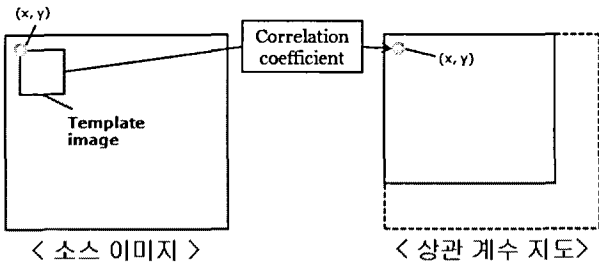


그림 5. 상관계수 지도 구성
Fig. 5. Construction of correlation coefficient map.

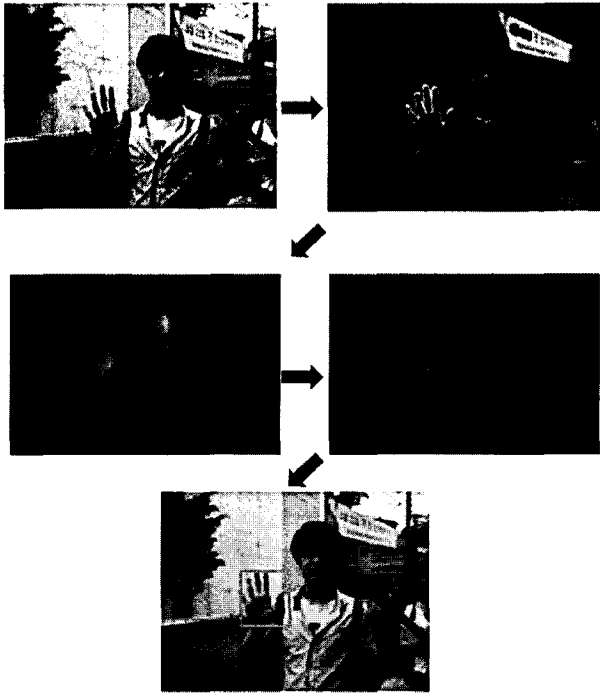


그림 6. 상관계수 지도를 통한 손 검출 과정
Fig. 6. Hand detection processing through correlation coefficient map.

그림 6를 통하여서 상관계수 지도를 만드는 과정을 알 수 있다. 왼쪽 위에 있는 영상부터 화살표 방향대로 처음 영상이 원 영상이고 다음 영상이 피부색을 검출한 영상이다. 세 번째 영상은 두 번째 영상을 그레이 스케일로 바꾼 후 상관계수 지도로 처리한 영상이고 네 번째 영상은 세 번째 영상에서 임계값이상의 영역 중에서 주요 방향 에지 기술자를 통해 변환한 영상의 상관계수 지도를 나타낸 것이고 마지막으로 이중에 최대값을 손 영역으로 검출한다.

3. 손 추적 과정

손 영역을 검출하는 것을 시점으로 하여 손을 추적한다. 손을 추적하는 것은 손 검출과는 다르게 움직일 때

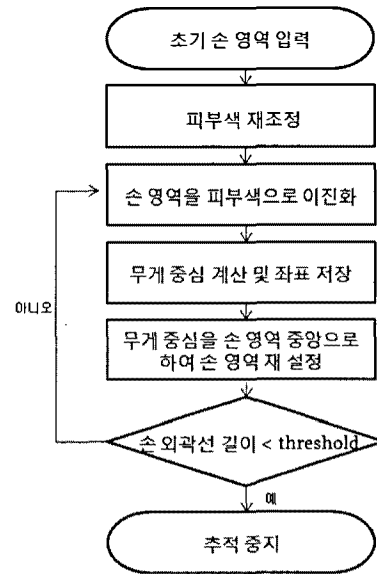


그림 7. 손 추적 과정
Fig. 7. Hand tracking processing.

자유도를 충분히 허용하는 방법을 사용하였다. 손을 찾으면 그 영역의 피부색을 YCbCr 범위로 환경에 맞게 재조정을 한다. 또한 처리 속도를 줄이기 위해 손 검출 후에는 그 영역 주위만을 변환하고 연산한다. 이 영역의 범위는 손 영역의 무게 중심을 추적하여 그 위치가 변할 때마다 재조정된다. 추적과정은 손의 외곽선 길이가 임계값보다 작으면 추적완료를 하고 손으로 그려진 패턴의 분석을 시작한다.

가. 피부색 미세 재조정

피부색을 미세 재조정 하는 이유는 손 추적에서는 손의 자유도를 위해 에지 정보를 사용하지 않고 피부색 정보만을 이용하므로 좀 더 세밀한 피부색 범위가 필요하기 때문이다. 그리고 피부색은 조명이나 광원의 위치 또는 개수에 영향을 받기 때문에 식 (2)와 같이 범용적인 피부색 범위에서 환경 적응적인 재조정이 필요한 것이다. 방법은 손이 검출되면 그 영역 중 식 (2)에서 정의한 피부색 범위에 해당하는 픽셀들의 Y, Cb, Cr 각각의 최대값과 최소값을 구한다. 그 범위로 피부색을 재조정한다. 이는 피부색의 정의를 큰 범위에서 더 작은 범위로 줄여준 것이다.

나. 무게중심 추적

손 추적 방법으로는 무게중심을 찾아가는 방법을 선택하였다. 무게중심을 구하는 방법은 다음과 같다.
① 손 영역으로 정해진 사각영역에서 피부색에 해당하

는 픽셀 부분을 1이라하고 나머지 부분은 0이라고 두고 이진화를 시킨다.

- ② 이진화 시킨 사각영역의 1이 되는 부분의 X 좌표값의 합과 Y좌표값의 합을 구하고 각각을 1에 해당되는 픽셀의 수으로 나누어 주면 무게중심의 좌표값을 구할 수 있게 된다.

한 프레임에서 무게중심이 구해지면 그 좌표를 중심으로 사각영역을 재조정하여 준다. 이러한 방법으로 프레임 마다 반복을 하게 되면 그림 8과 같이 손을 추적할 수 있게 된다.

이와 같이 손을 추적할 때 공간에 무엇인가 의미 있는 기호나 숫자를 그리게 되는데 영상에서 무게중심에 해당하는 픽셀만으로 자취를 그려나간다면 연속된 선을 얻을 수 없고 데이터화 시키는 데도 어려움을 겪을 수 있다. 따라서 무게 중심을 중심으로 일정한 크기의 원 모양을 범위로 그 자취를 남기게 된다면 연속된 선을 얻을 수 있다. 마치 볼펜의 볼 사이즈를 크게 한다고 생각하면 된다. 하지만 원의 크기가 너무 커지게 된다면 그리는데 있어서 동작이 커질 수 있으므로 적당한 크기를 선택한다. 그리고 추적된 자취는 전체 영상의 크기와 같은 크기의 8bit의 그레이 스케일의 영상에 그려진 자취부분만 255(흰색)로 두고 나머지는 0(검정색)으로 둔다.

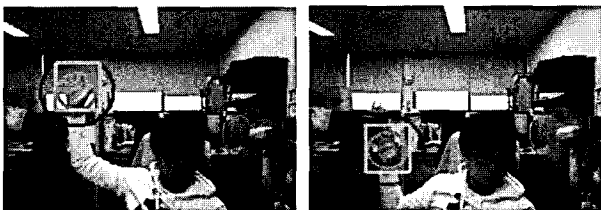


그림 8. 무게중심 추적에 의한 손 영역 검출
Fig. 8. Hand tracking by center of skin color blob.

다. 시작점과 끝점

손 추적의 시작은 손 영역 검출과 동시에 시작이 된다. 본 논문에서는 손가락 다섯 개를 모두 편 상태로 손을 인식하게 하였다. 일정한 사각 영역 안에서 손의 외곽선 길이는 거의 일정하다고 할 수 있다. 본 논문에서는 그 외곽선의 길이를 12명의 손 모양에서 추출하여 그 평균값을 임계값으로 그 이하의 외곽선 길이가 나올 경우에는 추적을 마치게 설정해 두었다. 따라서 주먹을 쥔 상태가 되면 외곽선의 길이가 임계값 이하가 되어 끝을 나타내는 신호가 되는 것이다. 외곽선은 Open-

CV 라이브러리를 사용하여 추출하였다.

4. 필기체 숫자 인식

검출한 손 무게중심 이동궤적의 영역범위를 설정하고 설정한 범위 안에 영상을 64비트의 이진화된 값으로 변환한 후 DNAC 알고리즘으로 미리 저장해 둔 레퍼런스 데이터와의 비교를 통해 사용자가 쓴 숫자가 무엇인지 분석한다.

가. 그려진 영상에 영역 재구성

그림 9와 같이 손 영역 중심 추적에 의하여 그려진 영역에서 가로와 세로의 크기를 구하고 그 영역을 8X8 픽셀 크기로 재조정하고 좌우를 바꾸어 준다. 왜냐하면 사용자 시점에서 문자를 썼기 때문이다. 최종적으로 레퍼런스 데이터와 비교하기 위해 흰색 픽셀이 있는 부분의 칸을 1로 두고 나머지는 0으로 둔다. 이 과정은 손으로 그려진 영역을 찾아 재구성하기 때문에 카메라 프레임 안 어느 공간에 써도 인식을 할 수 있다는 것이 이 시스템의 장점이다.

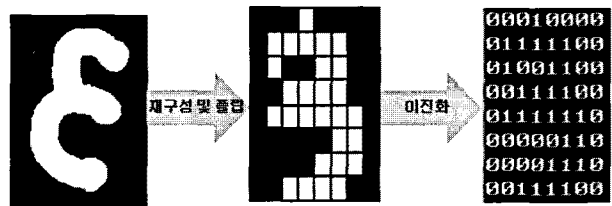


그림 9. 그려진 숫자 영역 재구성
Fig. 9. Reconstruction image of drawing region.

나. DNAC 알고리즘

DNAC 알고리즘은 DNA의 상보 결합의 특성을 이용한 생화학적인 DNA 컴퓨팅에서 모티브를 가져온 알고리즘이고 패턴 분류기로써의 그 성능은 이미 여러 논문을 통해 입증되어 있다.^[17-19] 이 알고리즘의 장점은 기존의 패턴 분류기가 학습에 오랜 시간이 걸렸다면 이 알고리즘은 학습 과정이 없이 처리할 수 있는 장점을 가지고 있으며 반복적인 루프 연산이 필요하지만 몇 개의 위치만을 비교함으로써 반복으로 인한 연산속도의 단점을 보완할 수 있다. 또한 하드웨어로의 구현이 비교적 간단하여 비교할 데이터가 많아질 경우 하드웨어나 GPU로 구현하여 연산 속도를 빠르게 할 수 있다.

DNAC 알고리즘 처리 방법은 다음과 같다.

퍼런스 패턴은 m개의 샘플로 이루어져 있고 길이는

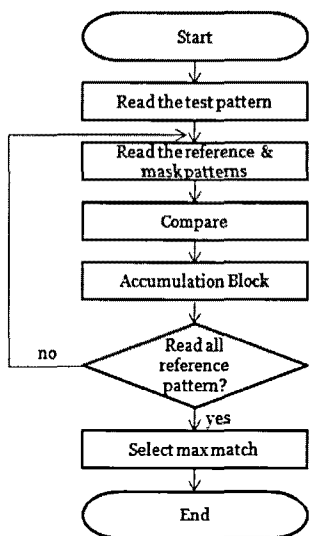


그림 10. DNAC 알고리즘 순서도^[17]
 Fig. 10. Block diagram of DNAC algorithm^[17].

n으로 되어 있다고 가정할 때 클래스를 확인하고자 하는 입력패턴이 들어오면,

- ① 패턴의 길이 n에 해당하는 2진 난수를 생성한다. 이를 mask라고 한다.
- ② 레퍼런스 패턴 샘플 1개와 mask의 1에 해당하는 위치의 입력패턴을 비교하여 준다. 여기서 비교한 위치의 개수를 오더(order)라고 한다. 이 오더는 일정한 개수로 유지 시켜준다.
- ③ 입력 패턴별 비교된 모든 데이터가 일치할 경우 레퍼런스 패턴에 해당하는 클래스의 계수기를 하나 증가 시켜준다.
- ④ ①~③을 모든 레퍼런스 패턴 샘플 m개에 적용한다.
- ⑤ ①~④를 적당한 수로 반복을 한 후 가장 높은 계수기 값을 가지는 클래스를 입력패턴의 클래스로 출력한다.

III. 실험

1. 손 검출 평가

가. 실험 환경

제안한 손 인식 알고리즘을 실험하기 위하여 손 영상 12장, 손을 포함한 영상 35장을 사용하였고 CPU(Intel Core TM2 Quad Q6600 2.4GHz), RAM(3.5G)의 PC환경에서 프로그램은 C와 Matlab을 사용하였다. 그리고

OpenCV library와 Libsvm을 사용하였다.

손만 있는 영상은 검정색 배경에 오른손만 찍은 영상이고 손을 포함한 영상은 여러 다른 배경과 다른 조명 아래에 찍힌 오른손이 포함된 영상이다.

나. 실험 결과

실험을 통하여 두 가지의 경우를 비교해 보았다. 첫 번째는 표 2에서와 같이 손 모델 영상을 학습하고 SVM으로 분류한 경우와 별도의 학습 과정이 없는 제안한 알고리즘으로 분류한 경우의 비교이다. SVM방법은 12장의 손 영상을 학습시키고 35장의 손을 포함한 영상에 적용한 결과 82%라는 검출률을 보였다. SVM은 Libsvm 라이브러리를 사용하여 Matlab으로 동작시켰다. 손 영상 학습은 주요 방향 에지 기술자로 변환 후 학습을 시켰고 학습된 SVM 모델을 슬라이딩 윈도우 방법을 통하여 손을 포함한 영상에 왼쪽 위부터 한 픽셀씩 차례로 이동하여 적합도가 가장 큰 영역을 손 영역으로 선택한다. 적합도는 Libsvm에서 제공하는 확률 값을 참고하여 선택하였다. 손을 포함한 영상에 손의 크기가 각각 다르므로 여러 크기의 학습 영상을 적용하였을 때 가장 적합하게 나온 경우를 선택하였다. 제안한 알고리즘의 경우는 12장 중 한 가지 영상을 뽑아 주요 방향 에지 기술자로 변환 후 학습 과정 없이 적용하여 보았다. 방법은 그레이 스케일의 템플릿 매칭으로 상관계수 값이 임계값 이상인 부분으로 영역을 줄인 다음 주요 방향 에지 기술자로 변환된 영상을 또 다시 템플릿 매칭 한 후 상관계수 값이 가장 높게 나온 영역을 표시 하였다. 검출률은 SVM과 동일하게 82%가 나왔다. 이 실험으로 사전 학습 과정이 없이도 샘플로 선택한 하나의 손 영상만을 가지고도 동일한 성능의 결과가

표 2. SVM과 제안한 알고리즘 검출률 비교
 Table 2. Comparison of SVM and the proposed algorithm.

	SVM	제안한 방법
인식개수/인식 대상 영상수	29/35	29/35
검출률	82%	82%

표 3. 기존 템플릿 매칭과 제안한 알고리즘 검출률 비교

Table 3. Comparison of the existing template matching and the proposed algorithm.

	템플릿 매칭	제안한 방법
인식개수/인식 대상 영상수	16/35	29/35
검출률	45%	82%

나왔음을 알 수 있다. 이는 또한 학습 방법보다 전처리 과정을 통한 데이터 가공이 중요함을 알 수 있고 그 데이터 가공에서 주요 방향 에지 기술자 성능의 우수함을 확인할 수 있다.

두 번째는 표 3에서와 같이 기존의 템플릿 매칭과 그림 4에서 제안한 방법과 같이 기존 템플릿 매칭에 주요 방향 에지 기술자 영상을 매칭한 결과를 합한 경우의 성능 비교이다. 기존의 템플릿 매칭 방법은 45%라는 낮은 검출률을 보였고 제안한 알고리즘인 기존 템플릿 매칭 방법으로 영역을 줄인 후 주요 방향 에지 기술자로 변환한 영상을 다시 템플릿 매칭을 하여 최대값을 구한 경우는 82%라는 검출률을 보였다. 그림 12는 피부색 검출을 통해 피부색이 아닌 픽셀은 검정색으로 둔

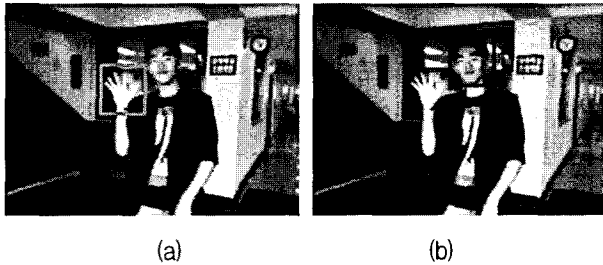


그림 11. (a) 제안한 알고리즘과 (b) 기존 템플릿 매칭 실패 사례

Fig. 11. (a) Proposed algorithm and (b) existing method template matching faulty.



그림 12. 외부 환경으로 인한 피부색 검출 실패

Fig. 12. Fail to detect skin color region of hand owing to external environment.

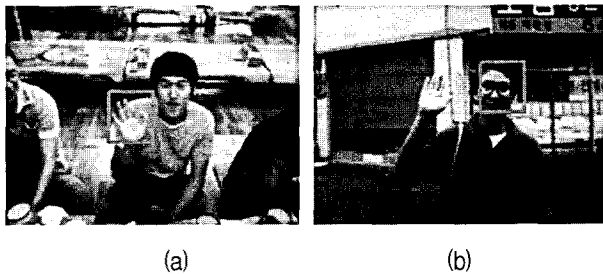


그림 13. 제안한 알고리즘 검출 (a) 성공과 (b) 실패 사례

Fig. 13. Example of (a) success and (b) failure to detect hand through the proposed algorithm.

경우인데 이와 같이 조명이나 외부 환경으로 인해 피부색 검출에 대한 오류가 생기는 경우가 있다. 이 때 그림 11(b)와 같이 기존의 템플릿 매칭은 색에 대한 의존도가 높기 때문에 얼굴을 손 영역으로 검출하지만 그림 11(a)와 같이 제안한 방법은 피부색과 에지를 모두 고려하기 때문에 손 영역을 정확하게 검출한다. 하지만 제안한 알고리즘의 경우도 회전에 약한 특성을 가지고 있기 때문에 그림 13(b)와 같이 손의 모양이 회전을 하여 샘플 손 모델과 다른 방향을 향해 있다면 검출하지 못하는 경우가 있다.

2. 손 추적 평가

가. 실험 환경

손 검출할 때 피부색을 범위를 다시 설정하는 방법의 성능 평가를 위해 다음과 같은 실험 도구를 사용하였다. PC환경과 개발 툴은 손 검출 평가 환경과 동일하다. 그리고 추적을 위한 카메라는 PC Cam(Logitech Quickcam Communicate Deluxe, 130만 화소, 30프레임/초) 을 사용하였다.

나. 실험 결과

그림 14(a)는 식 (2)에 해당하는 피부색 범위를 사용



그림 14. 피부색 범위를 (a) 재조정하지 않은 경우와 (b) 재조정된 경우 추적 비교

Fig. 14. Comparison with (a)case of non-readjusting skin color range and (b)case of readjusting skin color range.

한 경우이다. 피부색과 비슷한 물체를 만났을 때 그림 14(a)과 같이 오인식 하는 경우가 생긴다. 이와 같이 무게 중심을 따라 손 영역을 추적하는 방법은 회전에 강한 장점을 가진다. 하지만 예지 정보를 활용하지 않음으로 인하여 피부색과 비슷한 색을 가진 물체를 만났을 때 오인식을 일으키는 단점이 발생할 수 있다. 그래서 그림 14(b)와 같이 손을 검출할 때 피부색의 범위를 재조정하면 식 (2)에 해당하는 피부색 범위보다 더 적합한 범위를 가지므로 비슷한 색을 가진 물체를 구분할 수 있다.

3. 필기체 숫자 인식

가. 실험 환경

공간에서 카메라를 통한 문자 입력 방법의 성능을 평가하기 위하여 PC 환경과 개발 툴을 위와 같은 PC Cam을 사용한다. 레퍼런스 데이터는 UCI Machine Learning repository^[20]에서 Optical Recognition of Handwritten Digits Data을 사용하였고 형태는 다음과 같다. 필기체 숫자 영상은 각각 8X8의 셀로 나누고 각 셀마다 쓰여진 영역이 얼마나 차지하냐에 따라 0에서 16까지 정수로 된 그레이 스케일 값으로 되어 있다. 셀마다 임계값을 정하여 임계값보다 크면 1 아니면 0으로 이진화하여 각 숫자별로 64 비트로 저장한다. 그리고 마지막 65번째 비트에는 해당 숫자를 쓴 사람이 의도한 숫자가 무엇이었던지 저장되어 있다. 레퍼런스 데이터 샘플의 총개수는 3823개이고 각 숫자별로 380(±9)개로 거의 균등한 개수를 이루고 있다.

나. 실험 결과

각 숫자 마다 10회씩 총 100회 시행을 통해 평균 90%의 인식률을 보여주었고 그림 15와 같이 인식에 있

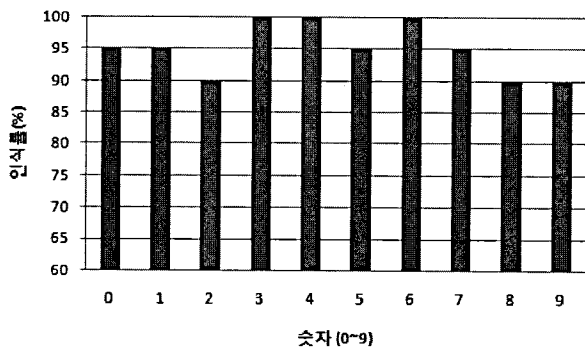


그림 15. 숫자 데이터 숫자별 인식률 비교
Fig. 15. Comparison recognizing ratio of each number.

어서 숫자마다 조금씩 차이가 있었다. 특히 2, 8, 9의 경우 쓰는 방식에 따라 민감하게 반응 하여 2의 경우 1이나 9로 오인식 되는 경우가 있었다. 8이나 9의 경우도 0이나 1로 오인식 되는 경우가 있었다. 이는 기하학적으로 비슷한 모양을 가지고 있는 경우 오인식이 있는 것으로 보인다.

IV. 결 론

본 논문에서는 손의 피부색 정보와 형태 정보를 결합하여 사용함으로써 기존의 피부색 정보가 주를 이룬 손 검출 방법보다 성능을 향상 시켰고 형태 정보를 구하는데 있어서 주요 방향 예지 기술자라는 방법을 이용하여 손 영상의 사전 학습 과정 없이도 손 영역을 검출함을 확인할 수 있었다. 또한 추적에 있어서는 검출 후 피부색 범위 재조정을 통해 회전에 강하지만 피부색과 비슷한 색에 대한 오인식을 줄였다. 마지막으로 3차원 공간에 손으로 필기체 숫자를 써서 인식하는 실험을 통하여 단일카메라를 이용한 손동작 입력장치로서의 가능성을 보여 주었다. 이는 시계열 데이터로 분류되는 제스처 인식 문제를 은닉 마르코프 모델과 같은 방법을 사용하지 않고 연산에 있어서 간단한 방법으로 동작의 크기나 시간에 상관없이 처리 하였다는 데에 또한 큰 의미가 있다. 하지만 TV나 휴대폰과 같은 기기에 실용화가 되기 위해서는 좀 더 많은 사람과 환경을 통한 검증이 있어야 할 것이다.

참 고 문 헌

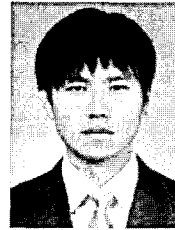
- [1] A.Jaimes, N.Sebe, "Multimodal Human Computer Interaction: A Survey," LNCS, vol. 3766, pp.1~15, 2005.
- [2] N.Sebe, I.Cohen, T.Gevers, T.Huang, "Multimodal Approaches for Emotion Recognition: A Survey," Proceedings of the SPIE, Volume 5670, pp. 56-67, 2004.
- [3] 배창석, 전병태, 윤호섭, 민병우 "손의 이동 궤적 분석에 의한 제스처 인식," 전자공학회 워크샵, 제 8권, pp.144~148, Jan. 1996.
- [4] Pragati Garg, Naveen Aggarwal, Sanjeev Sofat, "Vision Based and Gesture Recognition.," PWASET, volume 37, 2009.
- [5] A. Mulder, "Hand gestures for HCI," Technical Report 96-1, vol. Simon Fraster University, 1996
- [6] B. Stenger, A. Thayananthan, P.H.S. Torr, R.

- Cipolla, "Model-based hand tracking using a hierarchical Bayesian filter," IEEE transactions on pattern analysis and machine intelligence, September 2006.
- [7] 고민삼, 이광희, 김창우, 안준호, 김인중, "비전 기반 제스처 인식을 이용한 인터페이스 구현," 한국컴퓨터종합학술대회 논문집, vol. 35 No.1, 2008.
- [8] Tarek M. Mahmoud, "A New Fast Skin Color Detection Technique," PWASET, vol.33, September 2008.
- [9] R., Gonzales, and E., Woods, "Digital Image Processing," Prentice Hall, Inc, New Jersey, 2002.
- [10] B.D., Zarit, B.J., Super, and F.K.H. Quek, "Comparison of five color models in skin pixel classification," In Int. Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems, pages 58-63, Corfu, Greece, Sept. 1999.
- [11] D. Chai, and K.N. Ngan, "Face segmentation using skin-color map in videophone applications," IEEE Trans. on Circuits and Systems for Video Technology, 9(4): 551-564, June 1999.
- [12] W. T. Freeman and M. Roth. "Orientation histograms for hand gesture recognition," Intl. Workshop on Automatic Face and Gesture Recognition, IEEE Computer Society, Zurich, Switzerland, pages 296-301, June 1995.
- [13] K. Hotta, "Scene Classification Based on multi-resolution Orientation Histogram of Gabor Features," ICVS 2008, LNCS 5008, pp.291~301, 2008.
- [14] P. Buehler, M. Everingham, A. Zisserman, "Learning sign language by watching TV (using weakly aligned subtitles)," IEEE CVPR 2009.
- [15] N. Dalal and B. Triggs. "Histogram of oriented gradients for human detection." In Proc. CVPR, 2005.
- [16] 서승원, 선우명훈, "Template Matching을 위한 새로운 알고리즘 및 ASIC칩 구현," 전자공학회 논문지, 제3권, C편, 제1호, pp.15~24, Jan. 1998.
- [17] 권오혁, "DNA 특성을 모방한 디지털 패턴인식 하드웨어 설계," 인하대학교 정보통신공학과 석사학위논문, 2008.
- [18] 최선욱, "Multi-class 데이터 분류를 위한 DNA 컴퓨팅 기반의 패턴인식 하드웨어 설계," 인하대학교 정보통신공학과 석사학위논문, 2008.
- [19] J.-K. Kim, B.-T. Zhang, "Evolving hypernetworks for pattern classification," IEEE Congress on Evolutionary Computation (CEC 2007),

pp.1856~1862, 2007.

- [20] University of California Irvine, Machine Learning repository, <http://archive.ics.uci.edu/ml/>

— 저 자 소 개 —



윤진현(학생회원)
2008년 인하대학교 전기공학
학사
2008년~현재 인하대학교
정보통신공학과 석사과정.
<주관심분야 : 패턴인식, 영상처리>



이종호(정회원)
1976년 서울대학교 전기공학
학사
1978년 서울대학교 전자공학
석사
1986년 미국 아이오와 주립대학교
전기 및 컴퓨터공학 박사
1986년~1989년 미국 노틀담 대학교 조교수
1997년~1998년 인하대 집적회로설계센터 소장
2004년~2005년 브라운 대학교 두뇌 및
신경회로만 연구소 방문교수
1989년~현재 인하대학교 정보통신공학과 교수
2000년~현재 수퍼지능기술연구소 소장
<주관심분야 : 지능형 시스템, VLSI 설계,
Quality of life technology>