

PCA 기반의 이진 분류기와 손 추적을 이용한 손 검출

김진국*⁰ 민경원** 정찬기** 고훈석*

*고려대학교 전자컴퓨터 공학과

**전자부품 연구원 디지털미디어 연구센터

jjkim@ispl.korea.ac.kr⁰, { minkw, ckjung }@keti.re.kr, hsko@korea.ac.kr*

Hand Detection using PCA based Binary Classifier and Hand Tracking

Jinkuk Kim*⁰ Kyungwon Min** Chanki Jung** Hanseok Ko*

*Department of Electronics and Computer Engineering, Korea University

**Digital Media Research Center, Korea Electronics Technology Institute

요 약

본 논문에서는 인간과 컴퓨터 사이의 상호작용을 하는 방법중의 하나인 제스처를 인식할 때 필요한 정확한 손 검출 방법을 제안한다. 이를 위해 기존의 다수의 손 영상들 가장 잘 표현하면서도 효과적으로 압축할 수 있는 PCA를 이용해서 특징 벡터를 추출한다. 이어서 특징 벡터간의 Mahalanobis distance를 이용한 분류기에 가중치를 적용하여 사용한다. 또한 시간에 따른 연속적인 영상에서 검출된 이전 영상의 중심점의 위치와 중심점의 motion vector를 이용해서 손이 검출되지 않은 영상의 검출 성능을 보강한다.

1. 서 론

오늘날 인간과 컴퓨터가 정보를 주고 받는 방법은 여러 가지가 있다. 그런데, 정보를 주고 받을 때 접촉 없이 단지 인간의 몸짓이나 손짓 만으로 원하는 정보를 컴퓨터에게 전달하는 효과적인 방법이 제스처 인식이다 [1].

제스처를 인식하기 위해서는 우리가 인식할 대상의 위치 정보가 필요하다. 본 논문에서는 그 대상이 손이며 만약 입력 영상에 손만 포함되어 있다면 손의 색깔이나 모양과 같은 비교적 간단한 특징만으로도 손의 위치를 알아낼 수 있다. 그러나 손의 색깔이나 모양이 비슷한 후보가 여러 개 존재할 때 손을 구별해 내려면 후보들 중에 손이 어떤 것인지 알려줄 수 있는 분류기가 필요하다. 또한 연속된 입력 영상에서 손의 위치의 궤적에 대한 정보가 필요하므로 중간 중간에 손이 검출되지 않은 영상에 대한 보강이 필요하다.

따라서 본 논문은 안정적인 손 검출 성능을 얻기 위한 2가지 방법을 제안한다.

전처리 과정으로 먼저 피부색을 기반으로 후보영역을 찾는다.

첫째로 그 영상을 PCA를 이용하여 특징 벡터 추출한 뒤에 훈련된 특징벡터와의 Mahalanobis distance를 이용한 분류기를 설계한다. 이때 보다 높은 검출 성능을 얻기 위해서 분류기의 성분들에 가중치를 다르게 하는 방법을 사용한다.

둘째로 들어 오는 입력 영상이 시간에 따른 연속된 영상일 때 손 추적 기법을 통해서 중간에 손 검출에 실패한 영상에서도 이전영상에서 검출된 손 영역의 중심점의 위치와 중심점의 Motion vector를 이용하여 손 검출 성능을 보강하는 방법을 사용한다.

그림 1은 두 가지 제안한 방법을 통해서 구성된 손 검출 시스템의 과정을 보여준다. 이와 같은 과정에 따라서 2절에서는 PCA 특징을 추출하는 과정 나타내고, 3절에서는 분류기를 설계하는 방법에 대해서 4절에서는 손 추적으로 손 검출 성능을 보강하는 방법에 대해서 다룰 것이다.

2. 특징 추출

2.1 주요 성분 분석 (Principle Component Analysis)

주어진 영상을 표현하는 적절한 양의 특징을 추출하기 위해서 사용되는 방법 중에서도 주요 성분 분석(Principle Component Analysis)은 주어진 data의 정보를 잘 표현하면서도 그 정보량을 줄일 수 있는 효과적인 방법 중 하나이다 [2]. 관측된 data(영상 입력)를 벡터로 표현할 때 하나의 영상은 다음과 같이 표현된다. 영상의 pixel의 개수가 n 이라 할 때 다음과 같다.

$$x_i = [x[1] \ x[2] \ \dots \ x[n]] \quad (1)$$

여기서 i 는 영상이 들어온 순서를 가리킨다. 이렇게 들어온 입력 영상이 시간의 따라 여러 개가 있을 때 이것을 모아둔 전체 data를 m 개 영상에 대해 한꺼번에 다음과 같이 표현 할 수 있다.

$$X = [x_1 \ x_2 \ \dots \ x_m]^T \quad (2)$$

이 때 X 의 공분산 행렬을 구하면 다음과 같다.

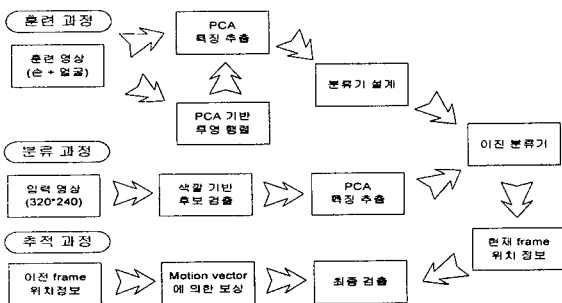


그림 1. 손 검출 시스템

$$\mathbf{m}_x = \sum_{i=1}^m \mathbf{x}_i \quad (3)$$

$$\mathbf{M}_x = [\mathbf{m}_x \ \mathbf{m}_x \ \dots \ \mathbf{m}_x] \quad (4)$$

$$\mathbf{C}_x = E[(\mathbf{X} - \mathbf{M}_x)(\mathbf{X} - \mathbf{M}_x)^T] \quad (5)$$

위의 (5) 식의 결과로부터 이 공분산 행렬의 고유 값 행렬과 고유 벡터 행렬을 얻을 수 있다. 이때 고유 값의 크기에 따라 정렬된 고유 값 벡터와 고유 벡터를 구하면 아래와 같다.

$$\mathbf{C}_x \mathbf{V} = \mathbf{V} \mathbf{D} \quad (6)$$

\mathbf{D} : diagonal matrix $\lambda_1 > \lambda_2 > \dots > \lambda_n$

$$\mathbf{V} = [\mathbf{e}_1 \ \mathbf{e}_2 \ \dots \ \mathbf{e}_n]$$

여기서 \mathbf{V} 는 고유 벡터 행렬, \mathbf{D} 는 고유 값 행렬을 나타낸다. 이렇게 얻은 고유 벡터를 basis로 하는 원래의 영상 data를 더 잘 나타내는 새로운 축을 구할 수 있다. [3]. 이때 정보량을 줄이기 위해서 고유벡터의 열 벡터 전체를 사용하지 않고, 적은 개수의 열 벡터만 선택할 수 있다. 이때 PCA는 이론적으로 최소의 평균 제곱 에러를 갖는 축을 선택하게 해준다.[4]

2.2 특징벡터 수의 설정

p 개 ($p < m$)만큼의 고유 벡터만 사용할 때, 새로운 투영 행렬을 다음과 같이 정의할 수 있다.

$$\mathbf{P} = [\mathbf{e}_1 \ \mathbf{e}_2 \ \dots \ \mathbf{e}_p]^T \quad (7)$$

투영행렬 \mathbf{P} 를 사용해서 임의의 입력 n pixel의 영상에 대한 특징을 추출할 수 있다. 임의의 n pixel의 영상 \mathbf{u} 에 대해서 특징 벡터는 다음과 같은 식으로 표현 할 수 있다.

$$\mathbf{f}_u = [f_u[1] \ f_u[2] \ \dots \ f_u[p]] = \mathbf{P}\mathbf{u} \quad (8)$$

이 특징 벡터는 원래의 영상보다 정보량이 적지만, 원래의 영상을 잘 표현할 수 있다. 아래와 같이 고유 값의 비를 이용하면 특징 벡터의 수 p 를 결정할 수 있다.

$$\frac{\sum_{i=p+1}^m \lambda_i}{\sum_{i=1}^p \lambda_i} < T \quad (9)$$

만약 $T = 5\%$ 라면 원래의 정보를 잘 표현 한다고 할 수 있다 [3].

3. 분류기의 설계

훈련 부분에서 손과 얼굴에 대한 훈련 영상들의 특징 벡터를 얻을 수 있다. 이 훈련 영상들의 특징 벡터와 후보 영상의 특징 벡터와의 Mahalanobis distance 차이를 비교함으로써 후보 영상이 손인지 얼굴인지를 판단할 수 있다 [5].

$$\mathbf{f}_{\text{mean}(k)} = \frac{1}{N_k} \sum_{i=1}^{N_k} \mathbf{f}_{\mathbf{u}(k,i)} \quad (10)$$

$$f_{\text{var}(k)}[j] = \frac{1}{N_k} \sum_{i=1}^{N_k} f_{\mathbf{u}(k,i)}[j]^2 - f_{\text{mean}(k)}[j]^2 \quad (11)$$

여기서 k 는 각각의 그룹을 나타내고 N_k 는 각 그룹 훈련 영상의 개수를 나타낸다. $f_{\mathbf{u}(k,i)}[j]$ 는 k 번째 그룹의 i 번째 영

상의 특징 벡터의 j 번째 값을 나타낸다. 이렇게 해서 평균 특징 벡터와 분산의 특징 벡터를 각각의 그룹에 대해서 구한 뒤에 후보 영상이 하나씩 들어 올 때 마다 특징 벡터를 추출하고, 분류 할 수 있다. i 번째 영상에 대해 다음과 같이 구한다.

$$R_k = \sum_{j=1}^p \frac{f_{\mathbf{x}_i}[j] - f_{\text{mean}(k)}[j]}{f_{\text{var}(k)}[j]} \quad (12)$$

여기서 보다 나은 검출 성능을 얻기 위해서 (12) 식에서 얻은 R_k 에 가중치를 두어서 그 중 최소 R_k 값을 갖는 k 를 찾는 방법을 제안한다. 즉, Decision Rule은 다음과 같이 결정 된다.

$$d = \arg \min_k \{\alpha_k R_k\} \quad (13)$$

여기서 α_k 는 각 그룹별 가중치를 나타낸다.

4. 손 추적을 통한 손 검출 성능 보강

입력 영상이 시간에 따른 연속적인 영상일 때 손 추적을 통해서, 검출 성능을 보강할 수 있다. 두 영상간의 시간차가 작다면 가속도는 0에 가까울 것이다. 이러한 가정하에서 다음과 같이 손 검출 성능을 보강할 수 있다.

손 영역의 중심점의 위치를 $\mathbf{C}(\mathbf{x}_i)$ 라 하면 Motion vector

$\mathbf{M}(\mathbf{x}_i)$ 는 다음과 같이 정의 될 수 있다.

$$\mathbf{M}(\mathbf{x}_i) = \mathbf{C}(\mathbf{x}_i) - \mathbf{C}(\mathbf{x}_{i-1}) \quad (14)$$

이 때 전 frame의 motion vector를 가지고 현재 frame에서 아무 것도 검출을 못하더라도 보강할 수 있다. 보강된 중심점의 위치는 다음과 같이 정해진다.

$$\mathbf{C}_{\text{compensation}}(\mathbf{x}_i) = \mathbf{C}(\mathbf{x}_{i-1}) + \mathbf{M}(\mathbf{x}_{i-1}) \quad (15)$$

보강된 중심점의 위치와 분류기를 통해 얻은 결과를 조합하면 더 나은 검출 성능을 얻을 수 있다. 이 방법은 손의 운동이 등속도 운동일 때 최적의 성능 보강을 얻을 수 있다.

5. 실험 및 결과

5.1 실험 환경

운영체제는 Windows XP에 Software는 Matlab을 사용하여 실험하였고, 훈련 영상은 30*30 크기의 손 영상 725개 얼굴 영상 725개로 1450개의 영상을 사용하였고, 후보영역을 손과 얼굴로 제한하기 위하여 배경이 단순한 연속적인 일련의 영상을 테스트 영상으로 사용하였다. 테스트 영상은 320*240 크기의 영상 540개를 사용하였다.

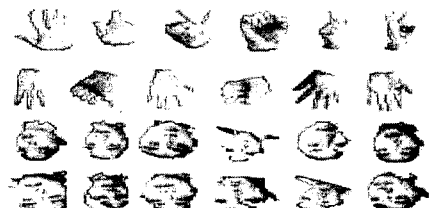


그림 2 사용된 훈련 영상의 예

5.2 실험 결과

5.2.1 훈련 영상에 대한 실험

5.2.1.1 특징 벡터 수 및 가중치의 결정

특징 벡터 수를 식 (9)에서와 같이 고유 값들의 비를 통해서 결정 하였다. $\tau = 2\%$ 일 때 특징 벡터의 수를 계산한 결과 고유 값의 비가 1.92%인 120개의 특징 벡터를 사용하기로 결정하였다. 또한 식 (12)에서 가중치를 결정하기 위해 가중치를 변화시켜 가면서 Equal error rate 을 이루는 지점을 찾아 보았다.

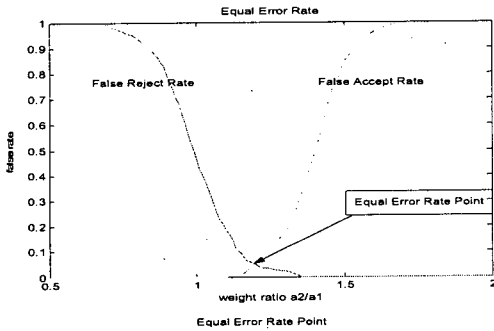


그림 3. Equal Error Rate Point

그 결과 Total Error rate이 최소가 되는 False Accept rate 과 False Reject rate 이 만나는 지점이 $\alpha_2/\alpha_1 = 1.2$ 가 되는 지점이라 는 것을 확인할 수 있었다. α_1 은 손 그룹의 가중치, α_2 는 얼굴 그 룹의 가중치이다.

5.2.1.2 Cross validation

1450개의 DB를 25개의 그룹으로 나누어서 각각의 그룹을 돌아 가면서 테스트 영상으로 사용하였고, 그 나머지를 훈련 영상으로 사용하였다. 각각의 테스트에서 1392개를 훈련영상으로 58개를 테 스트 영상으로 사용하였고 25번 실험에 대한 평균을 내서 아래와 같은 결과를 얻었다. 가중치는 1.2를 사용하였다.

표1. Cross validation 결과

False Reject rate	False Accept rate	Detection rate
4.6%	3.9%	95.4%

5.2.2 테스트 영상에 대한 실험

540개의 DB에 대한 테스트 영상으로 실험한 결과 후보영역이 각 각 2개씩 총 1080개의 후보 영역이 나왔으며, 각각의 후보영역에 대해 분류한 결과 다음과 같은 결과를 얻었다.

표2. 테스트 영상의 손 검출률

False Accept rate	Detection rate
0.2%	78.3%

5.2.3 손 추적에 의한 보상 실험

만약 분류기를 통해서 얻은 후보영역 중에 하나도 없거나, 복수 개가 선택된 경우 원하는 위치 정보를 온전히 얻을 수 없다. 하나도 없는 경우는 이전 프레임에서 위치 정보에 후보영역 중심점의 Motion vector를 더해 주는 방식으로 보상하였고, 복수 개의 경우 보상된 위치와 차이가 적은 것 하나만 선택하도록 하였다. 이 경우 인식률은 다음과 같이 정의 된다.

$$\text{인식률(\%)} = \frac{\text{손의 위치를 올바르게 찾은 영상}}{\text{테스트 영상의 수}} \times 100\% \quad (16)$$

여기서 찾은 영역의 중심 점에 손이 포함되어 있으면 올바르게 찾은 것으로 간주하였다. 이때 인식률은 다음과 같았다.

표3. 손 추적 전과 후의 손 인식률 비교

손 추적 전	손 추적 후
78.1%	90.2%

손 추적으로 손 인식률이 보상되었음을 확인할 수 있었다.

6. 결론

본 논문에서는 PCA를 이용한 이진분류기와 손 추적으로 손 검출 하는 방법 즉, 얼굴이나 손과 같은 피부색의 후보 군이 2개 존재하 는 경우에 대하여 손의 위치를 찾는 방법을 소개하였다.

이렇게 얻어진 손의 위치정보는 제스처 인식을 하는 특징으로서 이용할 수 있다. 이러한 손의 궤적을 이용한 제스처 인식을 통해 컴 퓨터는 인간의 손짓을 정보로 받아들이고 사람과 상호 작용을 할 수 있다.

Acknowledgement

본 연구는 산업 자원부 차세대신기술개발사업의 다중 생체정보 기 반 실감형 디바이스 개발 과제의 지원을 받아 수행되었습니다.

참고문헌

- [1] Vladimir I. Pavlovic, " Visual Interpretation of Hand Gestures for Human-Computer Interaction : A Review" , IEEE Transactions on Pattern Analysis and Machine Intelligence vol. 19, 677-695 [1997]
- [2] 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, 103-107, [1990]
- [3] D. Swets, " Using discriminant eigenfeatures for image retrieval" ; IEEE Transaction on Pattern Analysis and Machine Intelligence vol. 18, 831-836 [1996]
- [4] M. Turk and A. Pentland , " Face Recognition Using Eigenfaces" , IEEE Conference on Computer Vision and Pattern Recognition, 586-591 [1991]
- [5] Richard O. Duda, " Pattern Classification" 2nd Edit. [2001]