

필기습관 정보에 기반한 온라인 서명인식

성한호⁰, 이일병

연세대학교 컴퓨터과학과

생체인식연구센터(BERC)

{hanho⁰, yblee}@csai.yonsei.ac.kr

On-line Signature Identification Based on

Writing Habit Information

Hanho Sung⁰, Yillbyung Lee

Dept. of Computer Science

Biometrics Engineering Research Center, Yonsei University

요 약

생체인식 기술은 현재까지 많은 발전을 거듭하고 있으며 국내에서도 연구는 물론 표준화작업 및 데이터 베이스 구축이 활발히 진행되고 있다. 생체인식은 신체의 여러 부분을 이용하는 방법과 습관에서 비롯된 특징을 이용하는 방법이 있는데, 본 연구에서는 이 중에서 개인의 필기습관 정보를 이용하여 인식하였다. 본 연구에서는 필기습관에 주목하여 서명하는 사람의 습관이 잘 드러나는 펜의 기울임과 눌림, 펜의 방위각도 등의 성분이 표현되어지는 동적인 생체정보를 감지하고 특성을 추출할 수 있는 타블렛과 펜을 사용하여 서명정보를 추출한다. 이렇게 생성된 서명정보의 특징을 추출하기 위하여 패턴인식분야에 널리 활용하고 있는 주성분요소분석(PCA, Principal Component Analysis), 독립성분요소분석(ICA, Independent Component Analysis) 기법에 적용하였다.

생성된 두 특징벡터 사이의 거리를 Euclidean Distance를 이용하여 구하고 Nearest Neighbor를 비교하여 인식률을 알아보고 교차인식(Cross Validation) 기법 중 하나인 Leave-One-Out 방법을 이용한 분류 성능 측정을 통하여 데이터의 신뢰수준을 알아보았다.

1. 서 론

국내 생체인식 기술은 현재까지 많은 연구와 함께 발전을 거듭하고 있다. 이미 알려진 바와 같이 생체인식은 지문, 얼굴, 홍채, 망막, 정맥 등과 같은 신체의 일부분을 인식하는 방법이 있고 타이핑 습관, 걸음걸이 습관, 필기 습관 등과 같이 개인의 습관을 이용하는 방법이 있다.

본 연구에서는 필기습관에 주목하여 서명을 할 때 펜을 눌러 쓴 정도, 펜을 사용하는 위치, 펜을 얼마나 뉘어서 쓰는지 세워서 쓰는지, 펜을 오른손으로 잡는지 왼손으로 잡는지 등의 방위각과 기울임 정도에 대한 동적인 생체 정보를 감지하고 특성을 추출할 수 있는 타블렛과 펜을 사용하여 서명정보를 추출한다.

다섯 종류의 서명정보들은 각기 다른 범위로 값이 추출되므로 특징추출 시 용이하도록 0~255 사이의 실수 값으로 정보의 왜곡 없이 범위를 맞추고 보간법(Interpolation)을 이용하여 서명정보를 정규화 한다. 이렇게 추출된 서명정보를 현재 패턴인식 분야에 널리 활용하고 있는 주성분요소분석(PCA, Principal Component Analysis), 독립성분요소분석(ICA, Independent Component Analysis) 기법에 적용하여 특징을 추출한다.

추출된 생체정보를 한 서명에 대해 1행 N열의 특징벡터(20회의 서명 총 10개씩, 200×4000)를 학습데이터와

테스트데이터로 구성하여 PCA 과정을 거치면 입력데이터의 특징정보가 주성분으로 압축(200×196)되고 이렇게 추출된 특징정보들이 ICA를 통해 주된 축을 중심으로 성분들이 모이게 된다. 이렇게 함으로써 정보를 보다 효율적으로 이용할 수 있다.

생성된 두 특징벡터 사이의 거리를 Euclidean Distance를 이용하여 구하고 Nearest Neighbor를 비교하여 인식률을 알아본 결과 PCA는 87.5%, ICA는 88.5%의 인식률을 보였으며, PCA 과정을 거쳐 나온 200×196차원의 데이터에 대해 값의 변화가 많은 부분을 선택적으로 실험한 결과 196개의 특징차원에서의 인식률과 60개의 특징차원에서의 인식률이 동일하게 나왔으며 그 이후부터는 점차로 감소하였다. 이와 비교하여 ICA 적용 실험에서는 PCA 과정을 거쳐 압축된 196차원 중에서 22차원을 선택한 경우가 가장 높은 인식률이 나왔으며 이후는 점차 감소하는 결과를 보였다. 수집한 데이터의 신뢰수준을 알아보기 위해 교차인식(Cross Validation) 기법 중 하나인 Leave-One-Out 방법을 이용하여 분류 성능을 측정 한 결과 84.8%의 분류율을 얻었다.[7]

2. 본 론

현재까지 온라인 서명인식에서 사용되었던 방법의 공통적인 점은 특징추출과정에서 동적인 정보를 사용하였는데 정보를 가공하는 과정에서 손떨림이나 펜의 미세한 흔들림 등을 잡음으로 보고 제거하거나 기울기 정보의 극심한 변동을 일정 간격으로 표본화(Sampling)하여 데이터를 일정한 간격으로 선택하여 처리에 알맞게 가공하

* This work was supported (in Area-II : Multimodal Biometrics) by Biometrics Engineering Research Center, (KOSEF)

거나 크기를 임의로 적당한 크기로 줄이기도 하였다[김성훈98]. 분절 단위로 비교함으로써 전문 위조에 변별력이 있는 특징을 선택적으로 사용할 수 있다는 장점과 모델에 기반한 입력 서명을 분할하는 방법으로 일관성 있는 서명 분할의 어려움을 극복하기도 하였다[1].

본 논문에서는 현재까지 사용되어 온 방법론들이 필기자의 습관이 고스란히 담겨있는 서명하는 동안의 정보를 임의로 가공하여 일반적으로 잡음으로 처리하던 손떨림 정보들과 펜의 기울임 정보와 방위각 정보의 극심한 변동 등 사용할 수만 있다면 오히려 더 좋은 특징들이 될 수 있는 많은 정보들이 표본화(Sampling) 과정에서 손실을 가져올 수 있다는데 착안하였다. 서명을 시작해서 끝날 때까지의 생체정보를 숫자열의 텍스트로 받은 후 어떠한 왜곡도 없이 받아들인 서명정보를 정규화(Normalization)를 거친 후 '텍스트로 표현된 필기습관의 패턴 열'을 구성하여 근래에 음성인식[2][3]과 얼굴인식[4] 분야에 많이 사용되고 있는 PCA, ICA 기법을 사용하였다. 정보를 효과적으로 줄일 수 있는 PCA를 이용해서 특징을 얻은 후에 축을 달리하면서 특징을 분리하여 인식에 도움이 되는 ICA를 이용하여 정보의 손실을 없애고 정보 이용의 효율성을 높였다. 이렇게 필기습관만의 정보를 가지고도 비교적 높은 인식률을 얻을 수 있음을 보인다[5][6].

2.1 온라인 서명시의 생체정보 추출

온라인 서명 시 발생하는 동적인 움직임과 개개인의 저마다 길들여지고 익숙해진 필기습관은 누구도 모방할 수 없는 고유의 생체데이터이다. 필기 속도, 손에 쥔 펜의 기울임 정도와 글을 쓸 때 나타나는 미묘한 펜 기울기의 변화, 글자를 한 자 한 자 써 나갈 때 변화하는 펜의 눌러 쓴 정도의 변화. 그리고, 글을 쓰고 있는 면에 대해 펜이 어느 위치에서 쓰여지고 있으며 어떤 위치에서 시작하여 어떤 위치에서 서명을 마치는지에 대한 수많은 입력정보들은 모두 온라인 서명검증에 결정적 정보가 되는 생체특징에 대한 정보를 담고 있는 것이다.

펜의 눌림 정도와 펜의 타블렛에 대한 기울임 정도, 펜의 타블렛에 대한 방위각 및 펜의 좌표값(위치정보) 추출에 대한 자세한 내용은 참고문헌의 [5]와 [7]을 참조하기 바란다.

2.2 생체정보의 개인별, 성분별 비교

추출한 서명의 생체정보는 개인별로 독립적이며 각 생체정보의 성분 즉, 방위각, 기울기, 압력, 좌표 정보도 모두 독립적이다.

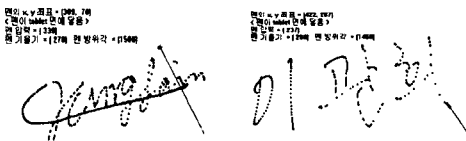


그림 1. 두 사람에 대해 받은 서명의 예

위의 그림 1에서 보듯 서명은 모두 개성이 있음을 알 수

있으며 좌측 상단의 '['] 안에 보이는 수치들이 현재 개인의 필기습관을 나타내며 이 값들에는 특징정보들이 고스란히 담겨있다. 아래의 그림 2에 위의 그림 1의 서명인에 대한 압력정보를 나타내었다.[7]

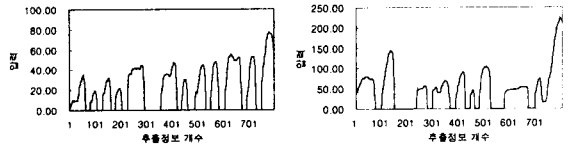


그림 2. 두 사람의 압력정보

위에서 보듯 압력정보는 독립적인 분포를 보인다. 그림에서 y축의 100은 0~255사이의 값으로 표준화한 것으로 약 400의 압력지수를 나타낸다.

2.3 서명정보의 전처리

압력수치가 증가(서명시작)하는 부분부터 더 이상의 입력이 없는 부분까지 데이터를 받아 서명을 시작해서 끝날 때까지 1차 표준화(Normalization)를 거쳐 데이터를 0~255 사이의 값들로 만들고, 아래의 그림 3과 같이 2차 표준화로 보간법(Interpolatin)을 사용하여 정보들을 균일하게 배치하여 생체정보를 추출하였다.[7]

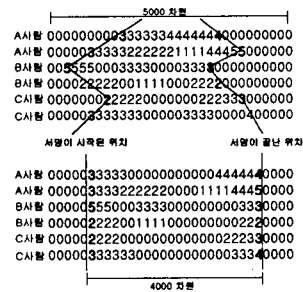


그림 3. 보간법을 이용한 서명정보의 정규화

2.4 PCA에 적용하기 위한 특징벡터 구성

추출한 생체정보들을 PCA와 ICA에 적용하기 위해서는 1행 N열의 구성이 필요하다. 데이터는 모두 5가지의 정보 즉, 방위각, 기울기, 압력, x좌표, y좌표의 정보를 모두 0~255 사이의 실수(소수점 2자리)로 변환된 값들을 각 800개씩 모두 4000개를 받게 된다. 아래의 표 1에 특징벡터의 구성을 나타내었다. 한 행은 한사람에 대한 서명정보를 담고 있다. 이 형태로 벡터를 구성하여 PCA와 ICA에 적용하여 특징을 추출하게 된다.[7]

표 1. 특징벡터의 구성

	방위각 정보	기울기 정보	압력 정보	x좌표 정보	y좌표 정보
1	· · · 800	801 · · 1600	1601 · · 2400	2401 · · 3200	3201 · · 4000
2	· · · 800	801 · · 1600	1601 · · 2400	2401 · · 3200	3201 · · 4000
·	·	·	·	·	·
·	·	·	·	·	·
N-1	· · · 800	801 · · 1600	1601 · · 2400	2401 · · 3200	3201 · · 4000
N	· · · 800	801 · · 1600	1601 · · 2400	2401 · · 3200	3201 · · 4000

2.5 PCA, ICA 기법을 적용한 특징추출

특징추출은 일반적으로 정보의 압축, 차원 감소 과정과 관련되며 이 특징추출에서는 인식을 위한 특징의 좋고 나쁨은 인식률로써 판단된다. 상관을 줄이기 위해 PCA(Principal Component Analysis)를 통해 pre_whitening하거나 filter bank를 더 좋은 성능을 나타낼 수 있도록 하기 위하여 웨이블릿(Wavelet)이나 ICA(Independent Component Analysis)를 이용하는 연구도 활발히 이루어지고 있다.

본 연구에서는 PCA, ICA를 온라인 서명인식에 이용하여 앞서 추출한 통계적으로 독립인 생체특징정보를 분석하고 압축하여 보다 간결하게 개인별 독립 성분으로 분리하여 검증에 사용하여 비교적 높은 인식률을 얻을 수 있음을 보인다.[5] 또한 ICA를 통해 독립된 성분을 찾아내고 분리하여 나타난 성분은 개개인의 결정적 생체특징임을 확인하고 이를 인식에 이용할 수 있음을 보인다.

3. 실험 결과

실험에서는 서명정보를 MS Windows 2000 Professional, MS Visual C++ 환경에서 타블렛(Wacom Intuos Tablet Model, GD-0405-U, USB)과 펜(Intuos Pen, GP-300E-00H)을 이용하여 20명을 대상으로 각각 20회의 서명을 획득하였다. 획득한 서명 중 10개는 학습데이터로, 나머지 10개는 테스트데이터로 사용하였다. 생체정보는 방위각, 기울기, 압력, 좌표(x좌표, y좌표)를 2초간 1000개씩 받아서 앞의 2.4절에서 설명한 형식대로 1×4000(여기서, 4000=5가지정보×800개로 Interpolation)의 배열을 구성하였다. 이렇게 얻어진 200×4000의 학습데이터와 테스트데이터를 Matlab 6.1 환경에서 PCA와 ICA에 각각 적용한 결과로 생성된 특징 벡터사이의 거리를 Euclidean Distance를 이용하여 구하고 Nearest Neighbor를 비교하여 인식률을 알아보았다. 수집한 데이터의 신뢰수준을 알아보기 위해 교차검증(Cross Validation) 기법 중 하나인 Leave-One-Out 방법을 이용하여 분류 성능을 측정하였다.

아래의 표 2에 생체정보의 조합에 따른 인식률 실험결과를 보았다.

표 2. 생체정보의 조합에 따른 인식률 비교

생체 정보		인식률(%)	
		PCA	ICA
1가지	방위각	83.00	80.00
	기울기	79.50	78.50
	압력	82.00	81.50
	좌표값(x, y좌표)	5.50	13.00
2가지	방위각, 기울기	84.50	83.50
	방위각, 압력	84.50	85.50
	기울기, 압력	84.50	84.00
	방위각, 좌표	81.00	82.00
	압력, 좌표	82.00	79.00
	기울기, 좌표	79.50	77.50
3가지	방위각, 기울기, 압력	87.50	88.50
5가지	방위각, 기울기, 압력, 좌표	87.50	87.50

위의 실험에서 PCA 적용 결과는 196차원에 대한 실험

결과이며 ICA는 PCA결과 생성된 196차원 중 인식률이 가장 높았던 22차원에 대하여 실험하였다. 위의 실험 결과에서 보듯 생체정보의 조합 수에 따라 항목별로 구분하였으며, 굵은 글씨로 쓴 숫자는 해당 항목에 대해 가장 높은 인식률을 보인 것이고, 가장 낮은 인식률을 보인 항목에는 밑줄을 그어 표시하였다.

아래의 그림 4에 각 특징차원별 PCA와 ICA의 인식률을 비교하여 나타내었다.

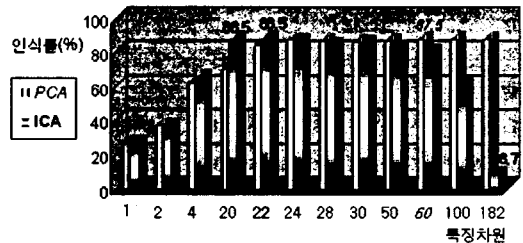


그림 4. 특징차원에 대한 PCA와 ICA의 인식률 비교

3.1 Leave-one-out 교차검증(Cross-validation) 결과

첫 번째 서명데이터를 Test set(1개)으로 정하고 이를 포함한 학습데이터(200개)와 테스트데이터(200개)를 새로운 Train set(400개)으로 만들어 ICA과정을 거친 후, 이 데이터가 해당하는 학습데이터 군으로 분류되는지를 알아보았다. 실험을 위한 테스트데이터는 1개의 서명데이터(학습데이터)를 포함한 모든 서명데이터를 사용하는 데, 이 학습데이터가 테스트데이터 내에 존재하는 자신의 그룹 또는 클래스로 분류되는지를 확인한 결과 84.8%의 분류율을 얻었다.[7]

4. 결론

본 연구를 통하여 개개인의 필기습관이 고스란히 담겨 있는 서명정보를 추출하고, PCA와 ICA를 이용하여 서명데이터의 생체특징을 분리하고 검증데이터로 활용된 결과 비교적 좋은 인식률을 얻을 수 있음을 확인하였다. 이로써 본 연구에서 제안하는 필기습관을 기반으로 한 인식 방법이 온라인 서명 인식에 효율적으로 사용될 수 있음을 알 수 있었다.

참고문헌

[1]T.Rhee, S.Cho and J.Kim, "On-line Signature Verification Using Model-Guided Segmentation and Discriminative Feature Selection for Skilled Forgeries," 6th Int'l Conf. on Document Analysis and Recognition, Seattle, WA, pp. 645-649, Sep. 2001.
 [2]박형민, 정호영, 이태원, 이수영, "잡음 환경에서의 음성 인식을 위한 독립 요소 해석 기법의 효율적 적용", 인공지능, 신경망 및 퍼지 종합학술대회, pp.179-186, 1999.
 [3]이종환, 정호영, 이태원, 이수영, "독립 요소 해석 기법을 이용한 음성 특징 추출", '99 인공지능, 신경망 및 퍼지 학술회의, pp.297-300, 1999.
 [4]곽노준, 최종호, 김유철, 신희정, "ICA에 기반한 얼굴인식", 2001 한국 뇌학회 학술대회 논문집, pp.121-122, 2001.
 [5]성찬호, 기근도, 윤성수, 이일명, "필기습관에 기반한 온라인 서명의 생체정보 추출 및 PCA와 ICA를 이용한 특징 추출," 2002 컴퓨터비전 및 패턴인식 연구회(CVPR) 추계워크샵 논문집, pp.48-50, 2002.
 [6]성찬호, 윤성수, 이일명, "펜 입력정보를 기반으로 한 온라인 서명의 생체정보 추출 및 ICA를 이용한 특징 추출," 한국정보과학회 가을 학술발표논문집, Vol. 29, No. 2, pp.577-579, 2002.
 [7]성찬호, "필기습관에 기반한 온라인 서명 인식", 연세대학교 석사학위논문, 2003.