

Statistical Fingerprint Recognition Matching Method with an Optimal Threshold and Confidence Interval

C. S. Hong¹ · C. H. Kim²

¹Department of Statistics, Sungkyunkwan University

²Research Institute of Applied Statistics, Sungkyunkwan University

(Received October 8, 2012; Revised November 9, 2012; Accepted November 14, 2012)

Abstract

Among various biometrics recognition systems, statistical fingerprint recognition matching methods are considered using minutiae on fingerprints. We define similarity distance measures based on the coordinate and angle of the minutiae, and suggest a fingerprint recognition model following statistical distributions. We could obtain confidence intervals of similarity distance for the same and different persons, and optimal thresholds to minimize two kinds of error rates for distance distributions. It is found that the two confidence intervals of the same and different persons are not overlapped and that the optimal threshold locates between two confidence intervals. Hence an alternative statistical matching method can be suggested by using non-overlapped confidence intervals and optimal thresholds obtained from the distributions of similarity distances.

Keywords: AUC, credit evaluation, cut-off, distance, FAR, FRR, minutiae, similarity, ROC.

1. 서론

지문인식(fingerprint recognition)은 사람마다 가지고 있는 신체의 특징 또는 습관 등을 이용하여 본인 여부를 확인하기 위한 생체인식시스템(biometrics recognition system) 중에 하나이다 (Yang 등, 2004). 기존의 도용, 분실, 복제의 여지가 많은 열쇠나 출입카드, ID, 비밀번호 등의 경우 범죄에 악용될 가능성이 높으며 휴대를 하거나 암호를 기억해야하는 불편함이 수반되나, 생체인식은 인간이 가진 생체정보를 이용하므로 탁월한 보안성과 휴대의 불필요성, 경제성 등 여러 면에서 기존의 방법에 비해 매우 유리하다. 생체인식에는 지문, 홍채 및 망막, 얼굴, 정맥, 서명, 음성, 장문, 손바닥, DNA 등이 사용되고 있다. 현재 생체인식은 개인용 컴퓨터 접근, 단말 액세스, 디지털 문서 보관, 전자 상거래, 인터넷 뱅킹 등 전자적 접근 제어뿐만 아니라 ATM 기계, 공항, 국경지대, 항만, 기타 보안을 요구하는 지역의 출입접근 제어 등 많은 응용분야에서 적용대상을 넓혀가고 있는 추세이다 (Park과 Lee, 2005). Tsutomu 등 (2002)과 Cappelli 등 (2006)은 여러 종류의 생체인식 기술 중에서 지문은 땀샘이 용기되어 일정한 흐름을 형성한 것으로 그 형태가 사람마다 다르고 태어날 때의 모습 그대로 평생 동안 변하지 않는 고유한 특성을 지니고 있기 때문에 지문을 이용한 생체인식은 현재까지 실용화의 면에서 가장 빠르고 활용성이 강하다고 하였다.

¹Corresponding author: Professor, Department of Statistics, Sungkyunkwan University, 25-2, Sungkyunkwan-ro, Jongro-gu, Seoul 110-745, Korea. E-mail: cshong@skku.edu

지문인식을 위한 알고리즘 중 지문인식 정합(matching)방법에는 특징점(minutiae)을 기반으로 한 정합 방법과 필터 뱅크(filter bank)를 기반으로 한 정합방법이 있다. 특징점을 기반으로 한 정합방법의 경우 지문 영상에서 지역적 특징점을 추출하여 특징점들의 상대적인 위치 관계를 이용하여 정합하는 방법이다 (Maio와 Maltoni, 1997). 보로노이 다이어그램(voronoi diagram)을 이용한 정합 (Kim과 Choi, 2004)과 그리드(grid)를 기반으로 한 정합 (Yang 등, 2004), AFV(adjacent feature vector)를 이용한 정합 (Tong 등, 2005) 등이 포함된다. 필터 뱅크에 기반 한 정합방법은 게이버 필터 뱅크(gabor filters bank)를 사용하여 지문의 융선(ridge)을 추출하는 방법이다 (Jain 등, 2000). 이 방법은 지역적(local) 특징뿐만 아니라 전역적(global) 특징 정보를 획득하여 사용할 수 있는 정합방법으로 Na와 Kim (2004)의 직교 좌표를 이용한 정합과 Jain 등 (2001)의 Hybrid 정합 방법이 있다. 이 외에도 지문의 특징점들을 지역적으로 정합하여 군집화하고 이를 이용하여 전체 지문 표면에 대한 변형 패턴을 모형화 하는 방법, 융선으로 서로 연결되어진 구조 정보를 활용하여 동일 융선상에 존재하는 특징점들에 대한 연결정보를 추출하고 이를 등록지문과 입력지문 간의 좌표 변환의 기준이 되는 특징점 쌍을 검출해서 정합시키는 방법 등 여러가지 지문인식 정합방법에 관한 연구는 꾸준히 이루어지고 있다.

본 연구에서는 특징점을 기반으로 하는 새로운 정합방법을 통계학적인 방법을 사용하여 제안하고자 한다. 지문인식 정합방법에서 통계학적 유사성(similarity) 측도를 사용한다. 신호탐지이론, 마케팅 판매, 의사결정론에 관한 데이터 마이닝 그리고 의학진단체계 및 신용평가 등에서 폭넓게 사용되는 혼합 분포의 판별력을 극대화하기 위하여 다양한 분류기준을 바탕으로 최적분류점(optimal threshold, 또는 절단점; cut-off value)을 발견하는 방법을 다양한 유사성 거리의 신뢰구간과 누적분포함수, ROC 곡선을 이용하여 제안한다.

본 논문의 구성은 다음과 같다. 2절에서는 지문인식모형과 정합방법을 소개하고 설명한다. 등록지문과 입력지문으로 구성된 새로운 지문인식모형을 설정하고 동일인과 비동일인을 분류하기 위한 유사성 거리 통계량을 제안한다. 3절에서는 지문인식모형을 따르는 다양한 자료를 생성하여 유사성 거리 통계량의 분포함수를 동일인과 비동일인에 따라 탐색하고 정의한다. 그리고 유사성 거리의 신뢰구간을 구하여 동일인과 비동일인을 판별하는 정합방법을 제안한다. 4절에서는 동일인과 비동일인의 유사성 거리 분포함수를 바탕으로 ROC 곡선을 구현하고 지문의 동일인과 비동일인을 분류하는 최적분류점을 구하여 이를 이용한 정합방법을 제안한다. 마지막 5절에서는 신뢰구간 그리고 최적분류점을 이용한 정합방법을 비교하면서 결론을 유도하고 향후 연구과제를 제시한다.

2. 지문인식모형과 유사성 거리

2.1. 지문인식모형

지문인식 시스템은 지문영상 전처리 단계(pre-process), 특징점 추출(minutiae extract) 단계, 정합 단계로 구성된다. 우선 지문영상 전처리 단계에서는 입력지문(input fingerprint) 영상의 잡음을 제거함으로써 지문의 융선과 골을 구별하고 등록지문(registered fingerprint)과의 차이를 최소화한다. 다음 단계인 특징점 추출 단계에서는 정합 단계에서 고려할 올바른 특징점을 추출하는 중요한 단계이다. 마지막으로 정합 단계에서는 두 지문 이미지의 특징점 분포를 비교하여 두 지문 사이의 유사성을 측정하는 과정이다. 유사성은 각 지문 특징점들의 일치 여부에 따라 점수로 계산되며, 이 점수가 미리 선정된 분류점을 초과할 경우 동일인으로 판정한다. 지문의 특징점을 정보로 하는 정합에서는 각 특징점들의 일치 정도를 점수로 계산하게 된다. 대표적인 정합률은 Park과 Lee (2005)이 제안하였다. 지문인식모형은 지문인식 시스템에서 주어진 두 지문 영상으로부터 동일인 또는 비동일인의 여부를 판단하는 역할을 수행하는 것으로서, 지문영상 전처리 과정과 특징점 추출과정 그리고 데이터베이스에 저장되어있는 정보

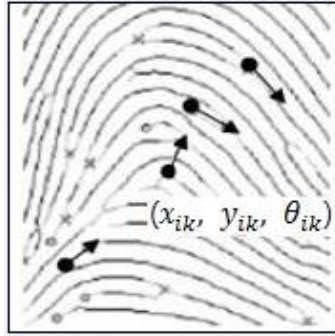


Figure 2.1. Fingerprint with minutiae

와 비교하여 유사한 지문인지 아닌지를 판단하는 정합과정을 거쳐 동일인과 비동일인을 판별하는 생체 인식모형 중의 하나이다.

지문인식 정합방법을 통계학적으로 접근하여 탐색하기 위하여 모의실험을 실행할 경우 지문의 골, 융선, 끝점, 분기점 등을 나타내는 특징점의 위치를 나타내는 좌표와 방향을 나타내는 각도가 필요하다. 본 논문에서는 등록지문과 입력지문에서 추출한 특징점의 좌표와 각도로 구성된 지문인식모형을 다음과 같이 설정한다.

등록지문 모형은 특징점의 좌표와 각도를 나타내는 $(x_{ik}, y_{ik}, \theta_{ik})$ 로 구성된다. 여기서 (x_{ik}, y_{ik}) 는 i 번째 등록한 사람의 k 번째 특징점의 지문 좌표이며($i = 1, 2, \dots, N$; N 은 등록지문 수, $k = 1, 2, \dots, K$; K 는 특징점 수), θ_{ik} 는 i 번째 사람의 k 번째 특징점의 각도를 나타낸다. Figure 2.1은 $K = 4$ 인 경우에 등록지문의 모형을 간단히 나타낸 예제이다. 지문은 단위 길이의 정사각형 내에 위치한다고 가정하여 등록지문의 특징점 $(x_{ik}, y_{ik}, \theta_{ik})$ 에서 좌표 (x_{ik}, y_{ik}) 는 평균과 분산이 동일한 크기의 균일분포와 정규분포인 $U(0, 1)$ 또는 $N(1/2, 1/12)$ 를 따르고, 특징점의 각도 θ_{ik} 는 균일분포 $U(0, 2\pi)$ 를 따른다고 식 (2.1)과 같이 가정한다.

$$\begin{aligned} x_{ik}, y_{ik} &\overset{iid}{\sim} \begin{cases} U(0, 1) \\ N\left(\frac{1}{2}, \frac{1}{12}\right) \end{cases}, \\ \theta_{ik} &\overset{iid}{\sim} U(0, 2\pi). \end{aligned} \quad (2.1)$$

그리고 입력지문 모형은 $(x'_{jk}, y'_{jk}, \theta'_{jk})$ ($j = 1, 2, \dots, M$; M 은 입력지문의 수)로 설정하는데, 등록지문의 특징점의 각 좌표와 각도에 오차항이 추가된 식 (2.2)와 같은 관계를 가진다고 가정한다.

$$x'_{jk} = x_{ik} + \epsilon_{jk}^1, \quad y'_{jk} = y_{ik} + \epsilon_{jk}^2, \quad \theta'_{jk} = \theta_{ik} + \epsilon_{jk}^3, \quad (2.2)$$

여기서 $\epsilon_{jk}^1, \epsilon_{jk}^2, \epsilon_{jk}^3$ 은 각각 독립적이고 동일한 정규분포 $N(0, 1/12)$ 를 따르는 확률오차이며, 확률오차의 분산도 등록지문의 좌표에 해당하는 x_{ik}, y_{ik} 의 분산과 동일하게 설정한다.

지문인식모형을 따르는 자료를 생성하기 위하여 특징점의 수 K 를 3, 5, 8로 정하고, 등록지문과 입력지문의 크기 N 과 M 을 30, 50, 100, 200으로 설정한다. 예를 들어 N 과 M 을 각각 30과 50으로 설정한다면, 정합하는 횟수 $30 \times 50 = 1,500$ 개 중에서 동일인의 지문은 30개이며 나머지 1,470개는 비동일인의 지문이다.

Table 3.1. Means of standardized similarity distances of the same and different person

same different	$K = 3$		$K = 5$		$K = 8$		
	N	Normal	Uniform	Normal	Uniform	Normal	Uniform
30		3.5133	3.5035	3.6747	3.6725	3.7694	3.7729
		4.7896	4.8613	5.2454	5.3151	5.4993	5.5791
50		3.7380	3.7353	3.9319	3.9684	4.0523	4.0799
		4.9666	5.0510	5.5207	5.6181	5.8963	5.9775
100		3.9924	3.9891	4.2626	4.2608	4.4458	4.4594
		5.1516	5.2599	5.8498	5.9522	6.3206	6.4260
200		4.2398	4.2346	4.5541	4.5687	4.7700	4.7884
		5.2882	5.4038	6.1112	6.2285	6.6958	6.7953

2.2. 유사성 거리

개체간의 유사성의 정도를 정량적으로 나타내기 위해 척도 중 많이 사용되는 다양한 거리 중에서 대표적인 유클리디안 거리(euclidean distance)와 맨해튼 거리(manhattan distance)를 이용한다. 특징점 사이의 거리(D_{ij}^1)와 특징점 사이의 각의 차이(D_{ij}^2)는 각각 유클리디안 거리와 맨해튼 거리를 이용하여 특징점의 거리를 다음과 같이 정의한다. 임의의 $i = 1, \dots, N, j = 1, \dots, M$ 에 대하여

$$D_{ij}^1 = \sum_{k=1}^K \sqrt{(x'_{jk} - x_{ik})^2 + (y'_{jk} - y_{ik})^2}, \quad D_{ij}^2 = \sum_{k=1}^K |\theta'_{jk} - \theta_{ik}|.$$

유사성 측정에 사용되는 거리의 척도들은 일반적으로 자료의 척도에 민감하기 때문에 표준화 유사성 거리(standardized similarity distance)를 임의의 $i = 1, \dots, N, j = 1, \dots, M$ 에 대하여 식 (2.3)과 같이 제안한다.

$$D_{ij} = \frac{D_{ij}^1 - \min(D_{ij}^1)}{\sqrt{\text{var}(D_{ij}^1)}} + \frac{D_{ij}^2 - \min(D_{ij}^2)}{\sqrt{\text{var}(D_{ij}^2)}}. \quad (2.3)$$

표준화 유사성 거리에서 평균 대신 최소값의 거리를 사용함으로써 비음수로 나타나고, 표준편차를 나눔으로 식 (2.3)에서 $i = j$ 인 경우의 D_{ii}^1, D_{ii}^2 는 동일인, 그리고 $i \neq j$ 인 경우의 D_{ij}^1, D_{ij}^2 는 비동일인 유사성 거리 D_{ii}, D_{ij} 의 분산은 2로 수렴된다.

3. 신뢰구간을 이용한 정합방법

지문인식모형으로 설정한 식 (2.1)와 식 (2.2)를 바탕으로 등록과 입력지문의 크기 N 과 M 을 동일한 크기로 설정하여 표본크기 N 과 특징점의 수 K 에 따라 식 (2.3)에서 제안한 동일인과 비동일인의 표준화 유사성 거리가 정규분포와 균일분포 별로 어떻게 나타나는지 살펴본다. Figure 3.1과 Figure 3.2 그리고 Table 3.1은 등록과 입력지문의 크기 N 과 M 을 30, 50, 100, 200으로 하고 특징점 수 K 를 3, 5, 8로 고정하여 모의실험을 각각 1,000번씩 반복수행하여 얻은 표준화 유사성 거리의 평균이다. Table 3.1에서 입력지문이 등록지문과 동일한지의 여부를 살펴보기 위하여 특징점의 수에 따른 분포별 동일인(same person)과 비동일인(different person)의 유사성 거리 D_{ii}, D_{ij} 의 평균에 대한 특징을 탐색한다.

Table 3.1을 바탕으로 표본크기가 증가함에 따라 표준화 유사성 거리를 살펴보기 위하여 Figure 3.1은 분포 종류별로 구현하였고 Figure 3.2는 특징점의 수 별로 동일인과 비동일인의 거리를 구현하였다.

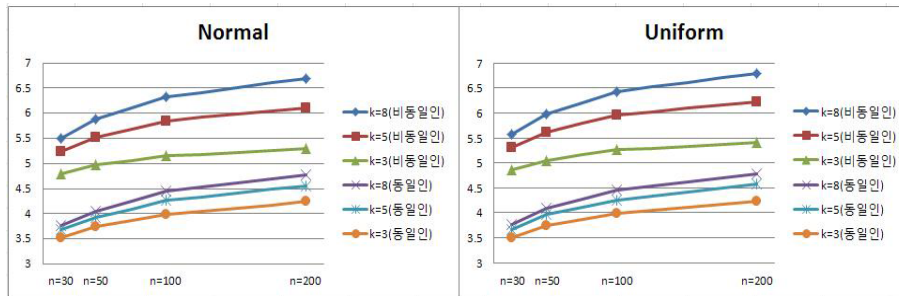


Figure 3.1. Standardized similarity distance w.r.t. minutiae

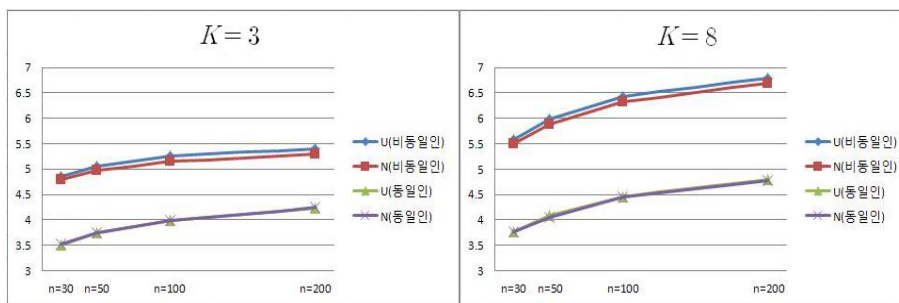


Figure 3.2. Standardized similarity distance w.r.t. distribution

Table 3.1과 Figure 3.1을 통하여 정규분포와 균일분포의 경우에 비동일인의 표준화 유사성 거리의 평균은 동일인에 비해 크다는 것을 확인할 수 있다. 이것은 표준화 유사성 거리가 적으면 등록과 입력지문 간의 유사성이 높으므로 등록과 입력지문이 동일하다고 판단됨을 의미한다. 등록과 입력지문의 표본 크기와 특징점의 수가 증가할수록 동일인과 비동일인의 표준화 유사성 거리는 증가한다. 정규분포에서 표준화 유사성 거리의 증가 속도는 균일분포에서의 증가속도보다 빠르다는 것을 파악할 수 있다. 그리고 Figure 3.2로부터 특징점의 수 $K = 3$ 일 때 분포종류에 따라 동일인의 유사성 거리는 유사하나 정규분포인 경우의 비동일인의 유사성 거리는 균일분포인 경우보다 작다. 이런 현상은 $K = 8$ 일 때도 유사하다. 등록과 입력지문의 크기가 증가할수록 유사성 거리의 증가속도는 현저하게 증가하는 것을 탐색하였다.

특징점의 수 K 가 3과 8인 경우에 등록과 입력지문의 크기 N 과 분포종류에 따라 동일인과 비동일인의 표준화 유사성 거리의 평균(Table 3.1)과 분산을 구하고 이를 바탕으로 동일인과 비동일인의 표준화 유사성 거리의 95% 근사신뢰구간을 구하여 Table 3.2에 나타내었다. 예를들어 등록과 입력지문의 수는 100이며 특징점의 수가 3, 그리고 특징점의 좌표가 균일분포를 따르는 경우에 동일인과 비동일인의 신뢰구간은 각각 (3.7119, 4.2663), (4.9827, 5.5371)이고, 등록과 입력지문의 수는 200, 특징점의 수가 8 그리고 정규분포인 경우의 동일인과 비동일인의 신뢰구간은 각각 (4.5740, 4.9660), (6.4998, 6.8918)임을 알 수 있다. Table 3.2를 살펴보면, 각각의 경우에 동일인과 비동일인의 표준화 유사성 거리의 95% 근사신뢰구간이 중복(overlap)되지 않는다는 것을 발견한다.

그러므로 동일인과 비동일인의 표준화 유사성 거리의 95% 근사신뢰구간을 이용하여 동일인 여부를 판단하는 정합방법으로 사용할 수 있다. 예를들어 $N = M = 30$ 이고 $K = 3$ 인 경우에 표준화 유사성 거리가 3 이하이면 동일인으로, 표준화 유사성 거리가 4.3 이상이면 비동일인으로 간주한다. 또한

Table 3.2. Confidence intervals of the same and different person

C.I. <i>N</i>	<i>K</i> = 3		<i>K</i> = 8	
	Normal	Uniform	Normal	Uniform
30	(3.0072, 4.0193)	(2.9974, 4.0095)	(3.2633, 4.2754)	(3.2668, 4.2789)
	(4.2836, 5.2957)	(4.3553, 5.3674)	(4.9932, 6.0054)	(5.0730, 6.0852)
50	(3.3460, 4.1300)	(3.3433, 4.1273)	(3.6603, 4.4443)	(3.6879, 4.4719)
	(4.5746, 5.3586)	(4.6590, 5.4430)	(5.5043, 6.2883)	(5.5855, 6.3695)
100	(3.7152, 4.2696)	(3.7119, 4.2663)	(4.1686, 4.7230)	(4.1822, 4.7366)
	(4.8744, 5.4288)	(4.9827, 5.5371)	(6.0434, 6.5978)	(6.1488, 6.7032)
200	(4.0438, 4.4358)	(4.0386, 4.4306)	(4.5740, 4.9660)	(4.5924, 4.9844)
	(5.0922, 5.4842)	(5.2078, 5.5998)	(6.4998, 6.8918)	(6.5993, 6.9913)

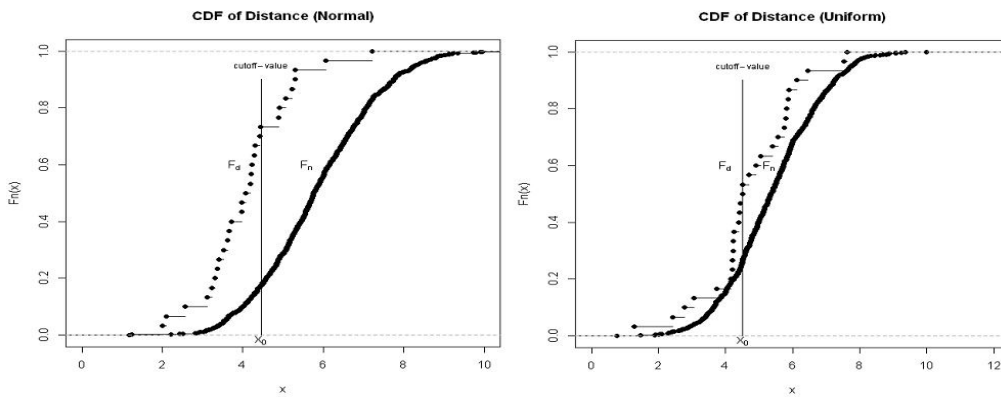


Figure 4.1. CDFs of the same and different persons

$N = M = 100$ 이고 $K = 8$ 인 경우에 표준화 유사성 거리가 5 미만이면 동일인으로, 표준화 유사성 거리가 6 이상이면 비동일인으로 판단한다.

4. 최적분류점을 이용한 정합방법

2절에서 제안한 지문인식모형에서의 표준화 유사성 거리 D_{ii}, D_{ij} 를 바탕으로 동일인과 비동일인 집단 간의 분류점(또는 절단점)을 설정하여 표준화 유사성 거리가 분류점보다 작으면 동일인으로, 분류점보다 크면 비동일인으로 간주하여 판별할 수 있기 때문에 분류점을 발견하는 것이 중요한 과제이다. 이것은 신용평가모형(credit evaluation model)에서 부도와 정상(default vs. non-default)의 예측과 생물통계(biostatistics)에서 정상과 환자(good vs. bad)의 판단 등에서의 동일인 문제이다. 따라서 표준화 유사성 거리 D_{ii}, D_{ij} 의 누적분포함수를, 예를들어 신용평가모형에서의 부도와 정상에 대한 누적분포함수 $F_d(\cdot), F_n(\cdot)$ 와 같이, 각각의 누적분포함수를 식 (4.1)과 같이 정의한다. 임의의 $x \geq 0$ 에 대하여,

$$F_d(x) = \frac{1}{N} \sum_{i=1}^N I(D_{ii} \leq x), \quad F_n(x) = \frac{1}{N(M-1)} \sum_{\substack{j=1 \\ i \neq j}}^M \sum_{i=1}^N I(D_{ij} \leq x). \quad (4.1)$$

등록과 입력지문의 수 $N = M = 30$ 그리고 특징점의 수 $K = 8$ 인 경우의 누적분포함수를 Figure 4.1에 표현하였다. Figure 4.1의 왼쪽 그림은 정규분포, 오른쪽은 균일분포의 경우이다. Figure 4.1 각각의

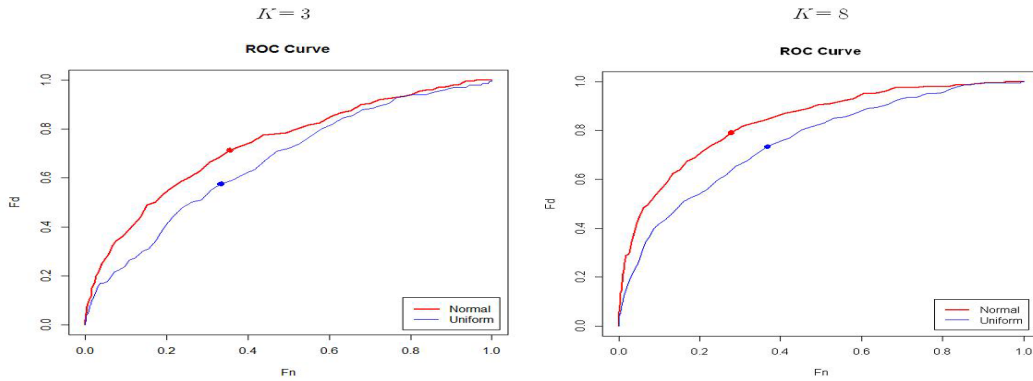


Figure 4.2. ROC curves

그림에서 왼쪽의 누적분포함수는 동일인의 누적분포함수, 오른쪽은 비동일인의 분포함수를 나타내며 임의의 x 에서 $F_d(x) \geq F_n(x)$ 를 만족한다.

표준화 유사성 거리의 누적분포함수 $F(x)$ 를 동일인의 분포함수 $F_d(\cdot)$ 와 비동일인의 함수 $F_n(\cdot)$ 의 혼합으로 식 (4.2)와 같이 정의할 수 있다.

$$F(x) = \lambda F_d(x) + (1 - \lambda) F_n(x), \tag{4.2}$$

여기서 $\lambda = N/NM$ 이며, λ 는 신용평가모형에서 총부도율(total probability of default)이며 지문인식 모형에서는 총정합확률(total probability of matching)이다.

지문 정합에서 시스템 성능을 평가하는 대표적인 방법인 FAR(false accept rate: 타인수락 오류율 또는 동일인 오분류율)과 FRR(false reject rate: 본인거절 오류율 또는 비동일인 오분류율)은 혼합 분포로부터 추출한 확률표본의 판별력을 나타내는 최적분류점을 추정하는 방법이다. 여기서 타인수락 오류율은 등록된 사용자의 지문과는 다른 사용자의 지문이 입력되었음에도 불구하고 동일인 사용자의 지문으로 잘못 판정하는 경우이고, 본인거절 오류율은 등록된 사용자의 지문과 같은 사용자의 지문이 입력되었음에도 타인이라고 잘못 판정하는 경우이다. 본 연구에서는 식 (4.2)와 같은 혼합분포함수로부터 분류점을 발견하기 위하여 Park과 Lee (2005)의 정합율보다 정교한 오류율 측도로서 타인수락 오류율 또는 동일인 오분류율(FAR)과 본인거절 오류율 또는 비동일인 오분류율(FRR)을 사용한다. 이것은 각각 제 I 종과 II 종 오류(type I and II error)로 나타내며 분류점 x_0 에서의 각각의 오분류율은 식 (4.3)과 같이 정의한다.

$$FRR = \frac{\text{동일인을 비동일인으로 오인식한 회수}}{\text{동일인의 매칭회수}} = \frac{1}{N} \sum_{i=1}^N I(D_{ii} > x_0) \equiv 1 - F_d(x_0),$$

$$FAR = \frac{\text{비동일인을 동일인으로 오인식한 회수}}{\text{비동일인의 매칭회수}} = \frac{1}{N(M-1)} \sum_{\substack{j=1 \\ i \neq j}}^M \sum_{i=1}^N I(D_{ij} \leq x_0) \equiv F_n(x_0). \tag{4.3}$$

표준화 유사성 거리 D_{ii}, D_{ij} 의 누적분포함수인 $F_d(\cdot), F_n(\cdot)$ 을 단위 크기의 사각형의 Y축과 X축으로 각각 표현한 ROC 곡선을 작성하여 Figure 4.2에 구현하였다. FRR과 FAR을 최소화 하는 분류점을 설정하는 최적분류 기준에는 Youden 지수 (Youden, 1950), SSS(sum of sensitivity and specificity; Connell과 Koepsell, 1985), DPR(differential positive rate; Ward, 1986), 대칭점(Symmetry

Table 4.1. Optimal thresholds

optimal threshold	$K = 3$		$K = 8$	
	Normal	Uniform	Normal	Uniform
N				
30	3.9030	3.9785	4.4575	4.5120
50	4.1451	4.1763	4.8358	4.8819
100	4.3570	4.4281	5.2665	5.3335
200	4.5701	4.6304	5.6249	5.6705

point; Moses 등, 1993; Pepe, 2003), (0, 1)까지 최단거리(The closest-to-(0, 1) criterion; Perkins와 Schisterman, 2006), 수정된 (0, 1)까지 최단거리(The amended closest-to-(0, 1) criterion; Perkins와 Schisterman, 2006), 전체정확도 (Total Accuracy; Lambert와 Lipkovich, 2008), 진실율(True Rate; Lambert와 Lipkovich, 2008; Hong과 Joo, 2010), 균형정확도 MVD(maximum vertical distance; Krzanowski와 Hand, 2009) 등이 있는데 그 중에서 많이 사용하는 Youden 지수, SSS, DPR, 수정된 (0, 1)까지 최단거리, 진실율, 균형정확도 MVD들은 Kolmogorov and Smirnov 통계량과 동일한 개념의 통계량이다 (Hong 등, 2012). 따라서 본 연구에서는 Kolmogorov and Smirnov 통계량($= \max \{F_d(x) - F_n(x)\}$)에 대응하는 최적분류기준을 사용하여 최적분류점을 발견하고 이 점을 Figure 4.1의 수평축에 x_0 로 나타내고 이에 대응하는 참조선(reference line)을 수직선으로 표현하였다. 이때 각각의 분포에 대하여 정규분포인 경우 최적분류점 $x_0 = 4.4575$ 이고 균일분포에 대한 최적분류점 $x_0 = 4.5120$ 이다. 그리고 최적분류점 x_0 에 대응하는 $(F_n(x_0), F_d(x_0))$ 의 좌표를 Figure 4.2의 ROC 곡선에 ‘●’으로 표시하였다. Figure 4.2를 살펴본 결과 특징점의 수 K 가 증가함에 따라 ROC 곡선은 (0, 1)점에 가까이 접근하고 이 현상은 특징점의 수와 표본크기가 증가할수록 판별력이 좋아지고 균일분포인 경우보다 정규분포인 경우에 판별력이 증가한다고 파악할 수 있다.

2절에서 제안한 지문인식모형에서 등록과 입력지문의 크기와 특징점의 수에 대하여 모의실험을 각각 1,000번 반복 시행하여 얻은 최적분류점을 분포별로 Table 4.1에 정리하였다. Table 4.1을 살펴보면 특징점의 수가 증가함에 따라 두 분포 모두 최적분류점이 증가하고 있음을 알 수 있으며, 정규분포에서의 최적분류점은 균일분포보다 작지만 산출된 최적분류점은 Table 3.2에 나타난 동일인과 비동일인의 신뢰구간에 포함되지 않고 두 신뢰구간의 사이에 위치하는 것을 발견하였다. 등록과 입력지문의 크기에 따른 분포별 최적분류점은 크게 차이점이 나타나지 않음을 알 수 있다. Table 4.1에서 $N = M = 30$, $K = 3$ 그리고 균일분포인 경우에 표준화 유사성 거리의 최적분류점은 3.9785이므로 표준화 유사성 거리가 3.9785 미만이면 동일인으로, 3.9785 이상이면 비동일인으로 판별하고, $N = M = 100$, $K = 8$ 그리고 정규분포인 경우에 표준화 유사성 거리의 최적분류점은 5.2665이므로 표준화 유사성 거리가 5.2665 미만이면 동일인으로, 5.2665 이상이면 비동일인으로 인식한다. 따라서 본 연구에서 제안한 지문인식모형에서 표준화 유사성 거리의 최적분류점을 지문인식 정합방법으로 사용할 수 있다.

5. 결론

본 연구에서는 지문의 특징점을 바탕으로 지문인식모형과 유사성 거리 측도를 제안하여 통계적인 지문인식 정합방법을 연구하였다. 특징점은 평면상의 좌표와 특징점의 방향을 나타내는 각도로 구성된다고 가정하고, 등록과 입력지문으로 구분하여 특징점의 분포를 정규분포와 균일분포를 따른다고 가정하면서 특징점의 개수와 표본수에 대한 모형을 설정하였다. 동일인과 비동일인의 유사성 거리의 분포와 특징을 탐색하여 등록과 입력지문의 수가 증가하고 특징점의 수가 증가할수록 표준화 유사성 거리의 분포를 탐색하였다.

통계적인 지문인식모형을 바탕으로 신뢰구간과 최적분류점을 이용하는 두 가지 정합방법을 제안하였다. 우선 동일인과 비동일인의 표준화 유사성 거리의 95% 근사신뢰구간을 구하고, 동일인과 비동일인의 신뢰구간이 중복되지 않는 결과로부터 동일인 여부를 판단하는 정합방법으로 표준화 유사성 거리의 신뢰구간을 이용하는 방법을 제안하였다. 두 번째 방법으로 동일인과 비동일인의 표준화 유사성 거리의 누적분포함수를 정의하고 두 분포함수로 표현하는 ROC 곡선으로부터 본인거절(또는 비동일인) 오류율과 타인수락(또는 동일인) 오류율을 최소화하는 분류기준에 적합한 표준화 유사성 거리의 최적분류점을 지문인식 정합방법으로 사용할 수 있음을 제안하였다.

표준화 유사성 거리의 95% 근사신뢰구간을 바탕으로 예를들어, $N = M = 30$ 이고 $K = 3$ 인 경우에 유사성 거리가 3 이하이면 동일인으로, 4.3 이상이면 비동일인으로 간주한다. 균일분포인 경우에 표준화 유사성 거리의 최적분류점은 3과 4.3 사이에 위치한 3.9785이므로 표준화 유사성 거리가 3.9785 미만이면 동일인으로, 3.9785 이상이면 비동일인으로 판별할 수 있다. 또한 $N = M = 100$ 이고 $K = 8$ 인 경우에 표준화 유사성 거리가 5 미만이면 동일인으로, 표준화 유사성 거리가 6 이상이면 비동일인으로 판단한다. 정규분포인 경우에 표준화 유사성 거리의 최적분류점은 5와 6 사이에 존재하는 5.2665이므로 표준화 유사성 거리가 5.2665 미만이면 동일인으로, 5.2665 이상이면 비동일인으로 인식할 수 있다.

표준화 유사성 거리의 최적분류점은 중복되지 않은 동일인과 비동일인의 표준화 유사성 거리의 신뢰구간 사이에 존재하는 것을 발견하였으며 본 연구에서 제안한 두가지 정합방법을 사용하여 동일인과 비동일인의 지문을 인식할 수 있으며 표준화 유사성 거리의 신뢰구간을 이용하는 방법보다 최적분류점을 이용하는 방법이 보다 정교하게 판별 및 분류할 수 있음을 발견하였다.

본 연구에서는 지문의 특징점을 이용하는 지문인식 정합방법을 위하여 통계적 확률분포를 따르는 모형을 설정하고 유사성 거리를 정의하여 동일인과 비동일인의 유사성 거리 함수를 탐색하여 지문인식 정합방법을 제안하였는데 특징점의 수가 적은 제한적인 경우만을 연구하였다. 향후에는 특징점의 수가 현실적으로 많은 경우를 고려하고, 특징점의 확률분포함수를 다양한 경우에 살펴보아야 하겠다. 본 연구에서 제안한 통계적 정합방법을 여러 종류의 지문인식 정합방법과 나아가 생체인식 정합방법에도 적용하여 응용 범위를 확대할 필요성이 있으며, 지문인식의 실용성을 위해서 실제 지문영상자료 또는 지문영상 데이터베이스를 이용하여 본 연구에서 제안된 지문정합방법을 구현하는 연구를 향후 과제로 남겨둔다.

References

- Cappelli, R., Maltoni, D., Wayman, J. L. and Jain, A. K. (2006). Performance evaluation of fingerprint verification system, *IEEE Trans. Pattern Analysis and Machine Intelligence*, **28**, 3–18.
- Connell, F. A. and Koepsell, T. D. (1985). Measures of gain in certainty from a diagnostic test, *American Journal of Epidemiology*, **121**, 744–753.
- Hong, C. S. and Joo, J. S. (2010). Optimal thresholds from non-normal mixture, *The Korean Journal of Applied Statistics*, **23**, 943–953.
- Hong, C. S., Kim, G. C. and Jeong, J. A. (2012). Bivariate ROC curve, *The Korean Journal of Applied Statistics*, **19**, 277–286.
- Jain, A. K., Prabhakar, S., Hong, L. and Pankanti, S. (2000). Filterbank-based fingerprint matching, *IEEE Trans. Image Processing*, **9**, 846–859.
- Jain, A. K., Ross, A. and Prabhakar, S. (2001). Fingerprint matching using minutiae and texture features, *Appeared in Proc. of Int'l Conference on Image Processing(ICIP)*, **3**, 282–285.
- Kim, S. H. and Choi, T. Y. (2004). A fingerprint matching algorithm based on the Voronoi diagram, *The Institute of Electronics Engineers*, **41**, 247–252.
- Krzanowski, W. J. and Hand, D. J. (2009). *ROC Curves for Continuous Data*, Chapman & Hall/CRC, Boca Raton, Florida. Has Been Selected, *Clinical Chemistry*, **32**, 1341–1346.

- Lambert, J. and Lipkovich, I. (2008). A macro for getting more out of your ROC curve, *SAS Global Forum*, **231**.
- Maio, D. and Maltoni, D. (1997). Direct gray-scale minutiae detection in fingerprints, *IEEE Transactions on Pattern Analysis Machine Intelligence*, **19**, 27–40.
- Moses, L. E., Shapiro, D. and Littenberg, B. (1993). Combining independent studies of a diagnostic test into a summary ROC curve: Data-analytic approaches and some additional considerations, *Statistics in Medicine*, **12**, 1293–1316.
- Na, H. J. and Kim, C. S. (2004). The study on the extraction of the minutiae and singular point for fingerprint matching, *Proceedings of the Korea Multimedia Society Conference*.
- Park, J. J. and Lee, K. H. (2005). Fingerprint recognition information of ridge shape of minutiae, *The Korea Institute of Signal Processing and Systems*, **2**, 67–73.
- Pepe, M. S. (2003). *The Statistical Evaluation of Medical Tests for Classification and Prediction*, University Press, Oxford.
- Perkins, N. J. and Schisterman, E. F. (2006). The inconsistency of “Optimal” cutpoints obtained using two criteria based on the receiver operating characteristic curve, *American Journal of Epidemiology*, **163**, 670–675.
- Tong, X., Huang, J., Tang, X. and Shi, D. (2005). Fingerprint minutiae matching using the adjacent feature vector, *Pattern Recognition Letters*, **26**, 1337–1345.
- Tsutomu, M., Hiroyuki, M., Koji, Y. and Satoshi, H. (2002). Impact of artificial gummy fingers on fingerprint system, *Optical Security and Counterfeit Deterrence Techniques IV*, **467**, 275–289.
- Ward, C. D. (1986). *The Differential Positive Rate, a Derivative of Receiver Operating Characteristic Curves Useful in Comparing Tests and Determining Decision Levels*, *Clinical Chemistry*, **32**, 1428–1429.
- Yang, S. R., Kim, S. M. and Cho, S. B. (2004). A grid-based matching algorithm for high performance fingerprint recognition, *Korean Institute of Information Scientists and Engineers*, **31**, 727–729.
- Youden, W. J. (1950). Index for rating diagnostic test, *Cancer*, **3**, 32–35.