

최적화된 확률 모델을 이용한 다양한 품질의 지문분류

정혜욱¹ · 이지형^{1†}

Various Quality Fingerprint Classification Using the Optimal Stochastic Models

Hye-Wuk Jung · Jee-Hyong Lee

ABSTRACT

Fingerprint classification is a step to increase the efficiency of an 1:N fingerprint recognition system and plays a role to reduce the matching time of fingerprint and to increase accuracy of recognition. It is difficult to classify fingerprints, because the ridge pattern of each fingerprint class has an overlapping characteristic with more than one class, fingerprint images may include a lot of noise and an input condition is an exceptional case. In this paper, we propose a novel approach to design a stochastic model and to accomplish fingerprint classification using a directional characteristic of fingerprints for an effective classification of various qualities. We compute the directional value by searching a fingerprint ridge pixel by pixel and extract a directional characteristic by merging a computed directional value by fixed pixels unit. The modified Markov model of each fingerprint class is generated using Markov model which is a stochastic information extraction and a recognition method by extracted directional characteristic. The weight list of classification model of each class is decided by analyzing the state transition matrixes of the generated Markov model of each class and the optimized value which improves the performance of fingerprint classification using GA (Genetic Algorithm) is estimated. The performance of the optimized classification model by GA is superior to the model before the optimization by the experiment result of applying the fingerprint database of various qualities to the optimized model by GA. And the proposed method effectively achieved fingerprint classification to exceptional input conditions because this approach is independent of the existence and nonexistence of singular points by the result of analyzing the fingerprint database which is used to the experiments.

Key words : Fingerprint Classification, Stochastic Model, Biometric System

요약

지문분류는 1:N 지문인식 시스템의 효율성을 높이는 단계로 지문의 매칭 시간 단축과 인식의 정확성을 높여주는 역할을 한다. 지문 각 클래스의 융선 패턴은 한 개 이상의 클래스와 중복되는 성질을 가지기 때문에 지문분류 작업은 어렵다. 또한 잡음을 많이 포함하거나 예외적인 입력 상태인 경우에도 분류 작업은 어려워진다. 본 논문에서는 다양한 품질의 지문을 효과적으로 분류하기 위해 지문의 방향특징을 이용해 확률 모델을 설계하고, 이를 최적화 하여 지문분류를 수행하는 방법을 제안하였다. 지문 융선을 픽셀단위로 탐색하여 방향 값을 산출하고, 산출된 방향 값을 일정 픽셀 단위로 병합하여 지문의 방향특징을 추출한다. 추출된 방향 특징을 이용해 확률론적 정보추출 및 인식 방식인 마코프 모델을 이용하여 지문의 클래스별 마코프 모델을 생성한다. 생성된 클래스별 마코프 모델의 상태전이 행렬을 분석하여 클래스별 분류 모델의 가중치 항목을 결정하고 유전자 알고리즘을 이용하여 지문분류 성능을 향상시킬 수 있는 최적의 수치를 찾아낸다. 유전알고리즘에 의해 최적화된 분류 모델에 다양한 품질의 지문 데이터베이스를 적용하여 실험해 본 결과 최적화 되기 전의 분류 모델에 비해 우수한 분류성능을 보였다. 또한 실험에 사용한 다양한 품질의 데이터베이스를 분석해본 결과 제안한 방법은 특이점 유, 무 및 상태에 독립적으로 예외적인 입력상황의 지문에 대해 효율적으로 지문분류를 수행했다.

주요어 : 지문분류, 확률모델, 생체인식 시스템

* 이 논문은 2009년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업의 연구 결과입니다.(No. 2009-0075109)

2009년 12월 6일 접수, 2010년 3월 11일 채택

¹⁾ 성균관대학교 컴퓨터공학과

주 저 자 : 정혜욱

교신저자 : 이지형

E-mail: jhlee@ece.skku.ac.kr

1. 서론

최근 개인인증 및 신원확인에 사용되는 지문인식 시스템의 응용범위가 다양해짐에 따라 지문을 정확하고 효율적으로 인식하는 것의 중요성이 증가하고 있다. 특히 대용량 지문데이터 베이스로 구축된 1:N 지문인식 시스템의 경우 지문의 매칭 시간 단축과 인식의 정확성을 높여 주기 위해 높은 신뢰성이 요구된다^[1]. 지문분류는 이러한 지문인식 시스템의 효율성을 높이기 위한 단계로 지문 용선 패턴에 따라 Arch(A), Whorl(W), Left(L), Right(R)로 분류한다^[2]. 지문 각 클래스의 용선 패턴은 한 개 이상의 클래스와 중복되는 성질을 가지기 때문에 지문분류 작업은 어렵다. 또한 잡음을 많이 포함하거나 예외적인 입력 상태인 경우에도 분류 작업은 어려워진다. 이러한 문제를 해결하기 위해 지문이미지에 다양한 영상개선 방법을 적용 후 특징을 추출하여 분류하는 방법으로 연구가 진행 되어왔다. 전통적인 지문분류 방법은 지문 용선의 흐름에 따라 중심점(core), 삼각주(delta)로 정의된 특이점을 기반으로 한다. 그림 1은 입력방식에 따른 지문이미지를 나타낸다. 그림 1(a)는 회전 날인된 지문이미지로 전통적인 지문분류 방식에 사용된 지문이미지이다. 그림 1(b)~(d)는 광학식 라이브스캔 방식으로 입력된 지문이미지이다. 그림 1(a)에서 보는바와 같이 회전 날인된 지문이미지의 경우 기존 헨리식 분류방법에 의한 특이점이 모두 존재 하지만, 그림 1(b)와 같이 광학식 라이브스캔 방식으로 입력된 지문 이미지의 경우 입력센서의 크기차이에 의해 특이점 중 하나인 삼각주가 존재하지 않는다. 또한 그림 1(c)는 예외적인 지문 입력 상황에 의해 지문의 일부분만 인식된 경우이고 그림 1(d)는 품질이 낮고 삼각주가 존재하지 않고 중심점의 위치가 한쪽으로 치우친 경우이다.

지문 분류는 크게 특이점 기반 방법과 방향성 기반 접근 방법으로 나눌 수 있다^[1]. 특이점 기반 방법은 대표적으로 Poincare와 같은 특이점 추출 알고리즘을 사용하여 추출된 특이점의 개수와 위치를 통해 지문을 분류 한다^[2]. Karu와 Jain^[2]은 특이점 개수 및 위치정보를 이용해서 휴

리스틱 규칙을 세우고 이를 통해 지문분류를 수행했고 평균 93.1%의 분류율을 보였다. 이러한 특이점 기반 방법은 잡음을 많이 포함하거나 예외적인 상황에 입력된 부분 이미지에 적합하지 않다. 또한 회전 날인된 지문에는 잘 적용되지만, 삼각주가 인식되지 못하는 라이브스캔 방식으로 입력된 지문에는 적합하지 않다. 방향성 기반 방법은 지문 이미지에서 방향성분을 추출하여 방향 영역을 분할하거나 지문 용선의 흐름 패턴을 정의하여 지문을 분류한다^[3]. Cappelli과 Lumini^[4]는 특이점 추출 후 특이점 주위의 방향성분과 패턴을 이용해 지문분류를 수행했고 평균 90.7%의 분류율을 보였다. 이러한 방향성 기반 방법은 특이점 영역과 유사한 중심점 및 삼각주 영역을 추출 한 후 이 영역을 기준으로 클래스별 지문의 방향 구조를 분할하거나 방향 패턴을 정의한다. 따라서 지문분류에 직접적으로 특이점 정보를 사용하지는 않았지만, 특이점 정보를 기준으로 지문의 방향 특징을 추출했기 때문에 라이브스캔 방식으로 입력되어 삼각주가 존재하지 않거나 예외적인 상황으로 인해 중심점의 위치가 한쪽으로 치우쳐 입력된 지문에는 적합하지 않다.

본 논문에서는 특이점정보 및 영역을 이용하지 않고 지문의 방향성분만을 이용하여 지문 분류를 수행한다. 지문의 방향성분을 추출하기 위해 지문 용선을 픽셀단위로 탐색하여 방향 값을 산출한다. 산출된 방향 값을 일정 픽셀 단위로 병합하여 확률론적 정보추출 및 인식 방식인 마코프 모델을 이용하여 지문의 클래스별 마코프 모델을 생성한다. 또한 생성된 클래스별 마코프 모델을 최적화하여 변형하기 위해 32개의 기중치를 정의하고 유전 알고리즘을 이용하여 최적의 값을 찾아 적용한다. 이러한 방법은 예외적인 상황으로 인해 손상되어 잡음이 많은 지문, 특이점 정보가 존재 하지 않거나 지문의 일부분만 입력된 경우에도 확률적 추론에 의해 분류가 가능하다.

본 논문의 구성은 다음과 같다. 2장에서는 지문분류를 위한 지문 이미지의 특징추출 방법에 대해 기술하고, 3장에서는 지문 분류를 위한 확률모델에 대해 기술한다. 4장에서는 실험 결과를 기술하고, 5장에서는 결론 및 향후 연구 과제에 대해 제시한다.



그림 1. 다양한 품질의 지문 데이터

2. 특징추출

본 장에서는 지문분류에 사용되는 특징을 추출하기 위해 전처리 과정과 지문 용선으로 부터 방향 값을 추출 하는 방법을 설명한다.

2.1 전처리

잉크 및 다양한 센서에 의해 입력된 지문 이미지는 일반적으로 그레이 레벨의 해상도를 가진다. 본 논문에서 사용하는 FVC2000, FVC2002 데이터베이스는 광학식 라이브 스캔 센서로 입력된 그레이 레벨의 이미지이다.

본 논문에서 제안하는 지문분류 모델은 용선의 방향 값을 사용한다. 따라서 픽셀단위로 용선을 추적하여 방향 변화 특징 값을 산출한다. 이러한 특징을 추출 작업을 용이하게 하기 위해 다음과 같은 전처리 과정을 거친다.

1. 입력된 지문의 명암분포를 고르게 해주기 위해 평활화(smoothing) 방법을 적용 한다.
2. 용선의 방향 값 산출을 위해서는 지문이미지의 용선과 골을 구분해 하기 때문에 이진화(binanzation) 방법을 적용하여 용선과 골이 검정색과 흰색으로 표현되는 이미지로 변환한다.
3. 지문 용선을 픽셀 단위로 탐색하여 방향 값을 산출하기 위해 1픽셀(pixel) 두께로 변환하는 세선화(thinning) 처리를 한다.

2.2 회전 지문 보정

지문은 찍히는 상황에 따라 기울기의 차이가 난다. 지문이 회전되어 입력된 지문이미지의 경우 방향 특징이 변질될 수 있다. 이는 지문 분류율을 낮추는 요인이 된다. 따라서 회전 지문이미지의 위치보정에 지문 이미지의 길이를 비례로 타원 템플릿을 생성하는 Ellipse fitting 방법과 각도 조절 방법을 이용한다. 제안하는 방법은 10° 이상 회전되어 입력된 지문이미지의 경우 지문 분류율을 낮춘다고 가정하고 이 지문들에 대해 위치 보정방법을 적용한다. 그림 2는 전처리 과정을 거친 회전된 지문 이미지의 보정 과정이다. 그림 2(a)는 10° 이상 회전되어 입력된 지문이미지 원본이다. 그림 2(b)는 전처리 과정을 거친 이미지이고, 그림 2(c)와 같이 Ellipse fitting 방법을 이용하여 지문이미지에 맞는 타원을 찾고, fitting된 타원의 회전 각도인 θ 를 계산한 후 지문이미지의 위치를 회전 시켜 보

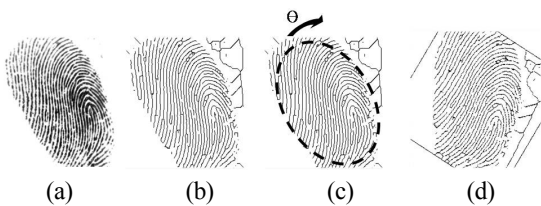


그림 2. 회전된 지문이미지의 보정과정

정한다. 그림 2(d)는 보정 후 지문이미지의 위치가 올바르게 바뀐 결과이다.

2.3 방향 특징 추출

전처리 과정 및 회전 보정 과정을 거친 지문이미지에서 지문분류에 사용하는 특징을 추출하기 위해서는 픽셀 단위로 용선을 탐색하여 방향 값을 산출해야 한다. 그림 3은 용선 탐색에 사용한 방향 정보 나타낸다. 그림 3(a)는 0°~360°의 방향을 용선 흐름 변화를 가장 적합하게 탐색 할 수 있는 22.5° 간격으로 나누어 16개의 방향 값을 나타낸다. 그림 3(b)는 5×5 윈도우상에 위치한 16개의 방향 값과 좌표 값을 나타낸다.

이와 같이 정의된 방향 값을 기준으로 5×5 윈도우를 이용하여 용선의 픽셀을 추적하여 방향 값을 산출 한다. 그림 4(a)는 용선 탐색에 사용되는 5×5 탐색윈도우이다. 이때 P는 용선 탐색의 중심이 되는 픽셀로 이 중심 픽셀을 기준으로 픽셀 탐색 우선순위를 결정한 후 16개의 방향 값을 그림 (b), (c)와 같이 산출 한다. 용선 탐색 중 분기점이 나오면 분기점을 기준으로 상단에 위치한 용선을 먼저 탐색하여 방향 값을 산출하고, 하단 부분에 존재하는 용선에 대해 동일한 과정을 반복한다.

그리고 용선의 길이가 20픽셀 미만인 경우 잡음으로 간주하여 탐색 용선에서 제외시킨다.

이러한 지문에서 용선의 방향 값 산출은 그림 5와 같이 256×256 크기의 지문 이미지 좌측에 수직선을 세워 수직선과 용선과의 교점을 시작점으로 정한 후 용선 탐색을 수행한다. 그림 5(b), (c)는 용선 탐색을 위한 수직선이 좌측에서 우측으로 이동하며 용선과 수직선의 교점을 찾

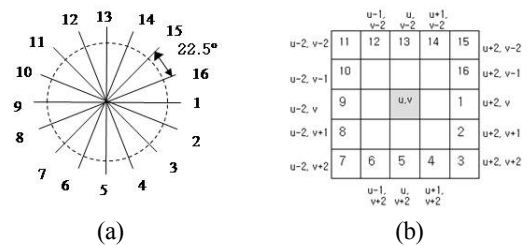


그림 3. 용선 탐색에 사용되는 방향 값

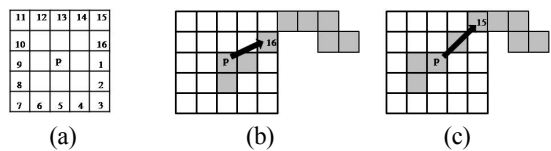


그림 4. 용선 탐색 윈도우

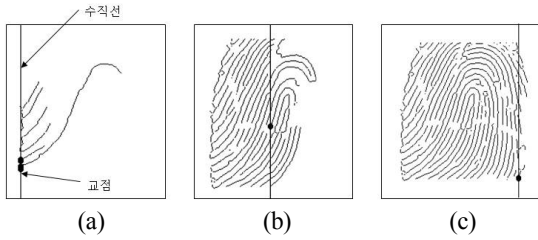


그림 5. 융선 탐색 과정

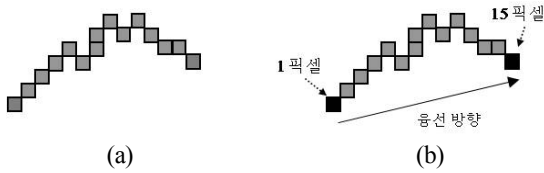


그림 6. 융선의 방향 값 병합

는 예를 보여준다. 이때 교점을 찾는 수직축의 스캔범위는 불필요한 탐색범위를 줄이기 위해 좌, 우, 상, 하 영역에서 각각 30, 30, 45, 45 픽셀은 고려하지 않는다.

위와 같은 과정을 거쳐 추출된 융선의 방향 값은 픽셀 단위의 정보이므로 픽셀간의 방향 변화 차이가 미세하기 때문에 지문 융선의 흐름 변화를 표현하기 어렵다. 따라서 융선의 방향변화 특징을 산출하기 위해 1픽셀 단위 방향 값을 15픽셀 단위로 병합한다. 그림 6(a)와 같은 융선의 픽셀이 있는 경우 그림 6(b)와 같이 1개의 융선에 대해 첫 번째 픽셀을 기준으로 15번째 픽셀이 변화하는 방향 값으로 산출했다. 이와 같은 융선의 방향 값 병합은 아주 작은 변화량을 보이는 1픽셀단위의 방향 값을 15픽셀 단위로 병합함으로써 융선의 방향 변화량 측정을 용이하게 할 수 있는 방법이다^[5].

3. 지문 분류 확률모델

본장에서는 지문 분류를 위한 확률 모델을 생성하고 이를 변형하여 최적의 지문분류 모델을 생성하는 방법을 설명한다.

3.1 마코프 모델(Markov models, MM)

마코프 모델은 확률 분포와 그것의 연산을 바탕으로 개발된 확률 모델이다. 마코프 모델의 성질은 관찰되는 상태가 시간에 따라 매 순간 변화하며 상태간 전이가 오로지 이전 n 개의 상태에만 관련이 있고 나머지 과거의 상태에는 독립적이다. 마코프 모델은 관찰 가능한 상태들의

집합인 관측열을 찾기 위해 마코프 체인을 사용하고 매개 변수는 식 (1)과 같다^[6]. 여기서 M 은 클래스 i 의 모델이라 했을 때 마코프 모델의 집합, S 는 마코프 모델 내부의 상태 i 에서 j 로 전이 확률, π 는 각 상태 i 가 초기에 나타날 확률이다. 식 (2)의 S 는 마코프모델의 상태전이 행렬, 식 (3)의 π 는 마코프 모델의 초기 행렬을 나타낸다. 또한 이러한 마코프 모델은 식(4)와 같은 제약조건을 가진다.

$$M = \{S, \pi\}, S = \{s_{ij}\}, \pi = \{\pi_i\} \quad (1)$$

$$S = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1N} \\ s_{21} & s_{22} & \cdots & s_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ s_{M1} & s_{M2} & \cdots & s_{MN} \end{bmatrix} \quad (2)$$

$$\pi = [\pi_1, \pi_2, \pi_3, \cdots, \pi_N] \quad (3)$$

$$\sum_{k=1}^N S_{ik} = 1, \sum_{i=1}^N \pi_i = 1 \quad (4)$$

본 논문에서는 지문영상으로부터 22.5° 간격으로 추출한 16개의 방향정보를 이용하여 16개의 상태를 가지는 마코프 모델을 설계하였다^[5]. 이러한 상태경로는 $t=1$ 에서 주어진 상태열을 두 개 이상의 다른 상태들로 중복되게 할당 가능하게 하여 방향 값이 다양하게 변화되고 중복성을 지니는 지문 융선의 특징을 정확하게 추출하게 하였다. 그림 7은 Arch클래스 지문의 마코프 모델이다. 그림 7(a)는 초기확률 행렬, 그림 7(b)는 상태전이 확률 행렬을 나타낸다.

3.2 유전자 알고리즘(Genetic algorithm)을 이용한 마코프 모델 변형

3.2.1 클래스별 셀 가중치 조정

마코프 모델을 이용해 생성한 클래스별 상태전이 행렬은 그림 8과 같이 한 개 이상의 클래스의 동일한 위치에 셀 값이 존재한다. 여기서 동일 색깔을 띠는 영역은 각 클래스에 중복되는 셀을 나타낸다. 이와 같이 4개의 클래스 중 한 개 이상의 클래스와 동일하게 위치하는 셀 값들은 각 클래스의 방향 특징을 모호하게 한다. 따라서 클래스의 중복영역을 각 클래스의 유일한 영역, 2개 클래스 간 중복영역, 3개 클래스 간 중복영역, 4개 클래스 간 중복영역인 경우로 나누어 32개의 가중치 항목을 정의했다. 이때 클래스별 중복되는 셀의 조정 기준은 중복되는 셀 중 1순위의 확률 값의 20%를 기준으로 나머지 셀들과의 차

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	0.089	0.133	0.1078	0.018	0.0001	0.0009	0.0027	0.0072	0.0045	0.0189	0.0234	0.1446	0.1482	0.1869	0.0039	0.0006
(a)																
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	0.4229	0.4667	0.0396	0.0001	0.0001	0.0021	0.0001	0.0001	0.0001	0.0062	0.0104	0.0062	0.0042	0.0125	0.0042	0.025
2	0.054	0.6705	0.2177	0.0096	0.0001	0.0001	0.029	0.0001	0.0001	0.0001	0.0077	0.027	0.0038	0.0077	0.0001	0.0001
3	0.0017	0.0596	0.7169	0.202	0.005	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0116	0.0017	0.0017	0.0017	0.0033
4	0.0039	0.0155	0.2674	0.6822	0.0155	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0078	0.0001	0.0039	0.0001	0.0039
5	0.0001	0.0001	0.0001	0.1961	0.0196	0.7647	0.0001	0.0001	0.0001	0.0001	0.0196	0.0001	0.0001	0.0001	0.0001	0.0001
6	0.0001	0.0001	0.0001	0.0001	0.0001	0.5556	0.2222	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.2222
7	0.0417	0.0001	0.0001	0.0001	0.0001	0.25	0.5833	0.0001	0.0833	0.0001	0.0001	0.0001	0.0417	0.0001	0.0001	0.0001
8	0.0357	0.0001	0.0001	0.0001	0.0001	0.0357	0.4286	0.3929	0.0357	0.0001	0.0357	0.0357	0.0001	0.0001	0.0001	0.0001
9	0.0001	0.0001	0.0001	0.0001	0.0001	0.0385	0.1923	0.3077	0.2692	0.0385	0.0001	0.0001	0.0001	0.0001	0.1154	0.0385
10	0.0001	0.0001	0.0001	0.0556	0.0001	0.0001	0.0278	0.2222	0.3333	0.1944	0.0833	0.0556	0.0001	0.0001	0.0001	0.0278
11	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0294	0.0588	0.0441	0.2059	0.5294	0.1176	0.0001	0.0001	0.0147	0.0001
12	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0185	0.0556	0.0185	0.0926	0.4444	0.2778	0.037	0.0185	0.0001	0.037
13	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0213	0.0213	0.0426	0.0213	0.1277	0.1064	0.234	0.2979	0.1064	0.0213
14	0.0075	0.0001	0.0001	0.0001	0.0001	0.0001	0.0015	0.0045	0.003	0.0075	0.0075	0.0075	0.0179	0.6403	0.3	0.003
15	0.0542	0.0214	0.0001	0.0013	0.0001	0.0001	0.0001	0.0013	0.0001	0.0013	0.005	0.0001	0.0113	0.1385	0.5101	0.2557
16	0.3227	0.1102	0.0176	0.0032	0.0001	0.0001	0.0001	0.0016	0.0001	0.0048	0.008	0.0032	0.0064	0.0096	0.0687	0.4441
(b)																

그림 7. Arch클래스의 마코프 모델

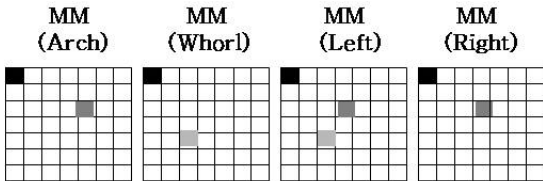


그림 8. 클래스별 동일 위치 셀 값

이 값을 산출하여 가중치의 증가 및 감소 항목을 결정했다. 그림 9는 1개~4개 중복 클래스에 대해 클래스별 셀 가중치 결정 조건을 나타낸다.

①~④는 중복되는 클래스의 셀별 확률 값의 우선순위를 나타내고, 가중치 조정은 가중치 값을 증가 또는 감소하는 두 가지 항목으로 설정 하였다. <각 클래스의 유일한 영역>은 0.001 이상인 경우 해당 셀의 확률 값을 증가 시킨다. <2개 클래스간 중복 영역>은 1순위 셀 값과 2순위 셀 값을 비교하여 그 값이 유사하면 확률 값을 둘 다 감소시키고 1순위 셀 값이 셀의 조정 기준에 의해 충분히 크면 1순위 셀의 확률 값을 증가 시키고 2순위 셀의 확률 값을 감소 시킨다. <3개 클래스간 중복 영역>은 1~3순위 셀의 값을 비교하여 셀의 조정 기준에 의해 3개의 셀이 모두 유사하면 확률 값을 감소 시키고, 1순위 보다 2, 3 순위 셀의 확률 값이 낮으면 1순위 셀의 확률 값을 증가 시키고, 2, 3 순위 셀의 확률 값을 감소 시킨다. <4개 클래스간 중복 영역>은 1~4순위 셀의 값을 비교하여 셀의 조정 기준에 의해 4개의 셀이 모두 유사하면 확률 값을 감소 시키고, 1순위 보다 2, 3, 4 순위 셀의 확률 값

```

/* ①>②>③>④: 공통 셀의 확률 값 크기 순위 */
/* ↑(가중치 증가), ↓(가중치 감소) */

<각 클래스의 유일한 영역>
if 0.001 이상 then ↑
if 0.001 이하 then 0.000001

<2개 클래스간 중복 영역>
if ① = ② then ①↓, ②↓
if ① > ② then ①↑, ②↓

<3개 클래스간 중복 영역>
if ①>③ && ①>②
then ①↑, ②↓, ③↓
else if ① = ② then ①, ②, ③↓

<4개 클래스간 중복 영역>
if ①>④ && ①>③ && ①>②
then ①↑, ②↓, ③↓, ④↓
else if ① = ② then ①, ②, ③↓, ④↓
else if ① = ③ then ①, ②, ③, ④↓
    
```

그림 9. 클래스별 셀 가중치 조정

이 낮으면 1순위 셀의 확률 값을 증가 시키고, 2, 3, 4 순위 셀의 확률 값을 감소 시킨다. 이러한 방법은 각 클래스의 방향특징을 보다 명확하게 할 수 있다.

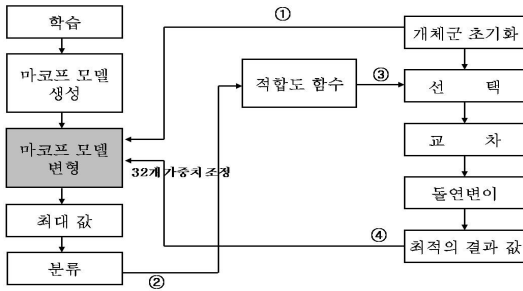


그림 10. 마코프 모델의 생성 및 변형

3.2.2 유전자 알고리즘

유전자 알고리즘은 자연적 진화 현상을 모방한 것으로, 환경에 잘 적응하는 유전자만을 선택하여 교차, 돌연변이 연산을 통해 적합도가 우수한 해를 탐색하고 세대가 되풀이 되면서 가장 우수한 해를 찾아내게 된다⁷⁾.

본 논문에서는 지문의 각 클래스를 대표하는 방향 변화 값을 최적화 하기위해 유전 알고리즘을 이용하여 마코프 모델로부터 생성된 클래스별 상태전이 행렬을 변형시킨다. 원본 마코프모델의 상태전이 행렬에서 32개의 가중치 조정 항목이 결정되면, 초기 개체군 100개를 랜덤으로 생성하여 마코프 모델을 수정한다. 수정된 마코프 모델들로 테스트 데이터를 분류하고 분류결과가 좋은 상위 10%의 개체군을 복사하고 나머지 90%의 개체군은 새로 생성한다. 생성된 개체군으로 다시 유전자 알고리즘 연산을 반복한다. 1000회 반복 연산 후 얻어진 가중치 32개 (최적의 분류 값을 산출하는 32개 가중치)를 원본 마코프 모델의 가중치 조정 항목에 적용해 최종적으로 최적화된 마코프 모델을 산출하고 지문분류에 적용한다.

그림 10은 유전자 알고리즘을 이용하여 마코프 모델을 변형하는 과정이다. 단계①에서는 마코프 모델의 변형을 위해 초기 가중치 값을 설정한다. 단계②에서는 분류결과를 적합도 함수를 이용해 평가한다. 이때, 적합도 함수는 가장 좋은 해와 가장 나쁜 해를 기준으로 하는 품질 비례 방법을 사용한다. 단계③에서는 산출된 적합도를 기준으로 가장 우수한 해의 10%로는 복사를 하고 나머지 90%에 대해서는 룰렛 휠(Roulette wheel) 방법을 적용하여 선택한다. 단계④에서는 교차와 돌연변이 연산을 통해 최적의 가중치 값을 찾아낸다.

4. 실험 및 결과

4.1 실험 데이터

본 논문에서는 다양한 품질의 지문 분류 실험을 위해



그림 11. 한 사람 지문의 8개 상태

표 1. 지문의 4개 클래스 분포

Database		A	W	L	R	Ambiguous data	Total
FVC 2000	DB1	152	184	256	224	64	880
	DB2	168	208	232	240	32	880
FVC 2002	DB1	88	144	288	288	72	880
	DB2	64	216	296	256	48	880

다양한 품질의 지문으로 구성되어 있는 FVC2000, FVC2002 데이터를 사용하였다^{8,9)}. 이들은 optical sensor로 입력된 그레이 스케일의 지문이미지로 110명으로부터 8번씩 입력된 880개로 구성되어있다. 그림 11은 한 명으로부터 입력 받은 지문이미지가 서로 다른 8개의 상태로 구성된 예이다.

본 연구에서는 지문을 4개의 클래스(Arch, Whorl, Left, Right)로 분류한다. 각 데이터베이스의 전체 데이터 880개 중에서 4개의 클래스에 대한 분포는 표 1과 같다. 여기서, 모호성 데이터란 4개의 클래스를 기준으로 했을 때 지문영상의 모호성 때문에 어느 클래스에도 해당되지 않는 데이터이다. 심지어 전문가가 정의하기 힘든 형태로 본 연구에서는 일반적인 클래스 분류를 대상으로 하기 때문에 본 데이터를 실험에서 제외 시켰다. 지문분류 모델 생성을 위한 학습 데이터와 지문분류 모델의 성능 평가를 위한 테스트데이터 구성은 다음과 같다. FVC2000, 2002 데이터베이스는 8개의 지문이 1개의 set이다. FVC2000의 경우 816개의 지문에 대해 102개의 set이 존재하고 이중 절반은 학습데이터로 나머지 절반은 테스트 데이터로 사용했다. FVC2002는 808개의 지문에 대해 101개의 set이 존재하고 이중 400개는 학습데이터로 나머지 408개는 테스트 데이터로 사용했다.

4.2 마코프 모델의 학습 및 변형

입력받은 지문이미지는 전처리, 특징추출 단계를 거친 후 마코프 모델로 만들어 진다. 생성된 4개 클래스 마코프 모델의 상태전이 행렬들은 각 셀 값의 비교 단계를 거쳐 32개의 가중치 항목이 결정되고, 이 가중치 항목은 유전자 알고리즘을 통해 최적의 분류모델로 변형된다. 유전자 알고리즘 실험에는 FVC2000 DB2에서 420개, FVC2002

표 2. 마코프모델의 분류성능 비교

데이터베이스	분류 방법	학습 데이터	테스트 데이터	정확도(%)
FVC2000	Markov model	FVC2000 DB1 (408)	FVC2000 DB1 (408)	66.5
	Markov model + GA	FVC2000 DB2 (420)	FVC2000 DB1 (408)	86.5
FVC2002	Markov model	FVC2002 DB1 (400)	FVC2002 DB1 (408)	69.5
	Markov model + GA	FVC2002 DB2 (420)	FVC2002 DB1 (408)	88.6

표 3. 지문의 품질별 분류 결과

데이터 베이스	낮은 품질	부분 입력	높은 품질	Total
FVC2000	89.7	81.9	87.8	86.5
FVC2002	88.1	88.0	88.8	88.6

DB2에서 420개의 데이터를 사용했고 1000번의 세대 반복수와 -0.3~0.3 사이의 고정 돌연변이 변수를 이용하였다. 유전알고리즘을 적용하여 최적화 시킨 마코프모델의 분류성능은 표 2에 나타내었다. FVC2000 DB1과 FVC2002 DB1의 테스트 데이터에 원본 마코프 모델과 GA를 이용하여 최적화 시킨 변경된 마코프 모델을 각각 적용한 결과이다. 비교결과 마코프모델의 최적화를 위해 유전알고리즘 연산을 적용 후 지문분류 성능이 평균 19.6% 증가했고 이는 효과적으로 마코프 모델이 최적화 된 것을 보여준다.

4.3 분류 결과 분석

본 연구는 지문의 다양한 품질 차이를 극복하여 지문 분류율을 높이는 방법을 제안한다. 따라서 기존의 특이점 정보를 사용한 지문분류 방법과는 달리 각 지문 용선의 방향 특징만을 이용한다.

본 연구에서 제안한 방법의 분류성능을 비교분석 하기 위해 품질별 분류 정확도를 표 3에 나타내었다. 이는 실험 대상의 지문을 품질별(낮은 품질, 부분 입력, 높은 품질)로 나누어 분류한 결과이다. FVC2000 데이터베이스의 경우 전체 분류 정확도가 86.5% 이다. 이때, 부분 입력 지문 경우 81.9%로 낮은 품질, 높은 품질 지문의 경우보다 정확도가 낮게 나왔다. 이와 같은 결과는 부분 입력 지문 데이터가 회전조정 단계에서 오류가 난 경우이다. 즉, 지문이미지의 방향 회전이 10°미만인 데이터인데도 불구하고 10°이상으로 판단하여 잘못 회전을 수행한 경우이다. 반면, FVC2002 데이터베이스는 낮은 품질, 부분 입력, 높은 품질의 지문에 대한 분류 정확도가 유사하게 나왔다. 이것은 지문의 회전조정 및 특징산출이 양호하게 수

표 4. 관련연구의 접근 방법 및 결과

접근방법	클래스 개수	에러율 (%)	분류에 사용한 특징
Karu and Jain 1996 ^[2]	4	8.6	특이점의 개수, 위치정보
A.K.Jain, S. Prabhakar, L. Hong 1999 ^[10]	4	5.2	특이점의 개수, 위치정보
R. Cappelli, D. Maio 1999 ^[3]	5	12.9	특이점 정보 기반 용선의 방향 특징 추출
LI-MIN, CHING-YU 2008 ^[11]	4	5.61	특이점 추출 후 용선의 방향 특징 템플릿 생성

행되었으며 지문의 품질에 영향을 받지 않음을 나타낸다.

제안한 방법의 성능을 분석하기 위한 기존 관련연구의 접근방법 및 분류결과는 표 4와 같다. 기존 연구 방법은 크게 두 가지로 구분된다. 첫 번째 방법은 특이점 개수 및 위치정보를 기반으로 지문 분류를 수행 한다. 두 번째 방법은 특이점을 추출 한 후 특이점 주위의 방향성분 및 패턴을 이용해 지문 분류를 수행한다. 그러나 이러한 특이점 정보 기반 지문 분류 방법은 회전 날인한 지문 데이터 즉, 중심점이나 삼각주의 개수 및 위치가 명확하게 나타나는 지문 데이터를 사용하는 경우에 적합하다. 최근 라이브 스캔 입력방식에 의해 획득한 지문 이미지는 특이점 정보의 손실 및 예외적인 입력상태의 지문이 존재한다. 따라서 본 연구에서는 광학식 센서 입력 방식으로 획득된 FVC2000, FVC2002 데이터를 이용하여 지문 분류 실험을 수행하고, 다양한 상태로 입력된 FVC2000, FVC2002 데이터베이스의 특이점 분포를 분석해 보았다. 그 결과 삼각주가 존재하지 않는 경우는 FVC2000 데이터베이스가 51%, FVC2002 데이터베이스가 55%의 분포를 보였다. 이는 입력센서의 크기가 작아서 삼각주의 위치까지 찍히지 않은 경우이다. 따라서 이러한 상태의 데이터베이스에 기존 특이점기반 연구 방법을 적용하게 되면 삼각주

가 존재 하지 않는 지문의 경우 분류가 불가능 하여 평균 지문 분류율이 현저히 낮아질 것이다. 또한 FVC2000 데이터베이스의 29%, FVC2002 데이터베이스의 23%는 중심점의 위치가 오류인 경우이다. 이것은 지문이미지가 한 쪽으로 치우친 경우로 예외적인 상황의 지문입력에 해당한다. 기존 특이점 정보를 기준으로 방향패턴을 분석하여 지문 분류를 수행하는 방법을 이러한 중심점 위치가 오류인 데이터에 적용 할 경우 평균 지문 분류율이 현저히 낮아질 것이다. 따라서 기존 연구 방법에 이러한 특이점 정보를 가지는 데이터베이스를 사용한다면 추가적인 에러 발생을 예측 할 수 있다.

본 연구에서 제시한 지문분류 방법은 삼각주가 존재하지 않거나 중심점의 위치가 오류인 데이터를 포함하여 실험 한 결과, 평균 분류율은 87.6%로 다양한 품질의 지문 데이터에 대해 효율적으로 지문 분류를 수행하는 것을 확인하였다. 지문 용선의 방향 특징만을 이용하여 예외적인 상태의 지문 클래스를 예측할 수 있는 확률 모델을 이용하여 다양한 품질에 무관하게 지문 분류를 수행하는 제안 방법은 특히 삼각주가 존재하지 않거나 중심점의 위치가 오류인 경우와 같은 특이점 정보가 불완전한 지문 분류 수행에 유리함을 알 수 있다.

5. 결 론

본 논문에서는 용선의 방향성 정보를 이용한 확률적인 새로운 지문 분류 방법을 제안하였다. 라이브스캔 방식으로 입력된 다양한 품질의 지문을 효과적으로 분류하기 위해, 용선의 방향특징을 이용하여 확률 기반 마코프 모델을 생성하고 각 클래스간 셀의 중복되는 확률 값의 영향을 줄이기 위해 유전자 알고리즘을 이용하여 가중치를 조절하는 방법을 제시했다. 또한 제안한 방법의 효율성 검증 위해 다양한 품질의 데이터를 이용해 실험과 분석을 하였으며, 이를 통해 제안한 방법이 특이점이 존재하지 않거나, 불분명한 경우 또는 지문이미지가 부분 입력된 경우, 품질이 낮은 경우에도 효과적으로 지문분류를 수행하는 것을 확인했다.

참 고 문 헌

1. D. Maltoni, D. Maio, A. K. Jain, S. Parbhankar, Handbook of Fingerprint Recognition, Springer, 2003.
2. K. Karu and A. K. Jain, "Fingerprint classification," Pattern Recognition, vol. 29, no. 3, pp. 389-404, 1996.
3. D. Maio and D. Maltoni, "A Structural Approach to Fingerprint Classification," Proc. 13th ICPR, Vienna, vol. 3, pp. 578, Aug. 1996.
4. R. Cappelli, A. Lumini, D. Maio, and D. Maltoni, "Fingerprint classification by directional image partitioning," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 21, no. 5, pp. 402-421, 1999.
5. H-W. Jung and J-H. Lee, "Fingerprint Classification Using the Stochastic Approach of Ridge Direction Information," IEEE International Conference on Fuzzy Systems, pp. 169-174, 2009.
6. Alberto Leon-Garcia, Probability, Statistics, and Random Processes for electrical Engineering, 2nd edition, PrenticeHall, 1994.
7. D. Goldberg, Genetic Algorithms in Search, Optimization, and Machine Learning, Addison Wesley, 1989.
8. D. Maio, D. Maltoni, R. Cappelli, J. L. Wayman, A. K. Jain, FVC2000: Fingerprint Verification Competition. Biolab internal report, University of Bologna, Italy, September 2000. Available from <http://bias.csr.unibo.it/fvc2000/>
9. D. Maio, D. Maltoni, R. Cappelli, J. L. Wayman, A.K. Jain, FVC2002: Fingerprint Verification Competition. Biolab internal report, University of Bologna, Italy, April 2002. Available from <http://bias.csr.unibo.it/fvc2002/>
10. A. K. Jain, S. prabhakar, L. Hong, "A Multichannel approach to fingerprint classification," IEEE Trans, Pattern Analysis and Machine Intelligence, vol. 21, no. 4, pp. 348-359, 1997.
11. Li-Min and Ching-Yu, "A Directional approach to Fingerprint classification," International Journal of Pattern Recognition and Artificial Intelligence, vol. 22, no. 2, pp. 347-365, 2008.



정혜옥 (wukj@skku.edu)

1999 한성대학교 정보전산학부 정보공학과 학사
2005 성균관대학교 정보통신대학원 정보보호학과 석사
2005~현재 성균관대학교 컴퓨터공학과 박사과정

관심분야 : 패턴인식, 지능시스템, 정보보호



이지형 (jhlee@ece.skku.ac.kr)

1993 한국과학기술원 전산학과 학사
1995 한국과학기술원 전산학과 석사
1999 한국과학기술원 전산학과 박사
2002~현재 성균관대학교 정보통신공학부 부교수

관심분야 : 지능시스템, 기계학습, 사용자 모델링