

매칭 점수 분포를 이용한 효율적인 1:N 지문 매칭 알고리즘

Efficient 1:N Fingerprint Matching Algorithm using Matching Score Distribution

김 경 민*, 박 중 조, 이 범, 고 영 진, 정 순 원
(Kyoung-Min Kim¹, Joong-Jo Park², Buhm Lee¹, Young-Jin Go¹, and Soon-Won Jung³)

¹Chonnam National University

²Gyeongsang National University

³RIA Soft

Abstract: This paper presents two adaptive fingerprint matching methods. First, we experiment an adaptive threshold selection of 1:N matching system in order to raise the reliability of the matching score. Second, we propose a adaptive threshold selection using fitting algorithm for high speed matching. The experiment was conducted on the NITZEN database, which has 5247 samples. Consequently, this paper shows that our suggested method can perform 1.88 times faster matching speed than the bidirectional matching speed. And, we prove that FRR of our suggested method decreases 1.43 % than that of the unidirectional matching.

Keywords: fingerprint recognition, matching algorithm, bidirectional matching, FRR

I. 서론

생체 기반 인증기술은 인간의 생체학적인 정보를 이용한 방법으로 지식기반, 소유기반방식에서 발생할 수 있는 분실 문제가 없고, 개인마다 독특한 특성을 가짐으로 보안 및 인식 분야에서 많이 활용되고 있다. 이상적인 생체특징으로 보편성(universal), 유일성(unique), 영속성(permanent), 정량성(collectable)을 가져야 하는데, 그 중 지문은 땀샘이 용기되어 일정한 흐름을 형성한 것으로 모양이 개개인마다 서로 다르고, 태어날 때의 모습 그대로 평생 변하지 않는다는 장점 때문에 식별 성능에 대한 신뢰도와 안정도가 높아 가장 효율적인 개인 인증 방법으로 이용되고 있다[1,2].

지문에 대한 개인 인증분야는 크게 분류(classification)와 매칭(matching)으로 나뉜다. 분류는 여러 모양을 갖는 지문의 형태들을 어떤 공통된 특징을 기준으로 나누는 작업이고, 매칭은 입력 지문과 데이터베이스 내의 등록지문과의 일치여부를 비교하여 특정 개인을 식별하는 작업이다. 또한, 매칭은 식별(identification)과 인증(verification)로 구분된다. 식별은 범죄 현장의 유류지문과 같은 입력 정보를 이용해 대량의 데이터베이스 중에서 동일한 등록 지문을 찾는 것이며, 인증은 지문과 함께 개인의 비밀번호나 카드를 함께 입력하여 등록된 본인의 지문과 동일 여부를 판별하는 것으로 출입 통제 시스템에서 자주 사용하는 방법이다.

매칭의 식별에서는 큰 데이터베이스 내에서 빠른 매칭과 정확도를 위해 개발이 되며, 매칭방법 중의 하나인 특징점을 이용한 지문 매칭의 경우, 특징점간의 기하학적 유사성을 분석하여 매칭 점수를 계산한다. 이러한 기하학적 유사성으로 매칭을 할 경우, 하나의 지문 데이터를 기준으로 다른 하나의 지문 데이터를 미리 정렬하는 과정이 필요하고, 이는 정렬 결과에 따라 두 지문의 유사도가 달라지므로 보통 특징점 기반 매칭기법에서는 양방향 매칭을 통하여 최종 매칭 점수를 구한다. 이때 이용한 양방향 매칭 기법은 단방향 매칭에 비해 매칭 신뢰도는 높지만, 매칭 시간이 단방향 매칭보다 두 배 이상 걸린다는 문제점이 존재한다[3].

이러한 특징점 기반 매칭기법의 문제점인 양방향 매칭시간이 오래 걸리는 단점을 해결하고, 단방향 매칭보다 신뢰도가 높은 매칭을 이루기 위해, 본 연구에서는 전반적인 지문인식시스템의 구성을 확인 후 특징점 추출에 대한 과정과 지문의 분류방법들을 살펴보고자 한다. III 장에서는 1:N 매칭방법의 문제점인 매칭 속도 측면을 해결하기 위한 매칭 알고리즘을 설명한 후, IV 장에서 대규모 지문데이터베이스 기반으로 1:N 매칭의 임계치에 따른 새로운 적응적 임계치 설정으로 인한 실험과 함께 새로운 적응적 임계치 설정을 위한 fitting기법 알고리즘을 활용하여 실험을 하고자 한다. 끝으로 V 장에서 결론으로 본 논문을 구성한다.

II. 본론

지문인식시스템은 일반적으로 지문 입력장치와 인식알고리즘으로 구성되며, 대규모 인식시스템에는 지문분류, 특징점 추출과정, 매칭과정 세 가지로 나누어진다.

한 사람의 지문이 어떤 분류에 속하는지를 결정하는 것은 지문 분류를 대규모 지문 데이터베이스에서 인덱스로

* 책임저자(Corresponding Author)

논문접수: 2011. 9. 30., 수정: 2011. 12. 18., 채택확정: 2012. 1. 6.

김경민, 이범, 고영진: 전남대학교 전기전자통신컴퓨터공학부

(kkm@jnu.ac.kr/buhmlee@jnu.ac.kr/jericho90@naver.com)

박중조: 경상대학교 제어계측공학과(jjpark@gnu.ac.kr)

정순원: (주)RIA Soft(jsw123@naver.com)

※ 본 논문은 2008년 전남대학교 교내학술연구비 지원에 의한 것임.

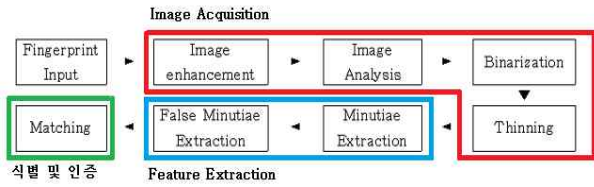


그림 1. 지문인식 시스템의 구성도.
Fig. 1. Scheme for fingerprints recognition system.

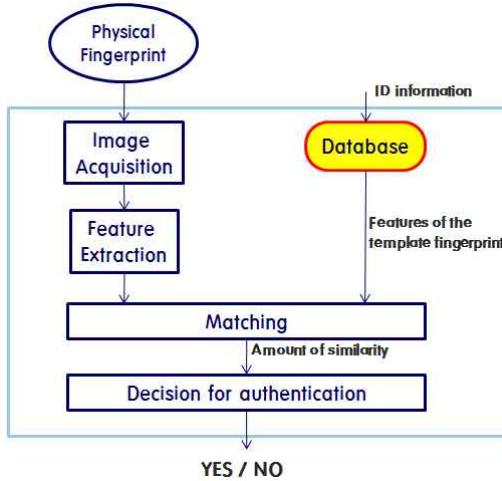


그림 2. 1:1 매칭시스템.
Fig. 2. 1:1 matching system.

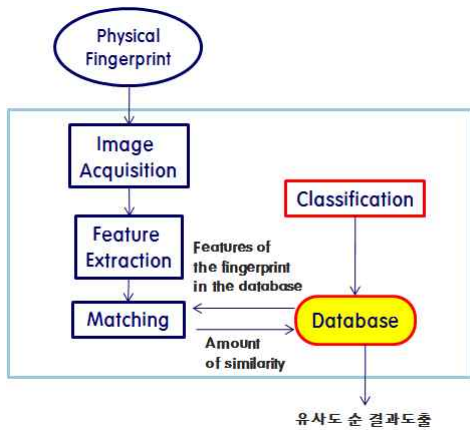


그림 3. 1:N 매칭시스템.
Fig. 3. 1:N matching system.

사용할 수 있어 매칭 시간을 단축할 수 있다. 대부분의 지문인식시스템은 데이터베이스의 크기가 작은 경우(1000개 이하의 지문영상)에는 높은 정확성을 나타내지만, 데이터베이스의 크기가 증가함에 따라 인식율은 떨어지게 된다[4]. 또한, 인식시스템의 처리속도도 감소하게 된다. 그러나 이런 문제는 데이터베이스에 있는 모든 지문영상들을 미리 정의된 분류기준에 의해 몇 개의 작은 부분으로 나누어 인식 속도와 정확성 문제를 해결 할 수 있다.

일반적으로 지문인식시스템은 매칭 방법에 따라 1:1 매칭시스템의 인증시스템(verification system)과 1:N 매칭시스

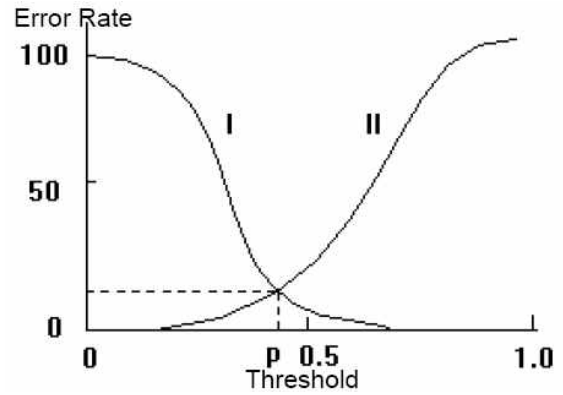


그림 4. 임계값 변화에 의한 오인식률의 증감.
Fig. 4. FRR variation for various threshold.

템의 식별시스템(identification system)으로 분류된다[5,6].

1:1 매칭 시스템은 지문 인증시스템(verification system)으로서 지문 데이터베이스에 본인의 지문과 ID를 등록하고, 본인 여부를 확인하기 위해 ID를 입력하고 그 ID의 지문과 일치 여부를 판단하는 인증시스템이다. 이 방법은 현재 노트북의 지문 인식 등 개인 사생활 보호 장치에 주된 용도로 사용되고 있다. 1:N 매칭시스템의 지문 식별시스템(identification system)은 획득 지문을 가지고 범죄자를 찾는 것으로 획득한 지문의 특성을 추출하여 기준에 구축된 방대한 양의 데이터베이스에서 그 지문에 가장 가까이 일치하는 후보를 골라내는 조회 시스템이다.

일반적인 매칭 방법은 국부 좌표계를 이용한 지문 매칭과 용선 방향을 이용한 매칭 그리고 MST (Minimal Spanning Tree)를 이용하는 매칭 등이 있다[6-8].

어떠한 종류의 지문 식별 및 인증방법을 사용하든지 지문인식의 목적상 타인의 지문을 동일 지문으로 인증하는 것은 치명적인 문제점을 초래할 수 있기 때문에 지문식별 방법의 효율성은 전체적인 오인식률을 최소화하는 것보다는 타인 지문을 동일 지문으로 오인식하는 그림 4의 II 오류를 0%로 하였을 때, 동일 지문을 타인 지문으로 오인식하는 I 오류를 얼마나 최소화 하느냐로 판정된다.

통상 지문의 특징점을 이용하여 지문 매칭을 수행하는 경우 특징점간의 기하학적 유사성을 분석하여 두 지문의 매칭 점수를 계산한다. 이러한 기하학적 유사도를 계산하기 위해서는 하나의 지문 데이터를 기준으로 다른 또 하나의 지문 데이터를 미리 정렬하는 과정이 필요하며, 정렬결과에 따라 두 지문의 유사도가 달라지므로 통상의 지문 매칭에 있어서는 양방향 매칭을 통하여 최종 매칭 점수를 구한다. 양방향 매칭의 경우, 단방향매칭에 비하여 매칭의 신뢰도부분에 있어서 좋으나 매칭에 걸리는 시간이 단방향에 비해 두 배로 걸린다는 문제점이 있다. 시간적 부분의 문제점을 해결하기 위하여 본 논문에서는 단방향 매칭과 양방향 매칭의 병행 시스템을 적용하여 보고, 새로운 임계치를 자동적으로 탐색하는 적응적 임계치를 연구하여 봄으로써 매칭 시간이 오래 걸리는 단점을 보완하고자 한다. 더불어 FRR에 대한 부분에 있어서 최적의 값을 찾기 위해 단방향, 양방향매칭 병행시스템을 제안한다[3].

III. 매칭점수 분포를 이용한 효율적인 1:N 시스템

1. 일반적인 단방향, 양방향 매칭방법[3]

그림 5는 지문의 매칭과정을 나타내며, 여기서 특징점 분석과 국부 유사도 분석은 유사도 계산을 위한 기하학적 정렬과정이라고 볼 수 있다. 이러한 기하학적 정렬과정 때문에 백터기반의 유사도 측정결과와는 다르게 매칭하고자 하는 두 지문데이터 중 어떤 것을 기준으로 정하느냐에 따라 정렬결과가 달라 질 수 있으며 이에 따라 두 지문 간에 계산되는 유사도 혹은 매칭점수도 바뀌게 된다.

그림 6은 지문매칭을 하였을 경우의 정렬결과를 보여준다. 그림 6에서 (a)는 등록지문, (b)는 입력지문, (c)는 등록 지문에서 추출한 특징점들을 기준으로 입력지문에서 추출한 특징점들을 기하학적으로 정렬한 결과, (d)입력지문에서 추출한 특징점들을 기준으로 등록지문에서 추출한 특징점들을 기하학적으로 정렬한 결과를 나타낸다.

$$MS_{Max} = Max(S_{st}, S_{ts}) \tag{1}$$

$$MS_{Average} = Average(S_{st}, S_{ts}) \tag{2}$$

(1)과 (2)에서 일반적으로 1:1 지문인증단계에서 MS는 매칭점수(Matching Score), S_{st} 는 등록되어있는 지문 특징점(template)들을 기준으로 입력지문(sample)에서 추출한 특징점들을 기하학적으로 정렬하여 구한 매칭점수를 나타내며, S_{ts} 는 반대로 입력지문(sample)에서 추출한 특징점들을 기준으로 등록되어있는 지문 특징점들을 기하학적으로 정렬

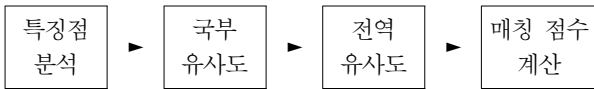


그림 5. 매칭 과정.
Fig. 5. Matching process.

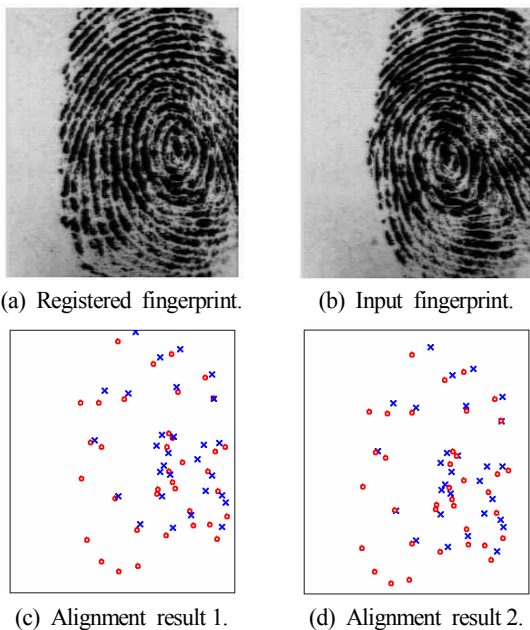


그림 6. 지문 매칭쌍 및 매칭을 위한 정렬결과.
Fig. 6. Fingerprints pair and alignment results.

하여 구한 매칭점수를 나타낸다.

여기서 $Max()$ 는 두 입력 변수 중 큰 값을 출력하는 함수이고, $Average()$ 는 두 입력변수의 산술 평균을 출력하는 함수이다. 본 실험에서는 (1)의 방법으로 매칭점수를 구하였다.

이와 같은 방법으로 그림 6의 지문데이터의 매칭점수를 [0, 9999]로 환산하여 구한 결과는 (c)의 경우 3623점, (d)의 경우 4279점을 얻어지며, 양방향매칭간의 차이가 있음을 알 수 있다[1].

이러한 이유로 실제 지문인식시스템에서 두 지문간의 매칭점수를 구하는 경우 그림 6(c)의 정렬결과로부터 구한 값과 (d)의 정렬결과로부터 구한 값의 최대치를 두 지문간의 최종 매칭점수로 사용하며, 알고리즘의 FRR (False Reject Rate), FAR (False Acceptance Rate)을 결정하는 여러 단계의 임계값도 다양한 지문 데이터베이스에 대한 양방향 매칭결과에 따라 설정한다[6].

지문인증을 위하여 하나의 입력지문 이미지와 하나의 등록지문 특징데이터를 매칭하는데 필요한 시간은 식 (3)과 같이 입력 지문이미지로부터 특징 데이터 추출을 위해 소요되는 시간과 양방향 매칭에 소요되는 시간의 합으로 표현된다.

$$T(\text{인증}) = T(\text{추출}) + T(\text{매칭}) \times 2 \tag{3}$$

여기서 T()는 처리에 필요한 시간을 나타내며 (1)의 최댓값을 구하는데 걸리는 시간은 무시할 수 있으므로 표시하지 않았다.

이는 사용되는 알고리즘에 따라 1:1지문 인증 시에는 단방향매칭에 비해 양방향매칭에 두 배의 시간이 걸린다 하더라도 전체 인증시간에 큰 영향을 미치지 않는다[3]. 하지만, 1:N지문인식 시에는 시간적 부분에 있어서 문제가 발생한다. 예를 들어 사용자 N을 1000명이라고 가정하면, (3)과 1:1지문인증방법에 의해 다음과 같이 구할 수 있다.

$$\begin{aligned} T(\text{추출}) &= 500ms \\ T(\text{매칭}) &= (N \times 10ms) * 2 \\ \therefore T(\text{인증}) &= 20.5sec \end{aligned}$$

이처럼 최악의 경우 사용자는 20.5초 후에 매칭 실패메시지를 받게 된다. 이러한 문제점을 해결하기 위하여 사용 지문인식 출입통제기 등의 경우에는 사용자가 ‘Time Out’ 시간을 설정할 수 있도록 하여 정해진 ‘Time Out’시간 내에 1:N 매칭이 실패하였을 경우, 매칭 실패메시지를 내보낸 후 사용자로 하여금 다시 지문을 입력할 수 있도록 하고 있다. 여기서 ‘Time Out’시간을 2초로 설정하였을 경우 특징 데이터 추출에 필요한 시간 500 ms를 빼면 약 1.5초의 최대 매칭시간이 주어진다.

이때, T_m 을 매칭 시간, T_{Mm} 을 최대 매칭시간, M_n 을 매칭 횟수라고 가정하여, $T_m = 10ms$, $T_{Mm} = 1.5s$ 이라는 조건을 가질 경우, M_n 은 다음과 같다.

$$M_n = \frac{T_{Mm}}{T_m} = \frac{1.5s}{10ms} = 150 \text{ 회} \tag{4}$$

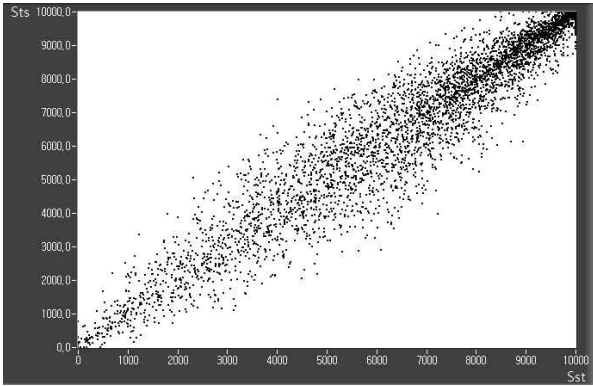


그림 7. 지문 매칭 점수 분포.

Fig. 7. Matching score distribution.

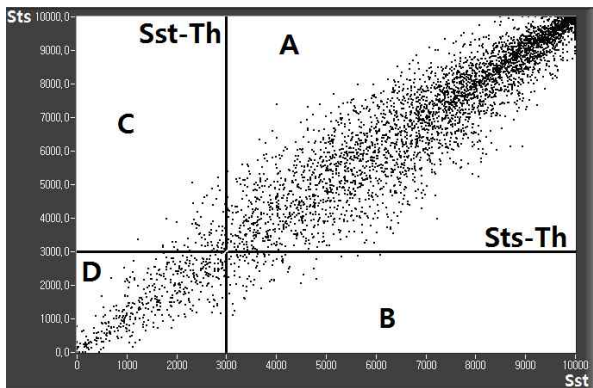


그림 8. 임계치에 따른 영역분할.

Fig. 8. Partitioning of region.

$$\text{단위시간당 매칭 회수} = \frac{M_n}{\text{평균매칭 회수}} \quad (5)$$

(4)와 (5)에 의해 양방향 매칭의 경우(평균매칭회수=2) 단 위시간당 75번의 매칭을 수행할 수 있다.

양방향 매칭 시, (1)에서 S_{st} 와 S_{ts} 의 매칭결과 값이 거의 차이가 없다면 양방향 매칭을 수행 할 필요 없이 단방향 매칭만으로도 오거부율이 증가하지 않은 상태에서 성공 적으로 매칭이 되며 매칭여부를 판단 가능하다. 이와 반대로 만약 S_{st} 와 S_{ts} 의 매칭결과 값이 차이가 크게 되면 단방향 매칭만으로도 매칭여부를 판단하는 것은 어려움이 따른다. 그러나 양방향 매칭 값의 차이는 양방향 매칭을 수행해 보아야만 구할 수 있으므로 이러한 사실만으로 기존의 양방향 매칭 방법과 차별되는 방법을 찾기는 어렵다.

그러나 많은 지문 매칭 쌍에 대하여 양방향 매칭 점수 차이의 분포를 구하여 보고, 이 분포와 매칭에 사용되는 임 계치를 이용하여 단위시간 내에 훨씬 많은 매칭을 수행할 수 있다.

그림 7은 실제 5247개의 지문 매칭 쌍에 대한 분포를 확인하기 위하여 S_{st} 와 S_{ts} 2차원공간상에서 데이터를 표기 하였을 때의 양방향 매칭 점수 분포를 보여주고 있으며, 그림 8은 지문 매칭 점수 분포에 대하여 임계치 3000으로 설정하였을 때 매칭영역을 나타내어 준다. 통상 지문인식에서

사용되는 임계치의 값은 매칭 알고리즘에 따라 차이를 보 일 수 있는데 본 논문에서는 이해를 돕기 위해 임계치를 3000으로 설정하여 설명한 뒤, IV 장에서 다양한 임계치의 값에 따른 데이터를 실험 및 연구하도록 한다.

만일 S_{st} 만을 기준으로 단방향 매칭을 수행 할 경우, A 영역과 B영역은 매칭 성공 매칭 쌍으로 분류되고 C영역과 D영역은 매칭 실패로 인식된다. 시간적 문제를 해결하기 위해 S_{st} , S_{ts} 둘 중 하나의 임계치를 기준으로 단방향 매 칭을 수행하였을 경우, 매칭 시간은 줄어들지만 오거부율인 FRR이 증가한다는 것을 알 수 있다.

1:N 매칭에서 양방향 매칭은 동일 시간 내에 75회의 매 칭을 수행하며 단방향 매칭은 동일 시간 내에 150회 매칭 을 수행한다는 것을 (4), (5)에 의해 알 수 있다. 단방향 매 칭의 수행은 단위시간 내에 양방향 매칭보다 많은 매칭을 수행할 수 있다는 장점이 있으나 매칭 점수의 신뢰성이 떨어지고 FRR이 커지며 FAR이 줄어드는 결과를 얻는다. 따라서 양방향 매칭을 가정하고 설정한 임계값을 단방향 매 칭에 맞추어 다시 설정해야 하는 문제가 있다.

이전 [3]의 논문에서는 매칭 시간적 문제와 오거부율의 문제를 해결하기 위해 보조 임계치 Th' 을 설정해 주었다.

그림 9는 보조 임계치 Th' 설정을 나타내는 그림이다. 이 처럼 기존의 임계치 Th 외에 새로운 임계치 Th' 를 새로 고 려하여 매칭 쌍에 대한 매칭을 수행하도록 한다. 기존의 임 계치의 값만을 통하여 매칭을 수행하였을 경우, 4개의 영역 으로 나누어진 것을 확인할 수 있었지만, 보조 임계치 설정 으로 인하여 그림 9처럼 총 6개의 영역으로 분할됨을 확인 할 수 있다.

만약 S_{st} 를 기준으로 단방향 매칭을 시킨다고 가정하였 을 경우, 위에서 언급하였듯이 A영역과 B영역에 존재하는 매칭 쌍에 대해서만 매칭 성공으로 판별되고 양방향 매칭 을 수행하였을 시 매칭 성공으로 인정되어야 하는 부분인 C영역이 매칭 실패로 구분된다. 이때, S_{st} 만으로 단방향 매

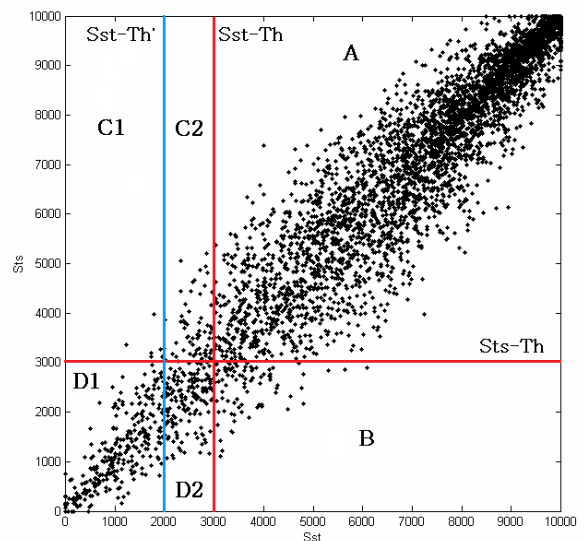


그림 9. 보조 임계치 Th' 의 설정.

Fig. 9. Decision of an auxiliary Th' .

칭을 수행한 뒤, 새롭게 추가된 임의의 임계치의 값 Th'에 의해 양방향 매칭을 시킨다면 C2영역에 존재하는 매칭 쌍은 기존의 매칭 실패에서 매칭 성공으로 바뀌게 된다.

이러한 Th'에 의한 양방향 매칭을 시켰을 경우, D2영역의 매칭쌍도 양방향 매칭의 대상이 되는데, 이는 전체적인 양방향 매칭을 수행하였을 때보다 적은 수의 매칭 쌍에 대한 양방향 매칭을 수행하기 때문에 기존의 양방향 매칭을 수행하였을 때보다 매칭시간이 줄게 된다. 또한, 단방향 매칭을 시켰을 때보다는 매칭시간을 늘어나지만, 전체적인 매칭 쌍의 일부분만을 양방향 매칭을 수행하기 때문에 매칭시간의 증가는 적다.

각 영역에 존재하는 매칭쌍의 개수를 $n(\)$ 이라 하면, 이때 발생하는 총매칭 횟수, 평균매칭횟수, 오거부율 FRR에 대해서는 (6)-(8)에 의해 구할 수 있다.

$$\text{총매칭횟수} = n(ABCD) + n(C2) + n(D2) \quad (6)$$

$$\text{평균매칭횟수} = \frac{\text{총매칭횟수}}{n(ABCD)} \quad (7)$$

$$\text{오거부율 FRR} = \frac{n(C) + n(D) - n(C2)}{n(ABCD)} \quad (8)$$

그림 9에서 생각할 수 있듯이 Th'이 Th에 가까울수록 C2 영역이 좁아지면서 매칭시간은 줄어들게 된다. 하지만, 매칭 실패율은 전반적으로 늘어 오거부율 FRR이 증가하게 된다. 이와 반대로 Th'이 Th의 값과 멀어질수록 C2영역이 커지면서 매칭시간은 증가하게 되지만, 오거부율 FRR의 경우 줄어들게 된다. Th'을 0으로 설정을 하여 매칭을 수행한다고 하더라도 전체적인 양방향 매칭을 수행하는 것이 아닌, C영역과 D영역에 대해서만 양방향 매칭을 수행하게 되므로 기존의 전체적인 양방향 매칭을 수행하였을 때보다 매칭시간은 줄어들게 되고 오거부율 FRR은 전체적 양방향 매칭을 수행하였을 때의 값인 D영역에 대해서만 나타나게 되어 기존의 방법과 비교하여 보았을 때 효율적인 것을 확인할 수 있다. 하지만, 새로운 임계치 Th'를 임의로 변경함으로써 데이터 분포에 따른 매칭횟수, 오거부율, 시간적 관계를 확인하는 것이 아닌 기존의 임계치 Th에 따른 최적의 적응 임계치 Th'이 자동적 설정이 요구된다.

최적의 적응 임계치 Th'을 자동적으로 찾기 위해 본 논문에서는 전체 지문 매칭 쌍이 $y=x$ 그래프형태에서 분산형태로 바뀐다는 점을 착안하여, 회귀분석방법을 이용하여 Fitting시켜 분산형태의 그래프를 $y=x$ 형태로 변경하고, 입력지문 혹은 매칭 계산 시 바뀌게 될 매칭점수를 예측하여 본다. 이 때 발생한 예측범위 값을 추적하여 최적의 임계치 Th'을 자동적으로 찾는데 적용하였다[7].

2. 매칭 분포특성을 이용한 임계치 설정

최소 자승법은 어떠한 데이터의 분포가 가우스 분포형태를 가질 경우 일직선 데이터로 표현하는데 적절히 사용된다. 본 논문의 데이터 분포는 직선 $y=x$ 를 기준으로 하나의 데이터 쌍에 대한 매칭점수는 큰 차이를 보이지 않으며, 데이터 쌍에 대한 분포는 가우스 분포형태를 가지게 된다. 이에 따라 역으로 분포형태에서 일직선형태로 표현하는 것

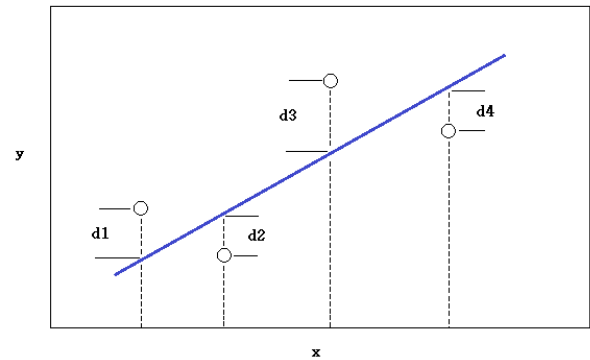


그림 10. 최소 자승법에 의한 직선.
Fig. 10. A regression line with least square method.

은 최소 자승법에 의해 가능하다. 데이터들이 분산되어 있을 때 통계적 방법을 응용하여 가장 가까운 선을 찾아내는 방법으로 가장 보편적으로 사용되는 기법이 바로 최소 자승법(method of least square)이다.

실험에 의한 n개의 실험값 $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ 에 대하여 이 점들을 지나는 어떠한 직선 $y = ax + b$ 를 그어 n개의 편차 d_1, d_2, \dots, d_n 을 얻었을 때, (d =데이터 값과 직선 값과의 차이이다) n개의 데이터 점을 지나는 최적의 직선은 편차의 제곱의 합이 최소가 되는 직선이다.

이 때 편차의 제곱의 합은 다음과 같이 구할 수 있다.

$$S = \sum_{i=1}^n d_i^2 = \sum_{i=1}^n (y_i - (ax_i + b))^2 \quad (9)$$

(9)의 결과는 \hat{a} 와 \hat{b} 의 최우추정치와 같다.

$$f(d_i) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{d_i^2}{2\sigma^2}\right) \quad (10)$$

우도함수는

$$L(d_1, \dots, d_n) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left(-\frac{\sum_{i=1}^n d_i^2}{2\sigma^2}\right) \quad (11)$$

이다. 이 우도함수는 (9)가 최소가 될 때 최댓값이 된다. 즉, S가 최소가 될 때, $y = ax + b$ 는 실험 데이터를 지나는 최적의 직선이 된다. 직선의 기울기 a와 절편 b를 결정하기 위해서는 (12)를 a와 b에 대하여 편미분하여 구한다.

$$\begin{aligned} \frac{\partial S}{\partial a} &= -2 \sum (y_i - ax_i - b) x_i = 0 \\ \frac{\partial S}{\partial b} &= -2 \sum (y_i - ax_i - b) = 0 \end{aligned} \quad (12)$$

(12)를 다시 정리하면 (13)과 같다.

$$\begin{aligned} \sum x_i y_i &= a \sum x_i^2 + b \sum x_i \\ \sum y_i &= a \sum x_i + nb \end{aligned} \quad (13)$$

(13)을 a와 b에 대하여 연립하여 풀면 (14)와 (15)처럼 a와 b를 얻는다.

$$a = \frac{n \sum_{i=1}^n x_i y_i - (\sum_{i=1}^n x_i)(\sum_{i=1}^n y_i)}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} \quad (14)$$

$$b = \frac{\sum_{i=1}^n y_i - a \sum_{i=1}^n x_i}{n} \quad (15)$$

구하여진 a 와 b 의 값에 의해 가우스 분포를 가지는 분포특성을 직선 형태로 회귀하게 할 수 있다. 본 논문에서 사용한 지문 매칭 쌍으로부터 $a = 0.967$, $b = 203.897$ 를 구할 수 있다. 매칭 쌍의 분포는 각 S_{st} 와 S_{ts} 로 매칭을 수행하여 나온 매칭점수이다.

템플릿 지문을 기준으로 입력지문을 매칭 시키는 S_{st} 경우나 이와 반대로 입력지문을 기준으로 템플릿 지문을 매칭 시키는 S_{ts} 는 모두 입력지문을 어떻게 입력받았느냐에 따라 매칭점수는 변하게 되며 이는 곧 매칭분포의 영향을 미친다. 이렇게 변형될 수 있다는 점으로 인해 직선형태의 회귀된 모형을 통하여 예측구간을 설정하여 주었다. 이 때 발생한 상한성분을 통하여 적응적 임계치 Th' 을 자동적으로 찾을 수 있도록 하였다.

예측 구간의 추론은 회귀직선의 불확실성과 오차항 내에 존재하는 불확실성에 의해 존재하게 된다. 입력지문의 매칭 점수는 실험에 의한 데이터로 인해 나오는 것이며 이는 어떠한 값이 나올지는 알 수 없는 불확실성을 가진다.

예를 들어 신뢰된 지문매칭 쌍 5247개의 데이터 분포 중 S_{st} 의 x축의 값이 4398점에 해당되는 점수에 대하여 예측구간은 다음과 같이 구할 수 있다.

$$\begin{aligned} x^* &= 4398 \text{ 이고, 예측구간 } 95\% \text{ 일 때} \\ y &= 0.967x + 203.897 \text{ 의 직선방정식에 의하여,} \\ y|_{4398} &\in (0.967 \times 4398 + 203.897 \pm 1383.8) \\ &= (3073, 5840.6) \end{aligned}$$

즉, S_{st} 의 매칭점수 4398점에서 입력지문에 의해서 매칭 점수가 변하거나 혹은 매칭과정 중 매칭점수가 바뀔 경우 95%의 신뢰수준으로 S_{ts} 의 매칭점수는 3073점과 5840.6점 사이의 값이 나오게 된다는 것을 알 수 있다.

3. 적응(Adaptive) 임계치 Th' 의 설정

본 논문에서는 효율적인 1:N 매칭을 수행하기 위해서 대규모의 지문 매칭 쌍 데이터를 분석하여 새로운 임계치 Th' 를 자동적으로 설정해줌으로써 동일시간 내에 많은 매칭을 시키고, 오거부율을 최소화하는데 목적을 두고 있다.

지문인식에서 매칭 성공이 되었다는 것을 판단하는 기준은 임계치 Th 의 값이다. 임계치 Th 의 설정은 매칭알고리즘에 따라 차이를 보이기 때문에 적응적(adaptive) 임계치 Th' 은 다양한 임계치 Th 값에 따라 자동적 설정이 요구된다.

이를 해결하기 위하여 3.2절에서 설명하였던 Fitting기법과 예측구간으로 활용하여 적응 임계치 Th' 를 설정하여 주었다. 예측구간으로 발생한 상한의 성분에 대하여 점 성분을 선 성분으로 바꿔줌으로써 발생하는 수식을 통하여 기

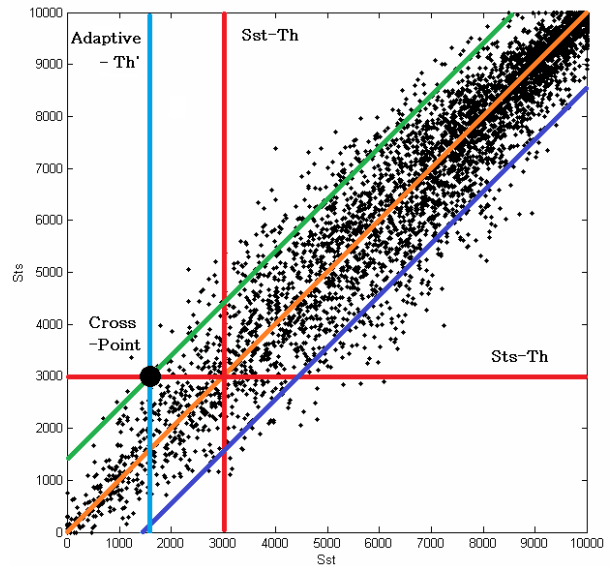


그림 11. 상한예측구간을 이용한 임계치 Th' 설정.

Fig. 11. Th' decision using the upper prediction intervals.

준의 양방향 매칭 수행할 임계치 Th 값과 만나는 점을 적응적 임계치 Th' 로 설정을 한다.

그림 11은 $Th = 3000$ 일 경우를 예로 보여주고 있다. 이 경우 S_{st} 를 기준으로 매칭을 수행할 경우, $S_{st}-Th$ 를 기준으로 단방향 매칭을 수행 한 뒤, 부분적 양방향 매칭을 수행할 $S_{ts}-Th$ 와 상한 경계와 만나는 점(x, y)에서 x성분의 값을 적응적 임계치 Th' 으로 설정을 하여 준다. 적응적 임계치 Th' 을 자동적으로 찾는 것에 대한 영향을 미치는 부분은 신뢰수준의 설정이다. 예측구간 설정 시 신뢰수준의 값은 fitting된 직선과 상한, 하한성분의 떨어질 범위를 설정하는 부분이다. 이는 즉, 신뢰수준의 값이 낮아질수록 상한, 하한성분의 값들은 직선과 가까워진다는 사실을 알 수 있다. 신뢰수준의 값은 예측범위를 설정하는 부분으로써 정성적 부분에 의해 구해진 값이 아니기 때문에 본 논문에서는 이 신뢰수준의 값을 변경하여 줌으로써 데이터의 분석 및 통계를 통하여 많은 매칭 쌍의 분포에 대한 정량적 조사를 실시하여 최적의 값을 찾는다.

IV. 실험 및 결과

본 논문은 신뢰된 니트젠 지문 매칭 쌍 5247개의 데이터로 실험을 실시하였으며, National Instruments LabVIEW 8.5.1 Software를 통한 통계, 분석 및 알고리즘을 구현하였다[3].

1. 단방향 매칭 및 양방향 매칭

표 1은 매칭쌍은 5247개, Th 는 3000, S_{st} 기준으로 임계치 Th 를 기준으로 단매칭한 결과와 S_{st} 임계치 Th 매칭 후 S_{ts} 임계치 Th 매칭 작업 수행을 수행한 양매칭한 결과를 나타낸다.

단방향 매칭을 수행 시 매칭횟수는 줄어들어 1:N 매칭 수행 시 시간적 이득을 안겨줄 수 있지만, 매칭이 성공될 경우의 매칭 쌍들 까지도 인식하지 못하면서 오거부율

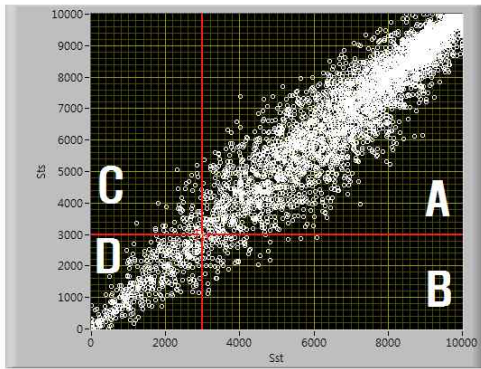


그림 12. 매칭 분포.
Fig. 12. Matching distribution.

표 1. 단방향 매칭과 양방향 매칭의 비교.

Table 1. Comparative table for unidirectional and bidirectional matching.

	단방향 매칭	양방향 매칭
총 매칭횟수	5247번	10494번
평균 매칭횟수	1회	2회
FRR	10.1%	8.65%

FRR이 증가한다는 것을 알 수 있다. 양방향 매칭을 수행하였을 경우, 임계치 Th 이하의 모든 매칭 쌍을 찾음으로써 오거부를 FRR을 최적의 값으로 찾지만, 매칭횟수가 단방향 매칭의 두 배로 증가하여 단위시간 내의 매칭 시간이 증가함을 알 수 있다.

2. 보조 임계치 Th' 에 따른 단방향, 양방향 매칭 병행시스템
단방향, 양방향 매칭 수행에 따른 문제점들을 해결하고자 논문[1]에서 제안한 임계치 Th' 을 설정하여 준다. 매칭 알고리즘에 따른 임계값 Th 의 다양함을 고려하여 본 실험에서는 통상 사용되고 있는 범위 1500, 2500, 3000을 기준으로 실험을 수행하였다. 이에 따라 새로운 보조 임계치 Th' 은 0, 500, 1000, 1500, 2000, 2500으로 설정하여 줌으로써 Th' 에 따른 이득을 파악하였다.

그림 13~15는 임계치 변화에 따른 $Th=0$ 일 때의 매칭 수행 및 분석실험 결과를 도시하였고, 표 2~7은 임계치 Th 에 따른 각 영역의 매칭 쌍의 개수와 결과를 나타낸다.

표 3~5에서 확인할 수 있듯이 Th' 이 Th 에 가까이 할수록 평균매칭 횟수는 줄어들지만 오거부를 FRR이 커지고, 반대로 Th' 를 '0'으로 가까이 할수록 오거부를 FRR은 줄어들지

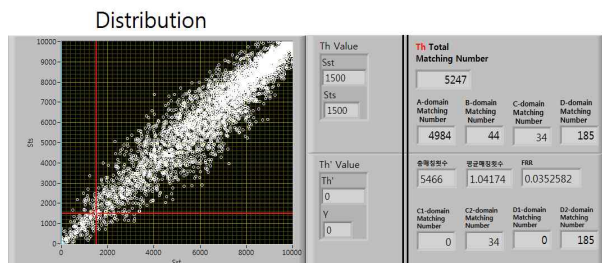


그림 13. $Th=1500, Th'=0$ 일 때의 매칭 수행 및 분석.
Fig. 13. Matching distribution with $Th=1500, Th'=0$.

표 2. 각 Th' 값에 따른 새로운 영역 매칭쌍 개수.

Table 2. New domain matching No. for various Th' .

매칭 쌍의 개수	$Th'=0$	$Th'=500$	$Th'=1000$
C1-domain	0	1	7
C2-domain	34	33	27
D1-domain	0	68	133
D2-domain	185	117	52

표 3. $Th=1500, Th'=0, 500, 1000$ 에 따른 결과 비교.

Table 3. Comparative analysis for $Th=1500, Th'=0, 500, 1000$.

결과	$Th'=0$	$Th'=500$	$Th'=1000$
총 매칭횟수	5466	5397	5326
평균 매칭횟수	1.04174회	1.02859	1.01506
FRR(%)	3.52582	3.54	3.66923

표 4. $Th=2500, Th'$ 값에 따른 수행결과.

Table 4. Results for various $Th'(Th=2500)$.

결과	$Th'=0$	500	1000	1500	2000
총 매칭횟수	5655	5586	5515	5436	5348
평균매칭횟수	1.07776	1.06461	1.05108	1.03602	1.01925
FRR(%)	6.67	6.67048	6.68954	6.72765	7.08977

표 5. $Th=3000$ 일 때, Th' 에 따른 변화.

Table 5. Results for various $Th'(Th=3000)$.

	$Th'=0$	500	1000	1500	2000	2500
총매칭횟수	5777	5708	5637	5558	5470	5369
평균매칭횟수	1.101	1.088	1.074	1.059	1.043	1.023
FRR	8.653	8.653	8.653	8.672	8.786	9.167

만 평균매칭횟수가 늘어남으로써 평균매칭시간이 늘어남을 확인할 수 있다.

다음은 위의 표 3~5에서 Th' 변화에 따른 평균매칭시간과 오거부를 FRR의 변이정도를 나타낸 그림이다.

그림 14에서 하얀색라인은 $Th=1500$, 빨간색라인은 $Th=2000$, 녹색 라인은 $Th=2500$ 을 나타내며, 평균 매칭 횟수는 각 변화에 따라 일정비율로 반비례의 관계를 유지하는 것을 확인할 수 있다. 그림 15는 Th' 변화에 따른 오거

Th별 Th' 따른 평균매칭횟수 비교

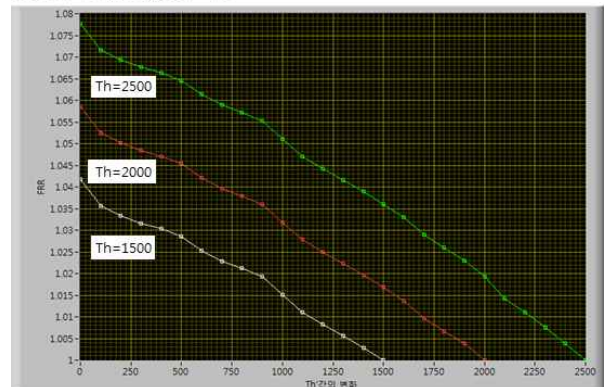


그림 14. Th' 변화에 따른 평균매칭시간.

Fig. 14. Average matching time based on Th' variation.

Th별 Th' 따른 FRR의 비교

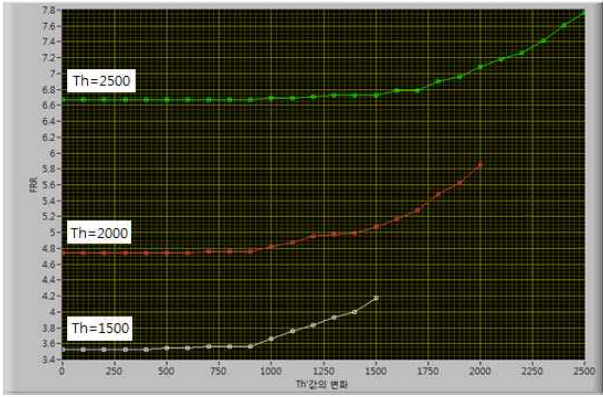


그림 15. Th' 변화에 따른 오거부율 FRR.

Fig. 15. FRR based on Th' variation.

부율 FRR의 변이를 나타낸다. 그림에서 하얀색라인은 Th=1500, 빨간색라인은 Th=2000, 녹색 라인은 Th=2500을 나타낸다. 평균매칭횟수에 있어서는 Trade off를 보이지만 이와 달리 FRR의 경우, 각 Th에 따라 Th'의 설정에 의해 급격히 변화를 이루는 값을 찾을 수 있다. 그림 15에서 Th=1500일 경우, Th'값이 900에서 1000이 되는 시점에서 변화가 일어나며, Th=2000일 경우, Th'값이 900에서 1000이 되는 시점에서 변화가 일어나기 시작했고, Th=2500의 경우, Th'값이 1500에서 1600이 되는 시점에 값의 변화가 일어나는 것을 확인 할 수 있다.

3. Fitting 기법에 의한 임계치 Th'의 설정

4.2의 실험에서는 Th'의 값을 임의적으로 Th'값을 가변시켜주면서 비교 및 분석에 사용되었다. 본 절에서는 Th'을 fitting 기법으로 도입하여 Th'값을 찾으려 실험을 하였다.

앞에서 언급한 것과 같이 지문 데이터가 가우스 분포형태를 갖기 때문에 선형회귀분석을 통하여 fitting시키며, 이후 예측구간을 선형적 직선으로 바꿔주어 임계치 Th와 만나는 지점을 찾아 주었다. 그림 16은 예측구간의 신뢰수준 값을 0.95로 설정하였을 경우를 나타낸다.

그림 17은 예측 구간을 선형 직선으로 바꿔준 그래프이며, 이에 발생한 $S_{ts} - Th$ 와 만나는 점을 찾아 주었다. 이때 찾아진 점을 적응적 임계치 Th'로 정의하였다. 예측구간

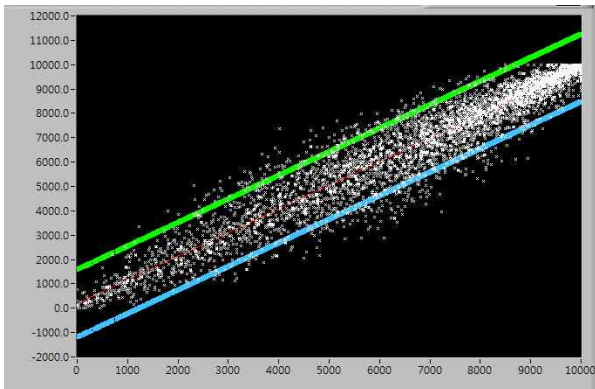


그림 16. Fitting 작업과 예측구간 설정.

Fig. 16. Fitting plotting and estimation setting.

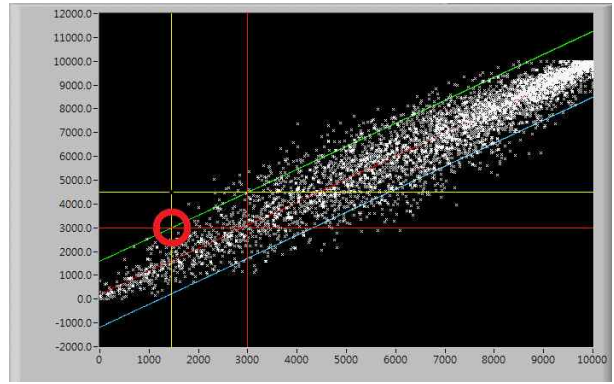


그림 17. 예측구간의 선형화와 임계치의 만나는 점(신뢰수준 0.95).

Fig. 17. Junction of threshold and linear regression.

의 범위는 직선과 떨어져 있을 신뢰수준의 값을 나타내며, 이 값은 예측적 신뢰범위를 설정해야하므로 정성적인 값을 찾는 것은 어려움이 따른다. 하지만, 어떠한 예측범위를 설정하더라도 적응적 Th'이 자동적으로 설정되어 기존의 단방향 매칭이나 양방향 매칭보다 양호한 성능을 보인다.

다음은 신뢰수준의 값에 따른 시스템의 분석을 위해 임계치 Th값을 3000으로 고정시키고 신뢰수준의 값을 변경시켜 주었다.

신뢰수준 값의 변경에 따른 실험을 토대로 선택적 적응 임계치 Th'를 선택하여 매칭에 사용된 임계치 Th에 적용하였으며, 본 논문에서는 신뢰수준을 0.95 범위로 예측하여 적응 임계치 Th'을 설정하여 주었다.

표 7은 각 임계치와 신뢰수준 값을 0.95로 했을 경우 설정된 적응적(adaptive) Th'값의 변화에 따른 총 매칭 횟수, 평균 매칭횟수, FRR을 나타낸다. 본 실험에서 Th=3000과 신뢰수준의 값을 0.95로 설정을 한 이유는 Th'의 값은 자동적으로 어떠한 Th값에 의해서도 설정이 가능해야한다는 목적을 가지고 있으며, Th'의 값은 신뢰수준의 값에 의해 변화를 갖는다는 것을 지금까지의 자료를 통하여 알 수 있다.

Th값을 3000으로 기준으로 하여 Th'을 설정하여 주었을

표 6. 신뢰 수준 0.95, 0.90, 0.80일 때의 실험결과 비교.

Table 6. Results for confidence interval 0.95, 0.90, 0.80.

	Th'=1459.62 (0.95)	Th'=1689.92 (0.90)	Th'=1955.38 (0.80)
총 매칭횟수	5314	5310	5302
평균 매칭횟수	1.01277	1.01201	1.01048
FRR (%)	8.82409	8.90032	9.05279

표 7. 각 임계치 Th에 대한 결과.

Table 7. Results for various Th.

	Th = 1500	Th = 2000	Th = 2500	Th = 3000
Adaptive Th'	0	425.512	942.568	1459.62
총 매칭횟수	5466	5492	5529	5563
평균 매칭횟수	1.04174	1.04669	1.05374	1.06022
FRR	3.52582	4.74557	6.67048	8.67162

표 8. Th=3000일 때 가변 Th'에 따른 비교.

Table 8. Comparison for various Th' (Th = 3000).

	Adaptive Th'	가변에 의한 Th'의 연구			
새로운 임계치	1459.62	1000	1500	2000	2500
총매칭횟수	5563	5637	5558	5470	5369
평균매칭횟수	1.06022	1.07433	1.05927	1.0425	1.02325
FRR	8.67162	8.65256	8.67162	8.78597	9.16714

때, Th'값은 1500일 때 효율적이라는 것을 그림 14~15에서 확인 할 수 있다. 이를 설정해주기 위해서는 신뢰수준의 값은 0.95로 설정을 해 줘야 1500에 근접한 값을 찾을 수 있다. 이때의 신뢰 수준의 값 0.95에 의한 값에 의해 Th=1500으로 설정하였을 경우, Th'은 0으로 설정하게 된다. 이 때 발생한 값에 의해 비록 Th'의 값은 현저히 줄어 들었지만, 그림 15에서 비교하여 보면, 900의 이하의 값은 FRR의 변동이 거의 없는 상태로써 효율적인 매칭을 수행하는데 이바지하게 된다. 이러한 이유로써, 최대의 임계치 Th를 3000으로 설정을 하여 매칭을 수행하며, 이 때의 최적의 Th'값을 설정하여 주기 위해 0.95의 신뢰수준을 기준으로 본 논문을 구성하였다.

단방향 매칭보다 효율적인 측면에 있어 기존의 단방향 매칭만을 수행하였을 시, 단위시간당 매칭 횟수 150회를 수행하며, 본 실험에서는 141.48회를 수행하였다. 단방향 매칭에 비해 단위시간당 매칭 횟수는 줄어들지만, 오거부율 FRR이 약 1.43 % 감소하였다는 이점을 얻는다. 즉, 약 6 %의 매칭 시간을 소요시킴으로써 오거부율을 최소화하는 효율적인 매칭을 수행할 수 있다.

또한 양방향 매칭보다 효율적인 측면에 있어 기존의 양방향 매칭만을 수행하였을 시, 단위시간당 매칭 횟수는 75회만을 수행하였지만, 본 논문에서 제안한 방법으로 인하여 같은 단위시간 내에 141.48회의 매칭을 수행할 수 있다. 이는 즉 약 0.8초 내에 양방향 매칭의 수행결과 같은 매칭회수를 가질 수 있어 빠른 매칭을 수행할 수 있다.

V. 결론

본 연구에서는 지문인식에 있어서 매칭시간을 줄일 수 있는 단방향, 양방향 매칭 병행시스템을 비교, 검토하였다. 본 병행시스템은 단방향 매칭을 기본으로 사용하되 매칭 성공 여부를 판단하는 임계치 Th'에 새로운 적응적(adaptive) 임계치 Th'을 추가시켜 줌으로써 Th와 Th' 사이에 존재하는 경우에만 양방향매칭을

수행을 시켜주는 방식으로 수행하였다. 이때, Th'값을 설정하여 주기 위해 분포 특성을 통계적으로 파악하여 선형 회귀를 통하여 직선으로 Fitting 시켜 준 뒤, 예측구간을 설정하고 신뢰구간 설정을 통하여 임계치의 값에 따라 자동적으로 새로운 적응적(adaptive) 임계치 Th'값을 찾아 주었다.

본 제안된 방법으로 인해 다음과 같은 기대효과를 가져올 수 있다.

첫째, 기존의 양방향 매칭의 매칭시간이 오래 걸린다는 점을 해결 할 수 있다.

둘째, 기존의 단방향 매칭과의 비교에 있어서도 양방향 매칭과 유사한 신뢰성을 가질 수 있다.

셋째, 임의로 가변시켜 주었던 임계치 Th'을 통계적 분석에 기초한 임계치의 값 Th에 따라 자동적으로 새로운 임계치 Th'을 찾을 수 있도록 설정 할 수 있다.

제안된 방법에 대한 효율성을 검증하기 위하여 대규모 지문 특징 데이터 베이스의 양방향 매칭 점수 분포를 구하고 이를 단방향, 양방향 매칭의 장점을 가지는 병행시스템을 적용해 보았으며, 실험을 통하여 단방향 매칭과 비슷한 단위시간당 매칭 횟수와 양방향 매칭 보다 약 1.88배의 매칭 횟수로 기존의 양방향 매칭보다 빠른 매칭을 수행하였으며, 오거부율 FRR의 경우 단방향 매칭보다 약 1.43 % 감소를 가져왔으며, 신뢰성이 높은 양방향매칭과는 유사한 오거부율 FRR을 가져왔다는 것을 실험을 통하여 유용성을 검증하였다.

향후 연구로는 다양한 지문 매칭 알고리즘에 따른 양방향 매칭 점수의 분포 특성을 파악하고 본 논문에서 제안한 단방향, 양방향 매칭 병행시스템 방법에 적용하는 것이 필요하며, 자동으로 적응적(adaptive) 임계치 Th'에 대한 최적의 값을 찾기 위해 신뢰구간의 값에 따른 매칭 쌍의 분포를 통계적으로 분석한 뒤, 최적의 Th'을 찾는 것이 필요할 것이다.

참고문헌

- [1] D. P. Mital and E. K. Teoh, "An automated matching technique for fingerprint identification," *Proc. of IEEE Industrial Electronics, Control, and Instrumentation*, Taipei, Taiwan, vol. 2, pp. 806-811, Aug. 1996.
- [2] A. Jain, L. Hong, and R. Bolle, "On-line fingerprint verification," *IEEE Trans. on Pattern Recognition and Machine Intelligence*, vol. 19, no. 4, pp. 302-313, 1997.
- [3] S. W. Jung, "Efficient 1:N matching scheme for fingerprint identification," *Journal of IEEK (in Korean)*, vol. 45, no. 5, pp. 173-179, 2008.
- [4] L. C. Jain. U. Halici, I. Hayashi, S. B. Lee, and S. Tsutsui, *Intelligent Biometric Techniques in Fingerprint and Face Recognition*, CRC Press, 1999.
- [5] A. K. Jain, L. Hong, and Y. Wan, "Fingerprint image enhancement: algorithm and performance evaluation," *IEEE Transaction on Pattern Analysis and Matching Intelligence*, vol. 20, no. 8, pp. 777-789, 1998.
- [6] D. Maltoni, D. Maio, A. K. Jain, and S. Prabhakar, *Handbook of Fingerprint Recognition*, 2nd Ed., Springer, 2009.
- [7] Harvey Motulsky & Arthur Christopoulos, *Fitting Models to Biological Data Using Linear and Nonlinear Regression*, Oxford University Press, 2003.
- [8] J. H. Chang and K. C. Fan, "A new model for fingerprint classification by ridge distribution sequences," *Pattern Recognition*, vol. 35, no. 6, pp. 1209-1223, Jun. 2002.

**김 경 민**

1988년 고려대 전기공학과 졸업. 1991년, 1996년 동 대학원 석사 박사. 2003, 2010년 캐나다 CENPARMI 방문교수. 현재 전남대 전기전자통신컴퓨터공학부 교수. 관심분야는 컴퓨터비전, 생체인식.

**박 중 조**

1981년 고려대 전기공학과 졸업. 1983년, 1995년 동 대학원 석사 박사. 2006년 캐나다 CENPARMI 방문교수. 현재 경상대 제어계측공학과 교수. 관심분야는 컴퓨터비전, 생체인식.

**이 범**

1981년 고려대 전기공학과 졸업. 1989년, 1995년 동 대학원 석사, 박사. 현재 전남대 공학대 전기전자통신컴퓨터공학부 교수.

**고 영 진**

2008년 전남대 전기및반도체공학과 졸업. 2011년, 동 대학원 석사. 2011년~ 현재 박사과정 관심분야는 컴퓨터비전, 로봇응용.

**정 순 원**

1990년 고려대 전기공학과 졸업, 1992년, 1997년 동 대학원 석사, 박사. 1999년~2009년 (주)니트젠 기술연구소 연구소장. 2010년~현재 Ria Soft 대표이사 관심분야는 생체인식.