

센서 어레이의 신호패턴 분류를 위한 각도 변이 기반 상태 천이 모델링 기법

김아람¹ · 이승재² · 김창화^{2†} · 김상경² · 박수현³

Angle Difference Based State Transition Modeling Technique for the Classification of Signal Pattern from the Sensor Array

Aram Kim · Seungjae Lee · Changhwa Kim · Sangkyung Kim · Soohyun Park

ABSTRACT

We propose a method to use a state transition model so that the sensing object can be distinguished through classification of signal patterns sensed by a sensor array. Focusing on the design of the model that is able to distinguish the sensed object more exactly, we present an idea in which the modeling elements, 'states' and 'transitions' are defined as each same-sized angle intervals into which the angle interval $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$ is divided and the angle differences between adjacent signal values on sampling signal value sequence sensed from the sensor array in the uniform time interval, respectively. In addition we show the usefulness of our model through experiments.

Key words : Sensor array, signal pattern classification, state transition model, ADSTM

요약

센서 어레이로부터 감지된 신호 패턴을 분류함으로써 감지 대상체를 구별하기 위해 본 연구에서는 상태 천이 모델을 이용하는 방법을 제안하였다. 센서 어레이의 신호 데이터를 패턴 모양의 특성을 나타낼 수 있는 상태 천이 모델로 변환하여 감지 대상체의 구별이 보다 정확하게 이루어 질 수 있도록 모델을 설계하는데 초점을 두면서, 모델링 요소인 '상태'는 각도 구간 $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$ 을 n 개의 일정한 크기의 구간으로 나누어 각 구간을 하나의 상태로 정의하고, '천이' 관계는 일정한 시간 간격으로 샘플링된 신호 데이터 간의 각도 변화로 각각 정의하여 각도변이 기반 상태천이 모델링을 고안하였으며 모델의 유효성을 실험을 통하여 검증하였다.

주요어 : 센서어레이, 시그널 패턴 분류, 상태천이모델, 각도변이기반상태천이모델링기법

1. 서론

센서는 일반적으로 "측정 대상물로부터 정보를 검지

또는 측정하여 그 측정량을 인식 가능한 유용한 신호로 변환하는 장치", "외계의 정보를 감지하여 신호처리하기 쉬운 전기나 빛의 신호로 변환하는 기능을 지닌 장치" 등으로 정의된다(김상진, 1998). 바이오 센서의 개발은 감지 대상체의 상태를 모니터링·분석 및 그 분석 결과에 따라 자동으로 제어하거나 의사결정을 행하는 시스템 개발을 위한 기초가 된다.

하나의 센서만 가지고 여러 대상체들 중 하나의 대상체를 식별하거나 혹은 감지 대상체의 상태를 식별할 수 없는 경우가 있다. 예를 들어 향수의 냄새는 여러 종류의 기체들로 구성되는데 향수의 냄새를 가지고 향수의 종류를 식별해야 할 경우에는 하나의 기체 감지 센서만으로는

* 본 연구는 정보통신부 및 정보통신연구진흥원의 대학 IT연구센터 지원산업의 연구결과로 수행되었음(IITA-2005-C1090-0501-0010)

2006년 7월 19일 접수, 2006년 8월 19일 채택

¹⁾ 강릉대학교 교육대학원 컴퓨터교육전공

²⁾ 강릉대학교 컴퓨터공학과

³⁾ 국민대학교 비즈니스IT학부

주 저 자 : 김아람

교신저자 : 김창화

E-mail; kch@kangnung.ac.kr

불충분하며 여러 종류의 기체들을 각각 감지하는 센서들이 필요로 된다(Jim Lipman, 1998; 홍형기 등, 1995; 이준섭, 2002).

센서 어레이는 하나의 센서 보드에 여러 개의 센서들을 장착하여 각 센서들로부터 감지된 데이터를 기반으로 여러 대상체들 중에서 하나의 대상체를 식별하거나 혹은 감지 대상체의 상태를 구별한다.

센서로부터 감지되는 신호 패턴은 감지 대상체의 현 상태를 나타내지만, 감지 대상체의 현 상태에 영향을 미치지 않는 환경적 요인들(예를 들어, 자장의 영향, 빛의 강도 센서 자체의 온도 등)에 의해 감지 동일한 상태에 대해 노이즈 등이 포함됨으로써 다른 패턴을 보이게 된다. 결국, 감지 대상체의 동일한 상태에 대해 여러 가지 환경 요인들 때문에 수 많은 여러 패턴들이 존재하게 되고, 이로 인하여 센서가 방출하는 신호 패턴을 가지고 감지 대상체의 현 상태를 분류하거나 분석하는데 대한 매우 어려운 문제점을 가지게 된다.

한편, 센서 어레이를 사용할 때에는 하나의 특정 센서로부터의 신호 패턴만을 중심으로 대상체 혹은 대상체의 상태를 구별하는 것이 아니라 각 센서들로부터 감지된 신호 패턴들의 총체적 패턴을 고려하여 대상체 혹은 대상체의 상태를 구별해야 한다.

본 논문은 센서 어레이의 각 센서들로부터 감지된 신호 패턴들을 기반으로 대상체 혹은 대상체의 상태를 구별하기 위한 일종의 상태 천이 모델링(State Transition Modeling) 기법을 제안한다. 센서 어레이 기반의 인식에 적용되는 기존 신호 패턴 인식 방법들은 신호 패턴의 특성을 잘 반영하지 못하거나 혹은 학습시간이 길어지는 경우가 많았는데 본 논문에서 제안한 방식은 신호 패턴의 특성을 잘 반영하여 학습뿐만 아니라 인식에서도 상대적으로 짧은 시간에 수행 가능하다.

본 논문의 구성은 다음과 같다. 2 장에서는 상태천이 모델과 유사한 모델을 기반으로 하는 모델링 방법들에 대하여 소개한다. 3 장에서는 본 논문에서 제안하고자 하는 기법인 각도 변이 기반 상태 천이 모델링 기법을 소개하고, 4 장에서는 가스를 기반으로 제안한 방법을 적용한 예를 제시하며, 5 장에서는 센서 어레이 기반의 신호 패턴 분류를 위한 프로세스 프레임워크를 제시한다. 6 장에서는 제안한 모델의 유효성을 알아보기 위한 실험 방법과 실험 결과를 제시하고 7 장에서 결론을 맺는다.

2. 관련 연구

센서 에레이 신호 기반의 학습과 인식에서 ANN(Artificial Neural Network) 기법이 많이 사용되어 왔다(강경아, 1997). ANN 기법은 입력층, 은닉층, 출력층 등으로 구분하여 supervised learning 시 학습 데이터의 출력 결과와 의도된 결과와의 오차를 보정하기 위해 하위 층으로 역전파(back-propagation)하는 과정을 반복하는 과정을 통하여 학습을 행한다. ANN 기법은 인식을 제고하기 위해 학습을 위한 많은 데이터 세트의 준비와 많은 학습시간 소요 등의 문제점들이 대두되어 왔다.

본 논문에서는 짧은 학습시간에 비해 정확한 인식률을 갖는 기법으로 상태 천이 모델링(State Transition Modeling) 기법을 적용하기 때문에 이 장에서는 본 논문의 아이디어의 기반이 되는 상태 천이 모델 기반의 학습과 인식 기법에 대해 소개하고자 한다.

• Pseudo 2D 모델링(Oscar E. Agazzi 등, 1993)

문자를 2차원 공간에 표현된 데이터로서의 관점을 가지고 일련의 전처리 과정을 거쳐 구분된 문자 모델을 Pseudo 2D HMM이라고 한다. 이 모델은 각 문자를 구분하고 서로 다른 크기의 문자를 정규화(normalize) 과정을 거쳐 일정한 크기의 픽셀맵(pixel map)에 표현된 문자를 Hidden Markov Chain으로 모델링 한 것으로 모델링에 대한 예가 그림 1에 나타나 있다.

이 모델링 방법은 2차원 맵에 표현된 문자 그래픽에서 흑(black)과 백(white)의 두 상태로 구분하고 x 축의 진행 방향에 따른 흑에서 백 상태로의 전환 혹은 백에서 흑 상태로의 전환을 천이 관계로 하여 상태 천이 모델을 만든다. 맵을 그림 1에서 보는 바와 같이 y축 방향으로 진행하면서 x축 방향으로 흑과 백의 상태 천이를 모델링 한다.

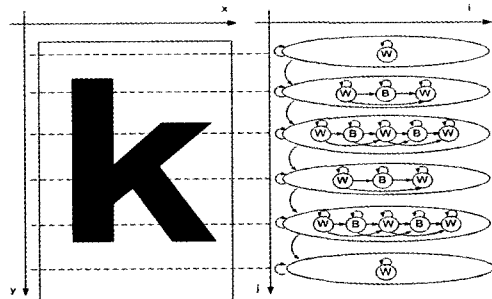


그림 2. 픽셀맵(pixel map)과 Pseudo 2D Hidden Markov 모델링

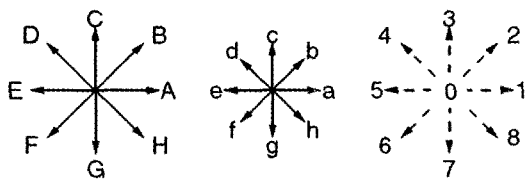


그림 3. 획의 방향을 이용한 substroke Hidden Markov 모델링

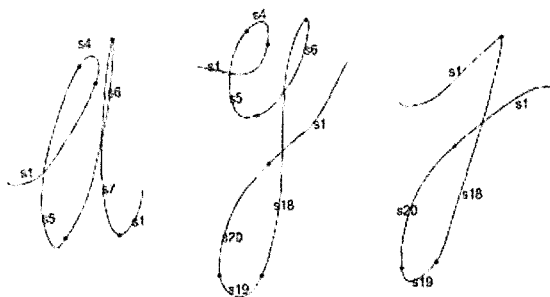


그림 4. Nebulous stroke 모델의 획 분할

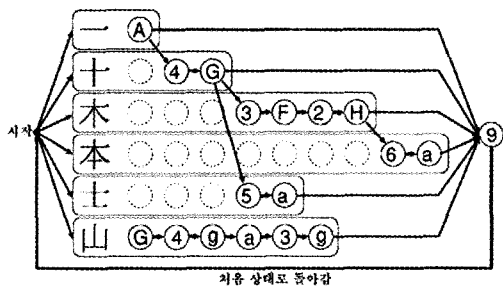


그림 3. Substroke 최적 문자 시퀀스 탐색 예시

- 서브스트로크(substroke) 모델링 (Hiroshi SHIMODAIRA 등, 2003)

서브스트로크 모델링은 일본어 간지를 모델링 할 때 사용한 방법으로 간지의 경우 계층적 언어구조를 가지고 있고, 같은 획이 여러 문자에서 공동으로 사용되는 경우가 많다. 따라서 다른 문자보다 모델 유닛으로 선택되는 획의 수가 상대적으로 적다고 할 수 있다. 이러한 적은 획수를 바탕으로 획의 방향과 길이, 펜을 들었을 때와 인식 판에 대었을 때의 펜의 이동 방향 등 3요소로 모델링을 할 수 있다. 그림 2에서와 같이 여덟 방향의 긴 획(A-H)과 여덟 방향의 짧은 획(a-h), 8방향의 펜의 이동(1-8), 펜의 접지상태(0) 까지 총 25개의 서브스트로크로 구성되어 있다. 그림 3은 간지의 각 문자 HMM을 나타내는 상태열의 수평방향 길들로 구성되어 있다. 각 열의 마지막 상태는 다시 문자열이 시작하는 처음상태로 돌아간다.

- Nebulous stroke 모델링(Jianving Hu 등, 1996)

필기체 문자의 획을 여러 단위로 잘게 나누어, 각 부분 획을 기준으로 모델링 한 경우도 있다. 그림 4에서와 같이 하나의 문자는 여러 개의 부분 획으로 구성되어 있다.

따라서 한 문자는 부분 획의 연속으로 나타낼 수 있다. 이를 HMM 모델링 한 것이 Nebulous Stroke 모델이다. 각 문자는 여섯 개, 여덟 개, 다섯 개의 획을 가진 'a', 'g', 'j'를 가리킨다. 'a'와 'g'는 s1, s4, s5, s6의 첫 4개의 부분 획을, 'g'와 'h'는 s18, s19, s20, s1의 마지막 4개의 부분 획을 공유한다. s1 부분 획은 3개 모두에서 사용된다. 문자를 여러 부분 획의 연속으로 생각하여 문자를 모델링하여 완성한다.

3. 각도 변이 기반 상태 천이 모델링

이 장에서는 센서 어레이를 이용하여 측정된 가스 패턴의 신호 그래프를 양자화 하여 각 요소 사이의 각도 천이 관계를 모델링에 반영한 각도 변이 기반 상태천이 모델링 (ADSTM : Angle Difference based State Transition Modeling) 기법을 소개한다. 감지 대상체의 구성 성분(혹은 상태)을 확인해보기 위해서 기존에 측정하여 보관해 둔 데이터들과 새로 측정된 데이터 사이의 성분(혹은 상태) 유사도를 구하는 방법을 사용한다. 유사도 계산에는 다양한 방법들이 사용되지만 본 연구에서는 상태 천이 모델링 방법에서 천이 행렬 간의 공간적 유사

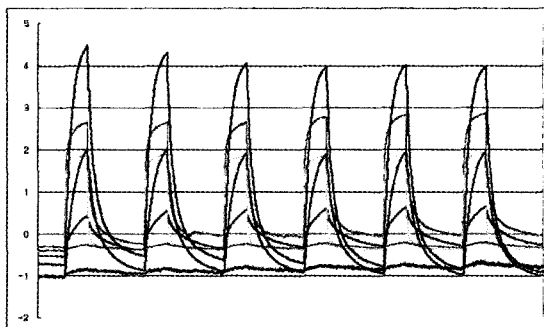


그림 6. 센서 어레이 신호의 형태

도를 측정하여 계산한다.

ADSTM 기법을 적용하기 위해서는 먼저 데이터의 특성을 살펴볼 필요가 있다. 본 연구에서 센서 데이터의 입력으로 받는 자료의 형태는 센서 어레이 여러 채널의 측정값들로 다차원 데이터이다. 그림 5는 동일한 가스를 인식할 때까지 반복하여 측정한 것을 그래프로 나타낸 것으로, 시간 경과(t)에 따른 각 센서 채널 측정값들의 변화(volts)를 좌표 평면상에 표현하였다. 이 그림에서는 채널이 8개로 2차원의 곡선 데이터가 8개 겹쳐진 다차원 데이터의 형태를 가진다.

3.1 의미 구간 추출

센서 어레이 신호들을 살펴보면 가스를 한번 측정할 때마다 증가와 감소의 과정을 반복하는 것을 볼 수 있다. 그림 6은 그림 5와 다른 가스를 측정하여 그래프로 표시한 것이다. 두 그림을 살펴보다도 센서 어레이 측정값들은 증가와 감소를 반복하는 것을 알 수 있다.

증가부분은 센서가 가스를 인지하여 안정된 결과값을 얻

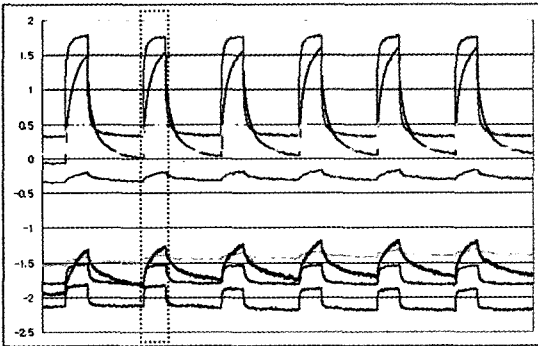


그림 7. 센서 어레이 신호 패턴에서의 의미 구간 선택

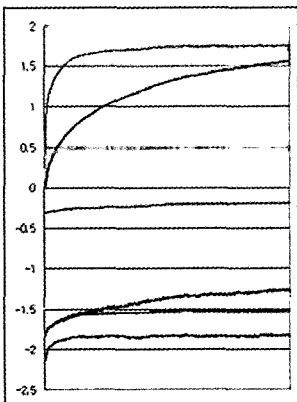


그림 8. 추출된 인식 구간

어가는 과정이고, 감소 부분은 센서 측정 과정이 종료되어 다시 가스가 없는 일반 환경의 센서 상태로 값이 회귀되는 과정이다.

따라서 센서의 신호가 의미를 갖는 구간은 가스가 발생되면 센서에 의해 인지되는 구간인 증가 과정이 표시되는 구간이라고 할 수 있다. 센싱값이 증가하기 시작하는 부분부터 최고값에 이를 때까지의 구간을 추출하는 것이다. 이러한 과정으로 그림 6에서 선택된 구간은 그림 7과 같이 표현할 수 있다.

3.2 추출된 구간의 선형 배치

본 논문에서 이용한 모델링 방법을 위해서는 추출된 인식구간의 다차원 데이터를 선형 데이터로 변환하는 과정이 필요하다. 이 변환과정에서는 데이터 변환 뿐 만 아니라 모델의 신뢰도를 높이기 위한 노이즈 제거 작업도 같이 수행한다.

그림 7에서 추출된 곡선 부분을 선형 데이터로 변환하기 위해 그림 8에서와 같이 선형으로 배치하여 다차원 데이터를 1차원 선형 데이터로 다루기 쉬운 데이터의 형태로 변환한다. 그림 8에서는 8개의 채널을 일정한 순서로 배열하였다. 각 채널은 색깔로 구분하였다.

3.3 양자화

다음 단계에서는 선형으로 변환된 데이터를 상태 천이 모델을 적용하기 위해 각 상태(state)로 나누고 각 상태 사이의 천이 관계를 얻어내출 시간 간격을 동일한 거리로 넓기 위해 데이터를 양자화 한다.

그림 8의 데이터 곡선은 매우 짧은 시간 간격으로 측정된 데이터를 점으로 나열한 결과, 곡선의 형태를 가지고 있다. 양자화 방법은 요소와 요소 사이의 추계 함으로써 곡선의 그래프 모양을 꺾은선의 모양으로 변환하는 것이

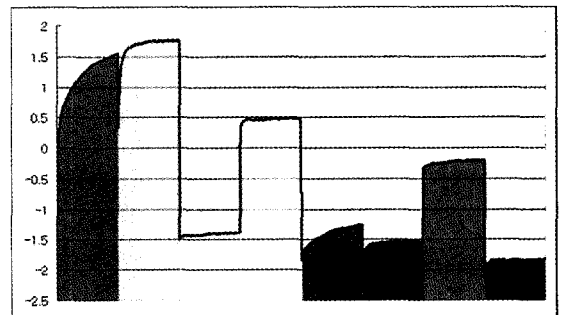


그림 9. 가스 신호 패턴 구간의 선형 배치

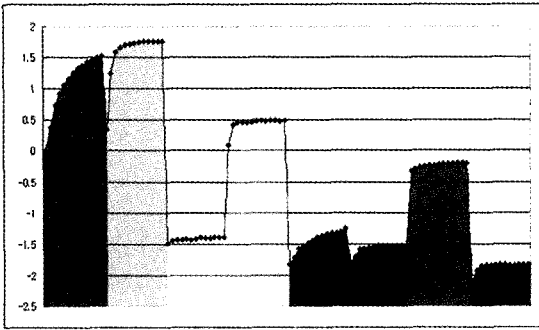


그림 10. 선형 가스 모델의 양자화

$$m = k \times l \text{ (식 2.1)}$$

다. 그림 9에서는 센서에서 각 채널별로 k 번 샘플링하여 추출하였다. 그림과 같이 패턴의 양자화로 각 요소간의 전이 관계를 추출할 수 있게 되었다. 센서 어레이에 포함된 센서 채널의 수를 l 번이라 할 때, 총 양자화 요소의 개수 m 은 식 (1)과 같이 나타낼 수 있다

$$m = k \times l \tag{1}$$

양자화 결과 모두 m 개의 점이 생성되었다. 각 점 사이의 x 축 거리는 동일 시간 간격으로 샘플링 하였기 때문에 일정하다. x 축 거리를 1로 놓았을 때 y 축의 증감(增減)값만으로 각 점과 점 사이의 기울기를 구할 수 있다. 양자화 된 i 번째의 측정값 q_i 는 식 (2)와 같이 표현할 수 있다.

$$q_1, q_2, q_3, \dots, q_m \tag{2}$$

3.4 ADSTM 구현 방법

ADSTM을 구성하는 최소 요소는 상태(state)와 상태들 사이의 천이(transition)관계이다. 각 측정값 사이의 관계를 천이 요소로 하여 모델을 만든다. q_i 에서 q_{i+1} 로 향하는 모든 벡터를 구하면 $\vec{q_1q_2}, \vec{q_2q_3}, \dots, \vec{q_{m-1}q_m}$ 와 같이 $(m-1)$ 개의 벡터를 얻을 수 있고, 이 벡터를 천이 벡터로 정의한다. 또한 본 논문에서 천이 벡터의 각은 x 축과 천이 벡터 사이의 각으로 정의한다. 본 논문에서 센서 어레이 신호를 모델링함에 있어 천이 벡터의 방향을 상태 천이 모델에서의 '상태'로 이용한다. 모든 천이 벡터의 x 축 값은 항상 양수이므로 천이 벡터의 방향은 항상 그림 10과 같이 좌표축의 원점을 기준으로 하였

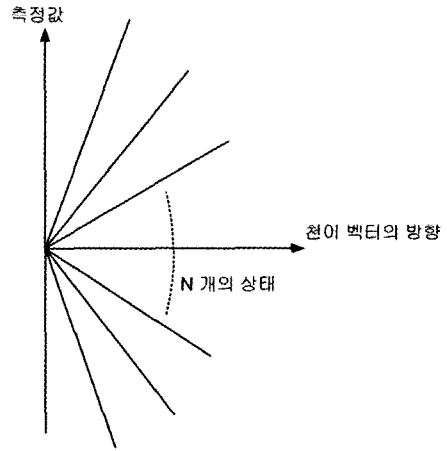


그림 11. 상태(state) 구분

을 때 1, 4 사분면 방향으로 구간 $(-\frac{\pi}{2}, \frac{\pi}{2})$ 에 존재한다. 따라서 구간 $(-\frac{\pi}{2}, \frac{\pi}{2})$ 을 n 개의 일정한 크기의 구간으로 나누어 각 구간을 하나의 상태로 정의한다. 이 때 n 은 계산상의 편의를 위하여 짝수를 취한다.

상태의 각 범위를 s_i 라 하면 i 번째 상태 s_i 는 식 (3)과 같다.

$$\begin{aligned} \frac{\pi}{n} \times (i-1) \leq s_i < \frac{\pi}{n} \times i \\ (0 < i \leq \frac{n}{2}) \text{ 일 때,} \\ -\frac{\pi}{n} \times (i - \frac{n}{2}) \leq s_i < -\frac{\pi}{n} \times (i - \frac{n}{2} + 1) \\ (\frac{n}{2} < i \leq n) \text{ 일 때,} \end{aligned} \tag{3}$$

$0 < i \leq \frac{n}{2}$ 는 천이 벡터의 기울기가 양수임을, $\frac{n}{2} < i \leq n$ 는 천이 벡터의 기울기가 음수임을 나타낸다. 표 1은 천이 벡터의 각을 상태 순서대로 표현한 것이다. 천이 벡터의 각을 천이 관계로 나타냈을 때 이 각의 크기는 센서의 측정값인 좌표 평면 내에서 천이 벡터의 x 값을 1로 정규화 하면 y 값만으로 아크탄젠트 함수를 이용하여 구할 수 있다. 기울기 값을 차례로 나열했을 때, 각 각의 천이 벡터 각 d_i 는 식 (4)와 같이 구할 수 있다.

표 1. 상태 기울기 범위

i	상태의 기울기 범위
1	$0 \leq S_1 < \frac{\pi}{n}$
2	$\frac{\pi}{n} \leq S_2 < \frac{2\pi}{n}$
\vdots	\vdots
$\frac{n}{2}$	$\frac{\pi}{2} - \frac{\pi}{n} \leq S_{\frac{n}{2}} < \frac{\pi}{2}$
$\frac{n}{2} + 1$	$-\frac{\pi}{n} \leq S_{\frac{n}{2}+1} < 0$
$\frac{n}{2} + 2$	$-\frac{2\pi}{n} \leq S_{\frac{n}{2}+2} < -\frac{\pi}{n}$
\vdots	\vdots
n	$-\frac{\pi}{2} \leq S_n < -\frac{\pi}{2} + \frac{\pi}{n}$

$$d_i = \arctan(q_i - q_{i-1}), 1 \leq i \leq m-1 \quad (4)$$

이렇게 계산된 천이 벡터의 각의 순차열을 이용하여 가스 신호 패턴 인식을 위한 상태 천이 모델이 만들어 진다.

구간 $(-\frac{\pi}{2}, \frac{\pi}{2})$ 를 n 개의 상태(state)로 나누었을 때, 각 상태마다 각의 범위는 표 1과 같이 나누어지고 천이 벡터의 각이 속하는 범위가 순차열에 표시된다. 생성된 순차열은 $n \times n$ 의 상태 천이 행렬을 만들어 상태로 변환될 수 있다. 상태 i, j 가 $1 \leq i \leq n, 1 \leq j \leq n$ 일 경

표 2. 상태천이 행렬

$i \backslash j$	s_1	s_2	s_3	\dots	s_n
s_1	a_{11}	a_{12}	a_{13}	\dots	a_{1n}
s_2	a_{21}	a_{22}	a_{23}	\dots	a_{2n}
s_3	a_{31}	a_{32}	a_{33}	\dots	a_{3n}
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
s_n	a_{n1}	a_{n2}	a_{n3}	\dots	a_{nn}

표 3. 샘플링 된 toluene 가스의 측정 데이터

시간	1	2	3	4	5
채널1	-0.273	0.151	0.370	0.508	0.624
채널2	-0.044	2.427	2.678	2.778	2.821
채널3	-0.787	-0.675	-0.673	-0.667	-0.444
채널4	-0.365	1.708	1.820	1.869	1.895
채널5	-0.845	-0.802	-0.770	-0.740	0.736
채널6	-0.928	-0.301	1.217	1.675	1.916
채널7	-0.313	-0.280	-0.250	-0.224	-0.204
채널8	-0.746	1.814	3.249	3.771	3.988

우, 상태 i 에서 상태 j 로 천이가 발생하여 a_{ij} 상태 천이 행렬이 생성되며 표 2와 같이 나타낼 수 있다.

4. 가스 센서 어레이 ADSTM 적용 예

3장에서 설명한 모델링 방법을 이용하여 실제로 센서 어레이에서 측정된 데이터를 이용하여 모델링을 실시해 본다. Toluene 가스를 측정 하였을 때, 의미 구간을 추출한 후 그림 7에서와 같이 한 번 측정한 구간의 데이터를 다섯 번 샘플링 한 양자화 결과를 표시해 보면 표 3과 같다. 각 값의 단위는 센서의 전압이다.

이 데이터들을 좌표 평면상에 채널 순서대로 나열하여 그래프로 나타내면 그림 11과 같이 나타낼 수 있으며 천이 벡터의 각은 표 4와 같다. 표에 표시된 기울기는 바로 앞 데이터와 이루는 각도를 표시한 것으로 그림 11의 꺾은선의 기울기와 같다. 또한 그림 11의 천이 벡터의 각을 계산하면 표 4와 같다.

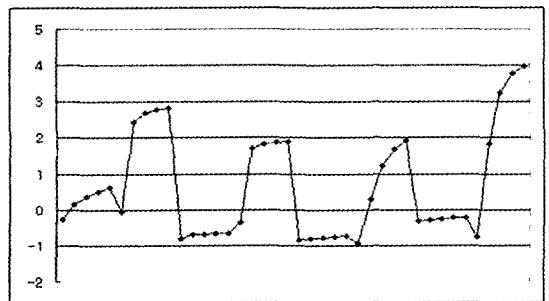


그림 12. 샘플링 과정을 거친 toluene 측정값

표 4. 상태천이 관계를 표현하기 위한 천이 벡터의 각 분포

측정치	각(°)	측정치	각(°)
-0.273	22.98	-0.845	2.46
0.151	12.35	-0.802	1.83
0.370	7.86	-0.770	1.72
0.508	6.62	-0.740	0.23
0.624	-33.74	-0.736	-10.87
-0.044	67.97	-0.928	50.87
2.427	14.09	0.301	42.49
2.678	5.71	1.217	24.61
2.778	2.46	1.675	13.55
2.821	-74.51	1.916	-65.84
-0.787	6.39	-0.313	1.89
-0.675	0.11	-0.280	1.72
-0.673	0.34	-0.250	1.49
-0.667	0.06	-0.224	1.15
-0.666	16.75	-0.204	-28.46
-0.365	64.25	-0.746	68.66
1.708	6.39	1.814	55.13
1.820	2.81	3.249	27.56
1.869	1.49	3.771	12.24
1.895	-69.95	3.988	-

3.4 절에서는 구간 $(-\frac{\pi}{2}, \frac{\pi}{2})$ 을 상태 천이 모델링에 서 상태의 개수로 나누었다. 예시의 편의를 위하여 나누 고자 하는 상태의 개수 $n=10$ 으로 하였다.

표 5. 상태 $n=10$ 일 때 상태의 범위

상태 (state)	범 위(각도)	상태 (state)	범 위(각도)
1	$0 \leq S_1 < \frac{\pi}{10}$	6	$-\frac{\pi}{10} \leq S_6 < 0$
2	$\frac{\pi}{10} \leq S_2 < \frac{\pi}{5}$	7	$-\frac{\pi}{5} \leq S_7 < -\frac{\pi}{10}$
3	$\frac{\pi}{5} \leq S_3 < \frac{3\pi}{10}$	8	$-\frac{3\pi}{10} \leq S_8 < -\frac{\pi}{5}$
4	$\frac{3\pi}{10} \leq S_4 < \frac{2\pi}{5}$	9	$-\frac{2\pi}{5} \leq S_9 < -\frac{3\pi}{10}$
5	$\frac{2\pi}{5} \leq S_5 < \frac{\pi}{2}$	10	$-\frac{\pi}{2} \leq S_{10} < -\frac{2\pi}{5}$

표 6. 각 측정치의 소속 상태

측정치	상태	측정치	상태
-0.273	2	-0.845	1
0.151	1	-0.802	1
0.370	1	-0.770	1
0.508	1	-0.740	1
0.624	7	-0.736	6
-0.044	4	-0.928	3
2.427	1	0.301	3
2.678	1	1.217	2
2.778	1	1.675	1
2.821	10	1.916	9
-0.787	1	-0.313	1
-0.675	1	-0.280	1
-0.673	1	-0.250	1
-0.667	1	-0.224	1
-0.666	1	-0.204	7
-0.365	4	-0.746	4
1.708	1	1.814	4
1.820	1	3.249	2
1.869	1	3.771	1
1.895	9	3.988	-

표 4의 기울기 값을 표 5의 상태 구분 기준에 맞추어 구분된 상태의 값으로 표시하면 표 6과 같다. 이를 상태 열(state sequence)로 바꾸면 표 7과 같이 나타낼 수 있다.

표 3의 데이터를 본 논문에서 고안한 각도 변이 기반 상태 천이 모델링 과정을 거쳐 표 7에서와 같이 최종 상태열로 변환되었다. 각 상태 사이의 상태 천이 관계를 다 이어그램으로 표현하면 그림 12와 같이 나타낼 수 있다. 각 상태 사이의 상태 천이의 횟수를 나타내었다. 수학적 연산을 위해 상태 천이 다이어그램을 행렬로 표현하면 표 8과 같이 나타낼 수 있다.

표 7. 최종 변환된 상태열

2 → 1 → 1 → 1 → 7 → 4 → 1 → 1 → 1 → 10 → 1 → 1 → 1 → 1 → 1 → 4 → 1 → 1 → 1 → 9 → 1 → 1 → 1 → 1 → 6 → 3 → 3 → 2 → 1 → 9 → 1 → 1 → 1 → 1 → 7 → 4 → 4 → 2 → 1
--

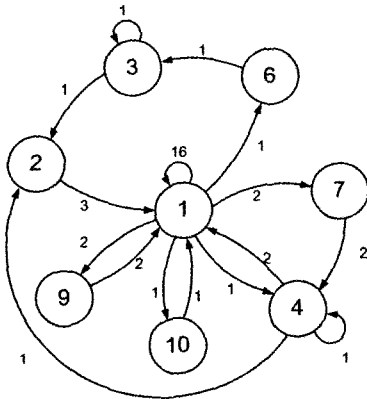


그림 12. 상태전이 다이어그램

표 8. 천이 횟수로 표현된 상태 천이 행렬

	s_1	s_2	s_3	s_4	s_5	s_6	s_7	s_8	s_9	s_{10}
s_1	16	0	0	1	0	1	2	0	2	1
s_2	3	0	0	0	0	0	0	0	0	0
s_3	0	1	1	0	0	0	0	0	0	0
s_4	2	1	0	1	0	0	0	0	0	0
s_5	0	0	0	0	0	0	0	0	0	0
s_6	0	0	1	0	0	0	0	0	0	0
s_7	0	0	0	2	0	0	0	0	0	0
s_8	0	0	0	0	0	0	0	0	0	0
s_9	2	0	0	0	0	0	0	0	0	0
s_{10}	1	0	0	0	0	0	0	0	0	0

표 8에서 생성된 상태 천이 행렬은 각 가스마다 서로 독립적인 형태를 가지게 된다. 동일한 가스에 대한 상태 천이 행렬은 상태의 수가 변하거나, 샘플링 회수가 변하더라도 비슷한 형태의 상태 천이 행렬을 가지게 된다. 따라서 천이 벡터의 상태 천이 행렬의 패턴은 가스를 분석하는데 사용할 수 있는 특징으로 사용될 수 있다.

5. 가스 신호 패턴 분류를 위한 프로세스 프레임워크

본 논문에서는 그림 13과 같이 가스 신호 패턴 분류를 위해 일반적인 프로세스 프레임 워크를 사용하였다.

그림 13에서 본 비와 같이 시스템은 크게 두 부분으로 구분된다. 센서 어레이에서 측정된 각 가스의 신호를 상태 천이 모델링 과정을 거친 후에 저장소에 보관하는 모델 생성

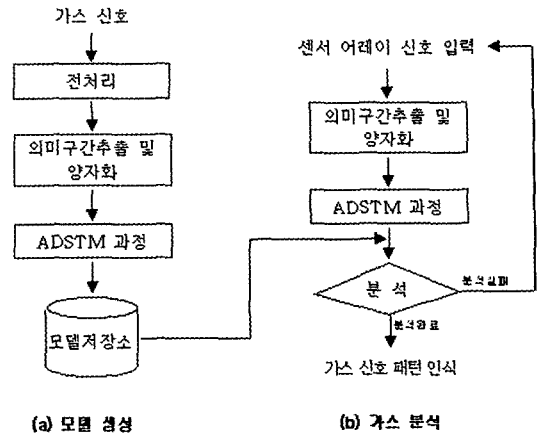


그림 13. 가스 신호 패턴 분류를 위한 프로세스 프레임워크

과정과 센서 어레이에서 측정된 새로운 가스 신호를 모델링 한 후에 분석하는 가스 분석 과정의 두 부분으로 나눌 수 있다.

각 상태 천이 행렬에는 상태 천이가 발생하는 곳에만 확률적 데이터가 입력된다. 변환의 기준은 상태 천이가 발생하는 위치에만 1의 값을 부여하는 것이다.

센서 어레이에서 새로 가스가 측정되어 생성된 상태 천이 행렬을 O , 모델 저장소의 비교 대상 모델의 상태 천이 행렬을 M 이라 하자. 각각 1의 값을 가지는 요소의 개수를 $n(O)$, $n(M)$ 이라 할 때, 두 행렬의 유사도(類似度)인 *similarity*는 식(2.5)와 같은 방법으로 계산할 수 있다.

$$similarity = \frac{n(O \cap M)}{n(O \cup M)} \tag{5}$$

여기에서 교집합(\cap)은 두 상태 천이 행렬에서 같은 위치의 행과 열에 동일하게 1의 값을 갖는 요소의 개수를 의미한다. 합집합(\cup)은 두 행렬에서 같은 위치에 어느 한 쪽이라도 1의 값을 가지고 있는 것이 개수를 의미한다. 이 유사도는 상대적 일치도 측정에 매우 유용한 방법이다.

8개의 실제 가스를 각각 여러 차례 측정된 자료를 바탕으로, 유사도에 대한 임계값(threshold)을 설정한다. 동일한 데이터를 이용하여 여러 번 데이터를 측정 후 각 데이터 간의 유사도 중에서 최소값으로 결정한다. 이 임계값을 기준으로 새 가스가 측정되어 유사도가 임계값을 넘지 못하는 경우 측정 과정을 반복한다. 일정 회수 이상 다시 측정하여도 분석이 불가능할 경우 분석 불가능한 가스로 처리한다.

6. 성능 실험 및 결과

이 장에서는 앞에서 설명한 모델링 방법을 이용하여 실제 측정된 데이터를 바탕으로 가스 모델에 대해 모델링을 실시하고, 새로운 측정 가스 데이터가 발생할 경우의 분석과정을 실험하여, 본 논문에서 제시하는 기법의 인식 효율성에 대하여 분석한다.

6.1 실험 방법

본 논문의 실험 대상 데이터는 여덟 가지 가스를 6회 반복 측정된 결과를 데이터로 사용하였다. 6회의 측정 결과 중에서 무작위로 4회의 측정 결과를 선택하여 모델링한 후에 임계값을 결정하고, 남은 2차례 측정 결과로 검증하는 과정을 반복하였다. 우선 본 논문의 모델링 방법의 변인 중에서 인위적으로 조절 가능한 상태의 개수(n)과 샘플링 횟수(k)의 값을 변화하여 가스 신호 패턴 인식의 기점인 임계값의 최적화 가능성에 대해 실험하였다. 다음으로 총 8개의 가스 모델 중에서 각 1개의 모델을 제외한 7개의 모델만 가지고 8가지 가스를 각각 분석하여, 생성한 모델에 만족하는 가스가 없을 경우의 측정치를 추정하였다. 1개의 모델을 제외하여 측정하는 과정을 8번 반복하여 8개의 모델이 각기 없을 경우에 측정치를 추정하여 임계값을 결정하는데 반영하였다.

6.2 샘플링 횟수의 고정 실험

첫 번째 실험은 샘플링 횟수(k)를 고정하고 상태의 개수(n)를 변화하여 실험하였다. 샘플링 횟수 $k=13$ 으로 고정하여, 채널의 수 $l=8$ 이므로, 양자화 요소의 개수 $m=104$ 로 고정하였다. 상태의 개수 n 을 20, 60, 100으로 변화하면서 분석의 속도와 정확도, 유사도 등의 변화를 살펴보고 임계값을 추정할 수 있도록 실험하였다.

그림 14는 8가지 가스 측정 결과 중에서 benzene 가스에 대한 실험 결과를 보여준다. n 값을 증가시킬수록 유사도 값의 크기가 감소되는 결과를 볼 수 있다. 유사도 값의 크기가 감소하는 것은 비교하고자 하는 가스 모델 사이에 인식할 수 있는 상대적인 값의 크기가 작아진다는 의미로, 인식의 가능성이 낮아짐을 의미한다. 인식의 확률을 높이기 위해서 n 의 값을 적당히 작게 해야 한다는 것을 n 값을 변화시킨 실험을 통해 유추할 수 있었다.

하지만, 분석 과정의 시간복잡도는 $O(N^2)$ 으로 n 의 값이 20에서 60으로 3배 늘어날 경우 3²만큼, 20에

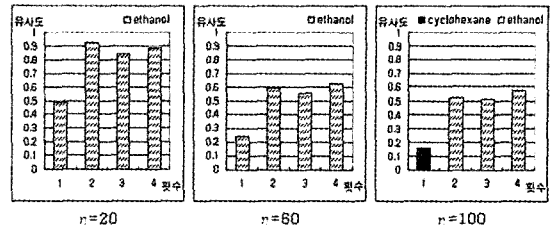


그림 15. 모델저장소에서 benzene을 제외하고 ethanol가스를 측정된 경우

서 100으로 5배 늘어날 경우 5²배만큼 커진다. 가스 분석에서는, 유독한 가스에 노출될 경우 측정 시간이 길어지면 인명피해가 발생할 가능성이 큰 경우에서처럼, 상황에 따라서 분석에 소요되는 시간이 알고리즘의 효율성을 나타내는 중요한 척도로 사용된다.

n 의 값이 증가할수록 확률 유사도 값이 작아지고 n 의 값이 커질수록 배수의 제곱만큼 분석 시간이 증가한다는 실험 결과에서 볼 때, n 의 값은 가스의 패턴 특성을 나타낼 수 있는 한도 내에서 작은 값을 선택하는 것이 임계값 설정이나, 가스의 패턴 구분에 유리하다.

가스 구분을 위한 임계값을 설정하는데 있어 중요한 요소는 두 가지이다. 첫째는 모델 저장소에 없는 미지의 가스를 측정할 때, 임계값이 미지의 가스에 대한 유사도보다 커야한다는 것이다. 둘째는 임계값을 정할 때 해당 가스를 정확히 구분해 내는 유사도 보다 작거나 같아야 한다.

$n=20$ 일 때의 실험 결과를 이용하여 임계값을 결정해 보겠다. 그림 15를 살펴보면 benzene을 모델 저장소에서 제외했기 때문에 benzene 가스는 구분해 낼 수 없어야 한다. benzene 가스를 측정된 항목을 살펴보면 유사도의 최대값이 toluene과의 유사도인 0.446809임을 알 수 있다. Benzene은 미지의 가스이므로 toluene으로 잘못 인식할 수 있기 때문에 임계값이 이 유사도보다 커야 미지의 가스인 benzene을 다른 가스로 인식하는 잘못된 결과를 피할 수 있다. 또한, 각 가스를 구분해 내는 유사도 중에서 앞의 조건인 미지의 가스 유사도 최대값보다 크고 실제 가스를 인식해낸 유사도 중에서 가장 작은 값은 ethanol을 첫 번째 측정했을 때의 유사도인 0.484848이 된다. 이 두 조건을 만족하는 임계값의 범위는 식 (6)과 같다.

$$0.446809 < \text{임계값} \leq 0.484848 \quad (6)$$

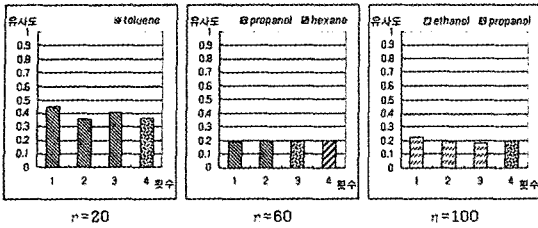


그림 16. 모델저장소에서 benzene을 제외하고 benzene가스를 측정할 경우

이 구간 사이의 값을 임계값으로 설정하면 적은 횟수의 비교를 통해 구분해 낼 수 있다. 따라서 이 구간 사이의 값을 benzene 모델을 제외한 7가지 가스 모델에 대한 측정 임계값으로 선택할 수 있다. 전체 모델에 대한 임계값을 도출하기 위해서는 각 가스를 제외했을 경우의 유사도 최대치를 구해야 한다. $n = 20$ 일 때 모델 저장소에서 자신의 가스가 제외되었을 경우에 해당 가스의 유사도를 살펴보았을 때 표 9에서와 같이 heptane과 hexane에서 유사도가 상대적으로 높아진 경우가 발생하였다. 대부분의 가스 모델의 경우에는 자신의 가스가 포함되지 않았을 경우에 다른 가스에 대한 유사도가 낮았지만, heptane과 hexane은 비슷한 성분을 가지고 있어 상대적으로 상대방 가스에 대한 유사도 확률이 높게 나타났다. 이 경우에는 임계값의 최대값이 매우 높아져서 가스를 잘못 인식하게 된다. 그러므로 예상보다 높아진 값을 낮추는 방법이 필요하다.

6.3 상태 개수 고정 실험

앞의 샘플링 횟수를 고정한 실험에서는 상태의 개수 n 의 적절한 크기에 대해 실험해 보았다. 실험 결과 n 을 20, 60, 100으로 실험한 결과 중에서 20일 경우 가장 나은 결과가 도출되었다. 하지만 heptane과 hexane의 경우에서처럼 임계값이 이번에는 최적의 샘플링 값을 얻고자 n 을 20으로 고정하고, 샘플링 횟수 k 의 값을 7, 13, 25로 변화시키면서 실험하였다. 실험결과 샘플링 횟수 k 의 값을 증가시켰을 경우보다는 감소시켰을 경우에 유사도가 커지거나 이전의 수준을 유지하는 유사도 분포를 보여주었다.

표 9를 살펴보면, 샘플링 횟수를 $k=7$ 로 작게 놓고 실험한 결과 $k=13$ 이었던 첫 번째 실험에서 문제가 된 heptane과 hexane 유사도 값이 비교 가능한 적정한 수준까지 낮아졌음을 볼 수 있다.

n 과 k 를 변경하는 실험을 바탕으로 $k=7, n=20$ 일 경

우 최적의 결과를 얻을 수 있었다. 이는 가스의 샘플링 구간이 짧아질수록(샘플링을 많이 할수록) 패턴의 특성을 잃기 때문에 적정한 샘플링 간격을 유지해야 패턴의 특성을 유지할 수 있다는 것을 의미한다. 이때 표 9에서의 음

표 9. k 값을 변화시킬 때 자신이 포함되지 않았을 경우 유사도의 변화

k = 13일 경우					
		Heptane		Hexane	
회수	최고유사도 가스	유사도	최고유사도 가스	유사도	
1	hexane	0.666667	heptane	0.548387	
2	hexane	0.620690	heptane	0.633333	
3	hexane	0.533333	heptane	0.633333	
4	hexane	0.666667	heptane	0.548387	
k = 7일 경우					
		Heptane		Hexane	
회수	최고유사도 가스	유사도	최고유사도 가스	유사도	
1	hexane	0.441176	heptane	0.470588	
2	hexane	0.470588	heptane	0.515152	
3	hexane	0.454545	heptane	0.515152	
4	hexane	0.529412	heptane	0.416667	

영 표시 값인 임계값을 모델에서 해당 가스를 제외했을 경우의 유사도 최대값 0.529412로 결정할 수 있다.

6.4 결정된 임계값을 이용한 가스 측정 결과

모델링 과정에서 $n = 20, k = 7$ 로 모델링 계수를 결정하고, 이때 결정된 임계값 0.529412를 적용하여 두 차례씩 새로운 가스 데이터로 실험하였다. 실험결과로 그림 16과 같이 모든 가스가 임계값을 넘어 구분해 낼 수 있다.

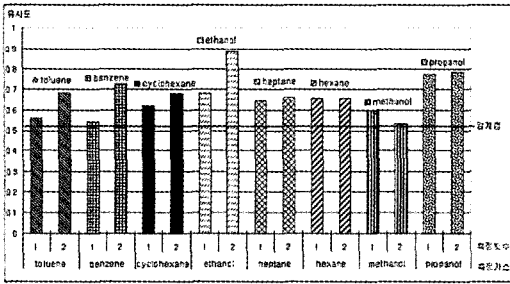


그림 17. 임계값 적용 후 새 가스 측정결과

7. 결 론

본 논문에서는 상태 천이 모델을 가스 신호 패턴 인식에 적용하여 만족할 만한 결과를 얻어낼 수 있는지를 살펴보았다. 실험 과정에서는 보다 나은 결과를 얻기 위해 상태의 수와 샘플링 횟수를 변인으로, 가스 신호 패턴 인식의 기준점인 임계값을 찾아내는 과정에 대하여 실험하였다. 샘플링 횟수를 고정하고 상태의 개수를 변인으로 한 실험에서는 상태의 수가 늘어날수록 유사도의 값은 감소하고 연산시간은 증가하였다. 상태의 개수를 고정하고 샘플링 횟수를 변인으로 한 실험에서는 샘플링 횟수 k 가 7, 13, 25로 변할 때 7의 경우에 가장 만족할 만한 임계값을 얻을 수 있는 유사도 분포를 구할 수 있었다. 두 차례의 실험을 통해 얻어진 임계값을 이용하여 새 가스에 대한 인식도를 알아볼 수 있었다.

실험을 통해 본 논문에서 설계한 모델링 방법이 가스 패턴을 구분하는데 있어 적용 가능한 방법이라는 것을 알 수 있었고, 인식 결과 가스 구분에 효과적으로 작동하였다. 또한 상태 천이 모델을 이용하여 모델링 한 결과 기기의 자원을 비교적 적게 소모하고, 빠른 시간 내에 모델을 생성할 수 있었다.

하지만, 모델링을 통한 가스 신호 패턴 인식에 있어 추정된 확률 값의 신뢰 여부를 결정하기 위한 임계값을 결정하는 과정을 통하여 본 논문에 사용한 가스들의 범위 내에서는 만족할 만한 결과를 얻었지만, 신뢰도를 보다 향상시키기 위하여 앞으로의 연구에서는 추가 실험을 통해 가스의 범위를 넓혀 신뢰도 높은 임계값을 결정할 수 있을 것이다.

본 논문은 가스 센서 인식 분야에서 기존에 거의 시도되지 않았던 상태 천이 모델을 적용하여 인식 가능한 모델을 보일 수 있었던 것이 하나의 성과라 할 수 있다. 보다 나은 신뢰도를 얻기 위해서는 보다 강건한 모델링 방법을 확립, 모델링 오류를 줄이기 위한 전처리 기법의 고안 등을 연구 과제로 남기며 결론을 맺는다.

참 고 문 헌

1. 강경아(1997), 백 프로퍼제이션과 홉필드 신경망을 이용한 형상인식 비교분석에 관한 연구, 조선대학교 대학원, pp. 13.
2. 김상진(1998), “자동화를 위한 센서”, 研學社, pp. 15-20.
3. 이준섭(2002), “필드 스크린을 위한 전자코 시스템의 구현”, 삼척대학교 산업대학원 전자공학 석사학위 논문, pp. 14-15, 18-21.
4. 홍형기 외(1995), “전자코 시스템의 기술동향”, 전기전자재료학회지 Vol.8. No.4, pp. 512-513.
5. Hiroshi SHIMODAIRA, Takashi SUDO, Mitsuru NAKAI and Shigeaki SAGAYAMA(2003), On-Line Overlaid-Handwriting Recognition Based on Substroke HMMs, Proceeding of the Seventh International Conference on Document Analysis and Recognition(ICDAR 2003), pp. 1-5.
6. Jianying Hu, Michael K. Brown and William Turin (1996), HMM Based On-Line Handwriting Recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence Vol 18 No 10, pp. 1039-1041.
7. Jim Lipman(1998), E-noses Nose Out Traditional Odor-Detection Equipment, EDN, pp. 62-64.
8. Oscar E. Agazzi and Shyh-Shiaw Kuo(1993), Hidden Markov Model based Optical Character Recognition in the Presence of Deterministic Transformations, Pattern Recognition Vol 26 No 12 Pattern Recognition Society, pp. 1814-1815.



김 아 람 (cciel@hanmail.net)

2003 아주대학교 정보 및 컴퓨터공학부 학사
2006 강릉대학교 교육대학원 컴퓨터교육전공 석사
2006~현재 강원도 평창교육청 지방교육전산서기보

관심분야 : 패턴인식, 무선센서네트워크



이 승 재 (silveree@kangnung.ac.kr)

1998 강릉대학교 물리학과 학사
2001 강릉대학교 컴퓨터공학과 석사
2006 강릉대학교 컴퓨터공학과 박사수료
2005~현재 강릉대학교 해양센서네트워크시스템기술연구센터 연구원

관심분야 : 해양센서네트워크, 분산데이터처리



김 창 화 (kch@kangnung.ac.kr)

1985 고려대학교 수학교육 학사
1987 고려대학교 전산학 석사
1990 고려대학교 전산학 박사
1989~현재 강릉대학교 컴퓨터공학과 교수
1994 ~ 1995 Toronto 대학교 Enterprise Integration 연구소 Post-Doc.
2002 ~ 2004 Texas A&M 대학교 Computer Science 학과 Visting Scholar
2005~현재 강릉대학교 해양센서네트워크시스템기술연구센터 소장

관심분야 : 해양센서네트워크, 분산데이터처리



김 상 경 (skkim98@kangnung.ac.kr)

1985 고려대학교 전자공학 학사
1987 고려대학교 컴퓨터공학 석사
2002 고려대학교 정보처리 박사
2003~현재 강릉대학교 컴퓨터공학과 교수
2005~현재 강릉대학교 해양센서네트워크시스템기술연구센터 연구책임자

관심분야 : 센서네트워크, 애드혹 라우팅



박 수 현 (shpark21@kookmin.ac.kr)

1988 고려대학교 컴퓨터학과 이학사
1990 고려대학교 대학원 전산학 이학석사
1998 고려대학교 대학원 컴퓨터학과 이학박사
1990~1999 (주)LG전자 중앙연구소 선임연구원
1999~2001 동의대학교 공과대학 소프트웨어공학과 교수
2002~현재 국민대학교 비즈니스IT학부 교수
2001~현재 한국SI학회 이사

관심분야 : 유비쿼터스 네트워크, 이동통신 시스템, Active Network