KIPS Tr. Software and Data Eng.
Vol.5, No.10 pp.503~510 pISSN: 2287-5905

http://dx.doi.org/10.3745/KTSDE.2016.5.10.503

Feature-Strengthened Gesture Recognition Model Based on Dynamic Time Warping for Multi-Users

Suk Kyoon Lee† · Hyun Min Um†† · Hyuck Tae Kwon†††

ABSTRACT

FsGr model, which has been proposed recently, is an approach of accelerometer-based gesture recognition by applying DTW algorithm in two steps, which improved recognition success rate. In FsGr model, sets of similar gestures will be produced through training phase, in order to define the notion of a set of similar gestures. At the 1st attempt of gesture recognition, if the result turns out to belong to a set of similar gestures, it makes the 2nd recognition attempt to feature-strengthened parts extracted from the set of similar gestures. However, since a same gesture show drastically different characteristics according to physical traits such as body size, age, and sex, FsGr model may not be good enough to apply to multi-user environments. In this paper, we propose FsGrM model that extends FsGr model for multi-user environment and present a program which controls channel and volume of smart TV using FsGrM model.

Keywords : Gesture Recognition, Dynamic Time Warping(DTW), Machine Learning

다중 사용자를 위한 Dynamic Time Warping 기반의 특징 강조형 제스처 인식 모델

이 석 균† · 엄 현 민†† · 권 혁 태†††

요 약

최근 제안된 FsGr 모델은 가속도 센서 기반의 제스처 인식을 위한 방법으로 DTW 알고리즘을 두 단계로 적용하여 인식률을 개선하였다. FsGr 모델에서는 유사제스처 집합 개념을 정의하는데 훈련과정에서 유사제스처 집합들을 생성한다. 제스처 인식의 1차 인식 시도에서 유사제스처 집합이 정의된 제스처로 판정되면, 이 유사제스처 집합의 제스처들에 대해 특징이 강조된 부분들을 추출해 DTW를 통한 2차 인식을 시도한다. 그러나 동일 제스처도 사용자의 신체 크기, 나이, 성별, 등의 신체적인 특징에 따라 매우 다른 특성을 보이고 있어 FsGr 모델을 다중 사용자 환경에 적용하기에는 한계가 있다. 본 논문에서는 이를 다중 사용자 환경으로 확장한 FsGrM 모델을 제안하고 이를 사용한 스마트TV의 채널 및 볼륨 제어 프로그램을 보인다.

키워드 : 제스처 인식, Dynamic Time Warping(DTW), 기계학습

1. 서 론

스마트폰의 대중화로 인해 스마트폰들의 센서들을 활용하는 연구들과 이들을 기반으로 애플리케이션 개발이 활성화되고 있다. 이들 중 가속도 센서는 제스처 인식에 사용되곤 하는데 기계학습, 패턴매칭 등의 분야의 DTW(dynamic time warping)[1], SVM(support vector machine), hMM(hidden Markov Model), 인공신경망 등의 알고리즘과 주로 사용되었다[2-11]. 최근 가속도 센서를 사용한 DTW 기반의 제스처 인식 방법으로 FsGr(Feature-Strengthened Gesture Recognition) 모델이 발표되었는데[2], 본 논문에서는 이를 다중 사용자 환경으로 확장한 FsGrM(Feature-strengthened Gesture recognition for Multi-users) 모델을 제안한다.

최근 가속도 센서 기반의 제스처 인식에 DTW가 사용된 연구 결과들이 발표되었다[2-7, 10]. 이들 중 대표적인 연구 결과로는 Liu의 uWave[3]를 들 수 있는데, 이는 개인화된 제스처 인식 알고리즘으로 여덟 개의 제스처들의 4000개 이상의 데이터들에 대해 95% 수준의 인식 정확도를 보였다. 남상하 외 3인은 스마트폰의 가속도 센서를 사용한 제스처 연구를 발표했는데, 훈련 방법으로 DTW를, 인식 방법으로

† 종신회원 : 단국대학교 소프트웨어학과 교수
†† 비 회 원 : 단국대학교 소프트웨어학과 학사과정
††† 정 회 원 : 단국대학교 컴퓨터과학과 박사과정수료
Manuscript Received : July 19, 2016
Accepted : August 9, 2016
* Corresponding Author : Suk Kyoon Lee(sklee@dankook.ac.kr)

는 DTW와 k-최근접 이웃 알고리즘을 사용하여 수행하였다[4]. Ko외 3인은 다수의 센서들로부터 생성된 다차원 데이터 시컨스들에 대해 시간적 융합을 시도하고 이에 대해 DTW를 적용하여 사용자의 행위를 추정하는 연구를 발표했고[5], Gillian외 2인은 다차원 DTW 알고리즘을 사용하여 음악적 제스처에 대한 인식을 시도했다[6].

FsGr 모델에서는 사용자의 제스처에 대해, 즉 제스처에 대한 스마트폰 가속도 센서 데이터에 대해 두 단계에 걸쳐 DTW 알고리즘을 적용하여 제스처 인식을 시도하는데, 첫 번째 단계에서는 모든 제스처들에 대해, 두 번째는 가능성이 높은 일부 제스처들에 대해 진행한다. **FsGr** 모델의 훈련 단계에서는 동작이 비슷해서 잘못 인식될 가능성이 높은 유사한 제스처들에 대해 유사제스처 집합들을 정의하고 DTW 기반의 1차 인식 시도의 결과가 유사제스처 집합이 정의된 제스처로 판정되면, 이 유사제스처 집합에 속한 제스처들에 대해 특징이 강조된 부분들만을 추출해 DTW 기반의 2차 인식 시도를 하여 인식률을 높였다[2].

그러나 동일한 제스처도 남여의 차이, 나이의 차이, 신체의 크기 등에 따라 다른 특성을 보이고 있어 이를 고려하지 않을 경우 제스처의 인식률의 제고에 한계가 있다. 본 논문에서는 사용자들의 다양한 특성들을 고려하도록 **FsGr** 모델을 확장한 다수 사용자를 위한 특징 강조형 제스처 인식(Feauture-Strengthened Gesture Recognition for Multi-users, **FsGrM)** 모델을 제안하고 이를 스마트 TV의 채널 및 볼륨 제어에 적용한 결과를 제시한다. 논문이 구성은 다음과 같다. 2절에서는 연구배경으로 **DTW** 알고리즘과 **FsGr** 모델의 개요를 소개하고, 3절에서 **FsGrM** 모델을, 4절에서는 이를 스마트 TV 응용에 적용한 결과를 제시하고 5절에서 결론을 맺는다.

2. 연구 배경

2.1 DTW와 제스처 인식

DTW(Dynamic Time Warping)은 음성인식, 데이터 마이닝, 제스처 인식 등의 시계열 데이터의 패턴 인식을 위한 알고리즘으로 길이가 동일하지 않은 시계열 데이터 시컨스들 사이의 유사도의 측정에 사용된다[1, 3, 5, 7, 10]. 이는 두 시계열 데이터 시컨스 $\boldsymbol{p} = p_1,p_2,\cdots,p_m$, $\boldsymbol{q} = q_1,q_2,\cdots,q_n$에 대한 비선형 대응(nonlinear alignment)을 통해 $\boldsymbol{p}$와 $\boldsymbol{q}$에 대해 누적 거리 비용을 최소화하는 일련의 대응(p_i,q_j) 시컨스를 구한다. 임의의 대응 (p_a,q_b)의 거리 비용 함수가 d(p_a,q_b)일 때, 함수 D는 다음과 같이 누적 거리 비용을 계산한다.

$$D(i, j) = d(p_i,q_j) + \min\{ D(i-1, j-1), D(i-1, j), D(i, j-1) \} \quad (1)$$

길이가 m, n인 두 데이터 시컨스 $\boldsymbol{p}$와 $\boldsymbol{q}$의 최소 누적 거리 비용은 D(m, n)로 표현된다. 본 논문에서는 제스처를 측정을 위해 가속도 센서 데이터를 사용하므로 거리 비용 함수 d는 두 가속도 센서 데이터의 유클리디언 거리를 계산하고, 데이터 시컨스 $\boldsymbol{p}$와 $\boldsymbol{q}$의 유사도는 $\boldsymbol{DTW(p, q)} = D$(m, n)로 정의한다. **G**를 식별하고자 하는 제스처들의 집합, **T**를 제스처 인식에 사용되는 대표 데이터 시컨스 (exemplar)들의 집합이라 할 때, 이는 각각 다음과 같이 표시한다.

$$\mathbf{G} = \{ g_1, g_2, \cdots, g_n \}, \mathbf{T} = \{ t_1, t_2, \cdots, t_n \} \quad (2)$$

T에 속한 모든 대표 데이터 시컨스는 훈련 과정을 통해 생성되며 간단히 대표 시컨스로 호칭한다. 제스처 g_i ($1 \le i \le n$)의 인식에는 대표 시컨스 t_i가 사용되는데 편의상 같은 인덱스 첨자를 사용한다. DTW 기반의 제스처 인식 방법들[2-5]은 임의의 데이터 시컨스 t의 제스처를, t와 **T**의 모든 대표 시컨스들에 대해 $Arg\,Min_{x \in T}\, DTW(t, x)$이 나타내는 대표 시컨스의 제스처로 추정한다. 즉 가장 작은 $\boldsymbol{DTW}$ 비용 함수 값을 보이는 대표 시컨스의 제스처를 t의 제스처로 결정한다.

2.2 FsGr 모델의 개요

FsGr 모델은 DTW 기반의 제스처 인식의 인식률을 높이기 위하여 제스처 g_i에 대해 동작이 유사하여 잘못 인식될 가능성이 높은 제스처들의 집합 즉 유사제스처 집합 $\mathbf{sG_i}$를 정의하고, 1차 DTW 인식의 결과로 판정된 제스처에 유사제스처 집합이 존재하는 경우 이들에 대해 2차 DTW 인식 작업을 수행한다. Table 1에서는 영어 알파베트 소문자의 인식 실험에서 사용된 유사제스처 집합들이 제시되는데, 제스처 b의 유사제스처 집합은 {p}이고 제스처 r의 유사제스처 집합은 {n, v}이다. **FsGr** 모델에서는 1차 DTW 인식 결과가 r인 경우, r과 {n, v}에 대해 2차 DTW 인식 작업을 진행한다. 유사제스처 집합은 **FsGr** 모델의 핵심 개념으로 이는 다음과 같이 정의된다. 임의의 데이터 시컨스 t와 대표 시컨스들의 집합 **T**에 DTW를 적용했을 때, t의 제스처를 g_i로 인식(판정)했으나 실제로는 g_k인 경우, g_k가 g_i의 유사제스처 집합 $\mathbf{sG_i}$에 속한다[2].

Table 1. Example of the Set of Similar Gestures

gi	$\mathbf{sG_i}$	$\mathbf{part\_bits_i}$
b	{p}	011010
r	{n, v}	001110
u	{y}	001010
v	{r}	000110

FsGr 모델에서는 제스처 g_i와 유사제스처 집합 $\mathbf{sG_i}$에 대한 2차 DTW 인식 작업을 위해 t_i와 $\mathbf{sG_i}$에 속한 제스처들의 대표 시컨스들을 재구성한다. 이때 차별성이 높은 부분들로 대표 데이터 시컨스들을 구성하기 위해 $\mathbf{part_bits_i}$를 사용한

다. Table 1에서는 6개의 비트로 구성된 **part_bits**들을 보이는데, 이는 원래 대표 시컨스를 6등분한 후 1로 설정된 부분만으로 대표 시컨스를 다시 구성함을 의미한다. 예를 들어 제스처 v와 이의 유사제스처 집합 {r}에 적용될 **part_bits** 000110은 v와 r의 대표 시컨스들을 6등분 한 후 네 번째와 다섯 번째 부분만으로 구성된 대표 시컨스들을 구성하며, 이들은 2차 DTW 작업에 사용된다.

제스처 g_i와 유사제스처 집합 $\mathbf{sG_i}$의 2차 DTW 인식 작업에 관한 정보는 **FsGr** 서브모델 $\mathbf{sM_i}$을 통해 표현된다. $\mathbf{sM_i}$은 $<g_i, \mathbf{sG_i}, \mathbf{sT_i}, \mathbf{part_bits_i}>$의 투플로 정의되는데, $\mathbf{sT_i}$는 g_i와 $\mathbf{sG_i}$에 속한 제스처들의 식별에 사용될 새로 구성된 대표 시컨스들의 집합이고, $\mathbf{part_bits_i}$은 g_i와 $\mathbf{sG_i}$에 속한 제스처들의 원래의 대표 시컨스들에 적용하여 $\mathbf{sT_i}$의 생성에 사용되는 비트 시컨스이다. **buildSeq**는 대표 시컨스 t에 대해 $\mathbf{part_bits_i}$를 적용하여 새로운 대표 시컨스를 반환하는 함수라 할 때, $\mathbf{sT_i}$는 다음과 같이 정의된다.

$$\mathbf{sT_i} = \{\ \mathbf{buildSeq}(\ t, \mathbf{part\_bits_i}) \mid t \in \mathbf{T} \wedge \text{gesture of } t \in \mathbf{sG_i} \cup \{g_i\}\ \} \quad (3)$$

SM을 모든 서브모델들의 집합, 그리고 **d**가 모든 서브모델에 적용할 **part_bits**의 길이라 할 때, **FsGrM** 모델은 **<G, T, SM, d>** 투플로 표현된다. 즉, **G**는 제스처들의 집합, **T**는 대표 시컨스들의 집합, 그리고 **SM**은 **G**에 관련된 **FsGr** 서브 모델들의 집합, **d**는 서브 모델에 적용할 **part_bits**의 길이를 결정한다. Table 1은 **d** = 6인 경우이다.

FsGr 모델의 제스처 인식 알고리즘은 인식 대상의 임의의 데이터 시컨스 t에 대해 1차 DTW와 2차 DTW의 두 단계로 진행된다.

1차 DTW: $ArgMin_{x \in T} DTW(t, x)$이 반환하는 대표 시컨스의 제스처 g_i에 대한 $\mathbf{sG_i}$가 공집합이면 g_i를 t의 제스처로 정하고 공집합이 아니면 2차 DTW로 진행한다.

2차 DTW: $\mathbf{sG_i}$의 **part_bits**이 a라 하면, $ArgMin_{x \in sTi}[\, DTW(buildSeq(t,a),x)]$이 반환하는 대표 시컨스의 제스처를 t의 제스처로 정한다.

3. FsGrM 모델

3.1 FsGrM 모델의 정의

사용자들은 나이, 성별, 신체의 크기 등에 따라 같은 제스처에 대해서도 차별적인 특징들을 지니고 있다. 따라서 사용자의 이러한 특성에 대한 고려 없이 제스처 인식을 시도할 경우 인식률이 현저히 저하될 수 있다. 따라서 본 절에서는 이러한 문제를 해결하기 위해 **다중 사용자들을 위한 특징 강조형 제스처 인식(FsGrM: Feature Strengthened Gesture Recognition for Multi Users)** 모델을 제안한다. **FsGrM** 모델에서는 제스처를 사용자의 차별적인 특징을 반영한 세분화된 제스처들로 표현한다. 사용자 u_j에 의한 제스처 g_i를 $g_i(j)$로 표현하며 $T(j)$는 u_j에 대한 제스처들의 대표 시컨스 집합을 나타낸다. 제스처 $g_i(j)$의 식별을 위해서는 대표 시컨스 $t_i(j)$가 사용되며 이들은 훈련을 통해 생성한다. **FsGrM** 모델은 사용자들의 집합 **U**, 제스처들의 집합 **G**, 모든 대표 시컨스들의 집합 **T**를 포함하며, **U**, **G**, **T**는 각각 다음과 같이 정의된다.

$$\begin{aligned} &\mathbf{U} = \{\ u_1, u_2, \cdots, u_m\ \}, \\ &\mathbf{G} = \bigcup_{i=1}^{m} G(i), \qquad \mathbf{T} = \bigcup_{i=1}^{m} T(i) \\ &\text{where} \quad G(j) = \{\ g_1(j), g_2(j), \cdots, g_n(j)\ \}, \\ &\qquad\quad T(j) = \{\ t_1(j), t_2(j), \cdots, t_n(j)\ \} \\ &\qquad\quad \text{for } 1 \le j \le m \end{aligned} \quad (4)$$

U에 속한 m명의 사용자들은 각각 차별적인 제스처 특징들을 보유하는 사용자들을 대표하며, $G(j)$는 사용자 u_j에 의한 n개의 제스처들의 집합, $T(j)$는 $G(j)$의 각각의 제스처에 대한 대표 시컨스들의 집합을 나타낸다. **T**는 모든 $T(j)$들의 합집합으로 m명의 사용자들의 n개의 제스처들에 대한 대표 시컨스들의 집합이다. 가령 $T(2)$는 사용자 u_2를 위한 대표 시컨스들의 집합을, $t_3(2)$는 u_2의 제스처 g_3, 즉 $g_3(2)$에 대한 대표 시컨스를 뜻한다. 문맥 상 사용자 구분이 중요하지 않을 경우 사용자 표기를 생략한다.

FsGrM 모델에서의 유사제스처 집합과 서브모델 등의 주요 개념들의 정의들은 **FsGr** 모델의 개념들에 사용자 개념을 포함하도록 확장된다. 일부 개념들의 확장된 정의는 그 의미가 명확하므로 다시 정의를 설명하지 않고 표기만 언급한다. 사용자 u_j의 제스처 g_i, 즉 제스처 $g_i(j)$에 대한 유사제스처 집합은 $\mathbf{sG_i(j)}$로, 이에 대한 대표 시컨스들의 집합은 $\mathbf{sT_i(j)}$로, 이에 대한 서브모델은 $\mathbf{sM_i(j)}$로 표현하며, $\mathbf{sM_i(j)}$은 $<u_j, g_i(j), \mathbf{sG_i(j)}, \mathbf{sT_i(j)}, \mathbf{part_bits_i(j)}>$의 튜플로 정의한다. 이때 $\mathbf{sT_i(j)}$는 $\mathbf{part_bits_i(j)}$가 적용된 대표 시컨스들의 집합으로 다음과 같이 정의된다.

$$\mathbf{sT_i(j)} = \{\ \mathbf{buildSeq}(t, \mathbf{part\_bits_i(j)}) \mid t \in \mathbf{T} \wedge \text{gesture of } t \in \mathbf{sG_i(j)} \cup \{g_i(j)\ \} \quad (5)$$

$$SM(j) = \{\ \mathbf{sM_1(j)}, \mathbf{sM_2(j)}, \cdots, \mathbf{sM_n(j)}\ \} \quad (6)$$

$$\mathbf{SM} = \bigcup_{j=1}^{m} SM(j) \quad (7)$$

$SM(j)$은 사용자 u_j에 대한 서브 모델들의 집합을, **SM**은

모든 사용자들을 위한 서브모델들의 집합을 나타낸다. 이때 **FsGrM** 모델 **M**은 **<G, T, U, SM, d>**의 튜플로 정의한다.

Table 2. Example of the Set of Similar Gestures in **FsGrM** Model

$g_i(j)$	$sG_i(j)$	$part\_bits_i(j)$
h(2)	{x(2)}	001000
p(2)	{b(2)}	011000
n(3)	{b(3), h(3), p(3), r(3), h(4)}	011101
y(3)	{q(2)}	010010
b(4)	{p(4)}	011010
n(4)	{p(6)}	001100
r(4)	{v(4)}	000110
u(4)	{y(4)}	001010
c(5)	{e(6)}	111000
p(5)	{b(5)}	000100
u(5)	{a(4), a(5), a(6)}	111100

Table 2에서는 소문자 알파벳 인식 실험에 **FsGrM** 모델을 사용하였을 때의 유사제스처 집합의 예를 보이는데, 이들은 Table 4의 실험에 사용되었다. 각 행은 제스처에 대한 서브 모델 정보 중 유사제스처 집합과 이에 적용된 **part_bits**을 보인다. 첫 번째 행은 사용자 u_2의 제스처 h의 유사제스처 집합은 동일 사용자의 제스처 x로 구성되어 있음을 의미한다. 세 번째 행은 사용자 u_3의 제스처 n의 유사제스처 집합은 사용자 u_3의 b, h, p, r과 사용자 u_4의 h로 구성됨을 보이는데, 이는 1차 DTW 결과가 n(3)일 경우, 이는 유사제스처 집합의 제스처가 n(3)으로 잘못 인식된 결과일 수 있음을 의미한다.

3.2 FsGrM 모델 기반의 제스처 인식 및 학습 방법

FsGrM 모델의 제스처 인식 방법은 **FsGr** 모델과 유사하다. 임의의 제스처 시컨스 t에 대해 인식 작업은 다음의 두 단계로 이루어진다.

1차 DTW: $Arg\,Min_{x \in T} DTW(t, x)$이 반환하는 대표 시컨스 $t_i(j)$의 서브 모델 $sM_i(j)$의 유사제스처 집합 $sG_i(j)$가 공집합이면 g_i를 제스처로 정하고 아니면 $sM_i(j)$에 대해 2차 DTW로 진행한다.

2차 DTW: $sM_i(j)$의 $part_bits_i(j)$이 a일 때, $Arg\,Min_{x \in sT_i(j)}[DTW(buildSeq(t,a),x)]$이 반환하는 대표 시컨스의 제스처를 t의 제스처로 정한다.

1차 DTW는 **FsGr** 모델에서와 동일하다. 단 하나의 제스처에 대해 복수의 대표 시컨스들이 존재하므로 최소 비용으로 결정된 대표 시컨스의 제스처와 사용자에 관한 서브 모델이 2차 DTW에 사용된다. 2차 DTW는 제스처 g_i에 연관된 모든 사용자들의 유사제스처 집합들이 아니라 1차 DTW에서 결정된 제스처 g_i와 사용자 u_j에 대한 서브 모델로 한정된다.

FsGrM 모델의 훈련 방법은 **FsGr** 모델의 훈련 방법을 사용자 별로 진행하도록 구성된다. 따라서 사용자 u_j별로 훈련을 진행하여 **G**에 대한 대표 시컨스의 집합 $T(j)$를 생성하는데 본 논문에서는 **FsGr** 모델의 기존 실험에서의 최소선택(minimum selection)을 사용했다. 각각의 $T(j)$를 통해 대표 시컨스들의 집합 **T**가 계산된 후, 각 제스처 $g_i(j)$에 대한 유사제스처 집합 $sG_i(j)$과 $part_bits_i(j)$을 구하고 이로부터 $sT_i(j)$을 도출한다. 이때 사용자에 따라 세분된 제스처들을 서로 다른 제스처로 간주하면 FsGr 모델의 훈련 알고리즘을 그대로 적용할 수 있다. 이에 대한 자세한 설명은 **FsGr** 모델의 훈련 알고리즘[2]을 참고하시오.

3.3 FsGrM 모델의 제스처 인식에 대한 성능 분석

사용자의 수를 m, 제스처의 수를 n이라 할 때 **FsGr** 모델은 훈련과정에서 n개의 대표 시컨스들을 생성하고 인식과정의 1차 DTW 작업에서는 입력된 제스처의 데이터와 n개 대표 시컨스들과 비교하나, **FsGrM** 모델에서는 훈련과정과 인식과정에서 m*n개의 대표 시컨스들을 생성하고 비교해야 되므로 비용이 증가한다. 따라서 **FsGrM** 모델을 사용할 경우, **FsGr** 모델을 사용했을 때의 작업에 대한 시간 복잡도보다 사용자 수 m의 곱만큼의 시간 복잡도가 늘어나게 된다. 그러나 **FsGrM** 모델을 사용할 시, 전체 사용자들이 아니라 특징적인 사용자들에 대한 학습을 가정하면 그 비용은 크지 않다. 또한 2차 DTW 작업은 서브 모델 내의 유사제스처 집합에 대해 이루어지므로 **FsGr** 모델과 **FsGrM** 모델의 시간 비용 즉 시간복잡도에서는 차이가 없다.

Table 3. Recognition Rates from Three Experiments

	실험1	실험2	실험3
1차 DTW 인식률	70.59	79.55	88.33
최종 인식률	71.28	81.28	88.78

본 절에서는 **FsGr** 모델과 **FsGrM** 모델을 통해 알파벳 필기체 소문자에 대한 인식을 시도하였다. 신체 조건이 각기 다른 여섯 명에 대해 열 번의 알파벳 필기체 소문자의 제스처 샘플들, 총 1,560개의 샘플들을 수집하고 이를 통해 훈련 및 인식 실험을 수행했다. 구체적인 실험 방법은 다음과 같다. **FsGr** 모델에 대해서는 한 명의 샘플로 훈련한 후 나머지 다섯 명의 제스처 샘플들로 인식 실험을 하는 경우(실험1)와 다섯 명의 샘플들로 훈련하고 나머지 한 명의 제스처 샘플로 인식 실험을 하는 경우(실험2), **FsGrM** 모델에서는 다섯 명의 샘플로 학습을 하고 나머지 한 명의 샘플로

Table 4. The Special Case of Experiment 3: Recognition Testing on u1's Samples and Training on Others' Samples

1차 DTW 실험 결과				최종 실험 결과		
인식 결과	횟수	실제문자/횟수	성공 (오인) 횟수	횟수	실제문자/횟수	성공 (오인) 횟수
a	11	a/10, u/1	10 (1)	11	a/10, u/1	10 (1)
b*	3	b/3	3 (0)	3	b/3	3 (0)
c*	12	c/10, o/2	10 (2)	12	c/10, o/2	10 (2)
d	10	d/10	10 (0)	10	d/10	10 (0)
e	10	e/10	10 (0)	10	e/10	10 (0)
f	13	t/3, f/10	10 (3)	13	t/3, f/10	10 (3)
g	10	g/10	10 (0)	10	g/10	10 (0)
h*	2	h/2	2 (0)	5	**b/1, h/4**	4 (1)
i	10	i/10	10 (0)	10	i/10	10 (0)
j	10	j/10	10 (0)	10	j/10	10 (0)
k	10	k/10	10 (0)	10	k/10	10 (0)
l	10	l/10	10 (0)	10	l/10	10 (0)
m	10	m/10	10 (0)	10	m/10	10 (0)
n*	22	p/2, b/2, h/8, n/10	10 (12)	19	p/2, **b/1**, **h/6**, n/10	10 (9)
o	8	o/8	8 (0)	8	o/8	8 (0)
p*	12	p/8, b/4	8 (4)	12	p/8, b/4	8 (4)
q	10	q/10	10 (0)	10	q/10	10 (0)
r*	11	b/1, r/10	10 (1)	11	b/1, r/10	10 (1)
s	10	s/10	10 (0)	10	s/10	10 (0)
t	7	t/7	7 (0)	7	t/7	7 (0)
u*	9	u/9	9 (0)	9	u/9	9 (0)
v	10	v/10	10 (0)	10	v/10	10 (0)
w	10	w/10	10 (0)	10	w/10	10 (0)
x	10	x/10	10 (0)	10	x/10	10 (0)
y*	10	y/10	10 (0)	10	y/10	10 (0)
z	10	z/10	10 (0)	10	z/10	10 (0)
합계	260		237 (23)	260		239 (21)

인식을 하는 경우(실험3)로 구성된다. 모든 실험은 모든 피험자들의 조합에 대해 진행되어서, 실험1에 사용된 총 샘플 수는 7,800개, 실험2와 실험3에서는 1560개이다. Table 3에서는 각 실험의 인식률 결과를 보인다.

1차 DTW 실행에 대한 실험1, 실험2, 실험3의 인식률은 각각 70.59%, 79.55%, 88.33%로 나타난다. 이는 DTW 만을 사용했을 경우로 한 명의 제스처 샘플로 훈련했을 경우보다는 사용자들의 구분 없이 훈련을 시켰더라도 다수 사용자들의 샘플들로 훈련했을 때 더 나은 성능을 보이고, 사용자 구분을 해서 훈련했을 경우 가장 높은 인식률을 보임을 나타낸다. 최종 인식률은 **FsGr** 모델과 **FsGrM** 모델이 사용된 최종 결과를 나타내는데 실험1, 실험2, 실험3의 인식률은 각각 71.28%, 81.28%, 88.78%로 1차 DTW만을 실행한 경우보다는 개선효과를 보인다. **FsGr** 모델의 경우, 다수의 사용자들을 훈련에 참여시켜도 각 제스처에는 하나의 대표 시컨스만이 존재하지만 **FsGrM** 모델은 각 제스처에 대해 훈련에 참여한 사용자 수만큼의 대표 시컨스들이 존재하게 되어 인식률이 높아지게 된다.

Table 4는 **FsGrM** 모델의 사용 예로 실험3의 일부 경우를 보인다. 이는 피험자들을 각각 u_1, u_2, u_3, u_4, u_5, u_6라 할 때, u_1의 제스처 샘플들로 인식 실험을 하고 나머지 사용자들의 샘플들로 훈련을 한 경우를 나타내며 표의 크기를 줄이기 위해 사용자 구분은 하지 않았다. 유사제스처 집합은 Table 2에 이미 제시되었지만 편의를 위해서 첫 번째 열의 알파벳 옆에 *로 표시했다. 참고로 2차 DTW 작업은 *로 표시된 행에 대해서만 이루어진다.

첫 번째 행은 유사제스처가 없는 경우이며 1차 DTW 결과로 열한번이 a로 인식되었지만 실제로는 열 번은 a, 한번은 u인 경우이며 이게 최종 결과로 이어진다. 열네 번째 행은 1차 DTW 결과 22번 n으로 인식되었는데 실제로 n인 경우는 열 번이고 나머지는 p, b, h가 잘못 인식된 결과를 보인다. 이 경우 22건에 대해 2차 DTW 실행을 하여 19번이 n으로 인식되고 나머지 세 번은 h로 두 번, b로 한 번 인식되어 두 건의 개선효과가 발생했다. 요약하면 1차 DTW 실행 결과 260개의 샘플들 중 237개가 정확히 인식해 1차 인식률이 91.15%, 2차 DTW 실행 결과 두 건의 개선효과가 발생하여 최종 인식률은 91.92%로 제시된다.

위의 실험들에서는 훈련한 피험자들의 샘플들을 인식과정

에 사용하고 있지 않다. 훈련한 피험자들의 샘플을 랜덤하게 인식과정에 사용하면 **FsGrM** 모델의 경우 인식률이 98%이상으로 나타난다. 만일 실험3의 테스트 데이터를 훈련 샘플이 아닌, 즉 훈련에 참가한 사람들의 제스처들의 새로운 데이터를 사용한다면 아마도 실험3의 최종 인식률 88.78% 보다는 높지만 훈련 샘플들을 사용한 경우 98%보다는 낮은 인식률을 보일 것으로 예상된다. 이는 알파벳의 난이도를 고려할 때 충분히 활용 가능한 수치로 보인다.

4. FsGrM 모델을 통한 스마트 TV의 채널 및 볼륨 제어

본 절에서는 **FsGrM** 모델의 사용 예로 스마트 폰의 가속도 센서를 사용하여 스마트 TV의 채널과 볼륨을 제어하는 프로토타입 시스템을 제시한다. 스마트TV에서 채널을 제어하기 위해서는 일련의 제스처 시컨스에 대한 인식이 필요하다. 가령 채널 102번을 선택하기 위해서는 1, 0 그리고 2를 순차적으로 한 번의 작업으로 인식해야 된다. 일반적으로 제스처 시컨스 내에서 각각의 제스처를 인식하기 위해서는 제스처의 시작과 끝을 구분할 수 있어야 하는데, 본 논문에서는 가속도 센서 데이터의 크기가 일정 이상이거나 미만인 경우를 포착하여 이를 제스처의 시작과 끝으로 구분하여 일련의 제스처를 인식하였다.

프로토타입 시스템의 구조는 Fig. 1에서 제시되는데 크게는 스마트 폰(안드로이드 OS) 애플리케이션과 스마트 TV (삼성, Orsay 플랫폼) 애플리케이션으로 구성된다. 스마트 폰 애플리케이션은 가속도 센서로부터 데이터를 수집하고 수집된 데이터 시컨스로부터 제스처 인식을 수행하여 결과를 송신한다. 스마트 TV 애플리케이션은 제스처 인식 결과를 수신하고 그 결과에 따라 채널 혹은 볼륨을 제어한다.

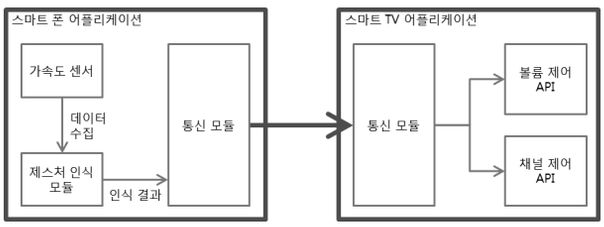


Fig. 1. Structure of Smart TV Control Prototype System

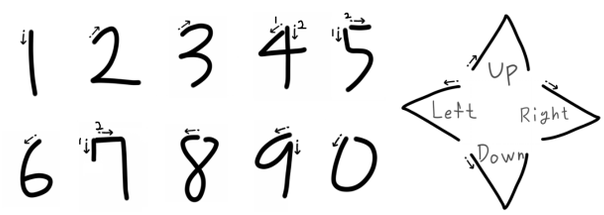


Fig. 2. Description of Gestures

볼륨 제어와 채널 제어는 **FsGrM** 모델을 기반으로 한 제스처 인식모듈의 인식 결과를 통해 이루어진다. 이를 위해 14개의 제스처들이 훈련되었는데 이들은 0~9의 아라비안 숫자와 상(Up), 하(Down), 좌(Left), 우(Right)를 뜻하는 심볼들로 구성된다. 각 심볼의 정확한 동작 방법은 Fig. 2에서 제시된다. 사용방법은 사용자가 제스처들을 수행하였을 때 그 인식 결과가 일련의 숫자들이라면 스마트 TV의 채널을 해당 숫자로 변경한다. 인식 결과가 Up이나 Down이라면 각각 채널을 1만큼 올리거나 내리고 Left나 Right라면 각각 1만큼 볼륨을 내리거나 올린다. Fig. 3은 제스처 인식을 위해 스마트 폰을 쥐는 방법과 모습을 나타낸다. 가속도 센서는 센서의 3축(x, y, z) 시계열 데이터 시컨스를 생성하므로 제스처를 그리는 방향과 방법 혹은 스마트 폰을 쥐는 방향이 정확해야 한다.

Table 5. Experiment Result of Recognition Testing for Controlling Channel and Volume of Smart TV.

1차 DTW 결과				2차 DTW 결과		
인식 결과	횟수	실제문자/횟수	성공 (오인) 횟수	횟수	실제문자/횟수	성공 (오인) 횟수
0	65	0/60, 6/5	60 (5)	65	0/60, 6/5	60 (5)
1	57	1/57	57 (0)	**58**	**1/58**	58 (0)
2	60	2/60	60 (0)	60	2/60	60 (0)
3	60	3/60	60 (0)	60	3/60	60 (0)
4	60	4/60	60 (0)	60	4/60	60 (0)
5	60	5/60	60 (0)	60	5/60	60 (0)
6	55	6/55	55 (0)	55	6/55	55 (0)
7	60	7/60	60 (0)	60	7/60	60 (0)
8	60	8/60	60 (0)	60	8/60	60 (0)
9	60	9/60	60 (0)	60	9/60	60 (0)
Down	61	1/1, Down/60	60 (1)	**60**	**Down/60**	60 (0)
Left	60	Left/60	60 (0)	60	Left/60	60 (0)
Right	60	Right/60	60 (0)	60	Right/60	60 (0)
Up	62	1/2, Up/60	60 (2)	62	1/2, Up/60	60 (2)
합계	840		832 (8)	840		833 (7)


(A) Front View


(B) Side View

Fig. 3. Pictures Showing How to Hold a Smartphone

스마트 TV의 채널 및 볼륨 제어 프로토타입 시스템의 성능 평가는 3절의 실험3과 동일한 조건으로 진행하였다. 6명의 840개의 샘플들이 사용되었는데 최종 인식률은 92.14%를 보인다. Table 5는 이 실험 결과를 제시하는데 0과 6, 그리고 1과 UP에서 약간의 오인 결과가 발생했다. 이를 반영해서 제스처의 동작 방법을 수정하면 인식률은 더욱 올라가고 3절에서 언급했던 것처럼 훈련한 사람들의 샘플을 인식에 사용할 경우는 99%이상의 인식률을 보일 것으로 예상된다.

5. 결 론

본 논문에서 제안한 FsGrM 모델은 기존의 FsGr 모델을 다중 사용자 환경으로 확장한 것으로 동일한 제스처에 대해서도 사용자의 신체적인 특징들을 훈련 과정에 포함시키므로 인식률을 높였다. 2차 DTW 인식 과정이 유사제스처 집합 내에서만 이루어지므로 시간 복잡도도 그다지 높지 않기 때문에 실용성이 있는 알고리즘이다. 본 연구에서는 스마트 TV의 채널 및 볼륨을 제어하는 프로그램의 예제를 제시했지만 이는 다양한 다른 분야에도 사용 가능하다. 또한 FsGrM 모델은 제스처 인식뿐 아니라 음성 인식, 데이터 마이닝 등 DTW가 사용되는 기존의 응용 분야에서 사용 가능해 보인다.

References

[1] R. Bellman, *Dynamic Programming*, Princeton University Press, Princeton, NJ, 1957.

[2] H. Kwon and S. Lee, "Feature-Strengthened Gesture Recognition Model based on Dynamic Time Warping," *KIPS Transactions on Software and Data Engineering*, Vol.4, No.3, pp.143-150, 2015.

[3] J. Liu, L. Zhong, J. Wickramasuriya, and V. Vasudevan, "uWave: Accelerometer-based personalized gesture recognition and its applications," *Pervasive and Mobile Computing*, Vol.5, Issue 6, pp.657-675, 2009.

[4] S. Nam, J. Kim, S. Heo, and I. Kim, "Smartphone Accelerometer-Based Gesture Recognition and its Robotic Application," *KIPS Transactions on Software and Data Engineering*, Vol.2, No.6, pp.395-402, 2013.

[5] M. Ko, B. West, S. Venkatesh, and M. Kumar, "Using dynamic time warping for online temporal fusion in multisensor systems," *Information Fusion,* Vol 9, Issue 3, pp.370-388, 2008.

[6] N. Gillian, R. Knapp, and S. O'Modhrain, "Recognition Of Multivariate Temporal Musical Gestures Using N-Dimensional Dynamic Time Warping," *Proc. of the International Conference on New Interfaces for Musical Expression*, pp.337-342, 2011.

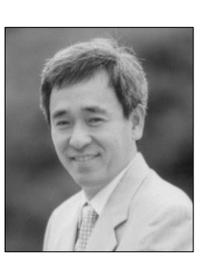
[7] M. Muller, "Information Retrieval for Music and Motion," Springer, 2007.

[8] S. Kim, G. Park, S. Jeon, S. Yim, G. Han, and S. Choi, "HMM-based Motion Recognition with 3-D Acceleration Signal," *KIISE Transactions on Computing Practices and Letters*, Vol.15, No.3, pp.216-220, 2009.

[9] S. Cho, W. Bang, J. Yang, "Two-stage Recognition of Raw Acceleration Signals for 3-D Gesture-Understanding Cell Phones," *Proc. of the 10th International Workshop on Frontiers in Handwriting Recognition*, 2006.

[10] Ahmad Akl, Chen Feng, and Shahrokh Valaee, "A Novel Accelerometer-Based Gesture Recognition System," *IEEE Transactions on Signal Processing*, Vol.59, No.12, Dec., 2011.

[11] Renqiang Xie and Juncheng Cao, "Accelerometer-Based Hand Gesture Recognition by Neural Network and Similarity Matching," *IEEE Sensors Journal*, Vol.16, No.11, 2016.



이 석 균

e-mail : sklee@dankook.ac.kr

1982년 서울대학교 경제학과(학사)

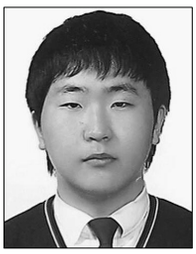
1990년 U. of Iowa, 전산학(석사)

1993년 U. of Iowa, 전산학(박사)

1993년~1997년 세종대학교 전임강사

1997년~현 재 단국대학교 소프트웨어학과 교수

관심분야 : 데이타 모델, 불완전 정보관리, 시각질의어, 문서의 변화 탐지 및 버전 관리, 기계학습



엄 현 민

e-mail : uhm0311@naver.com
2013년~현 재 단국대학교
소프트웨어학과 학사과정
관심분야 : 패턴 매칭, 기계학습

권 혁 태

e-mail : ceo@dgmit.com
2001년 단국대학교 전산통계학과(학사)
2003년 단국대학교 컴퓨터과학과(석사)
2007년 단국대학교 컴퓨터과학과
(박사과정수료)
2009년~현 재 디지엠정보기술(주)
대표이사
관심분야 : 멀티스크린 융합모델, 제스처 인식기반 사용자 경험