

동작 유사도와 적응 추이를 이용한 한국 수화 인식에서의 사용자에 대한 적응

Incremental User Adaptation in Korean Sign Language Recognition Using Motion Similarity and Prediction from Adaptation History

정성훈, Seong-Hoon Jung, 박광현, Kwang-Hyun Park, 변증남, Zeungnam Bien
한국과학기술원 전자전산학부 전기 및 전자공학전공

요약 최근 들어 손 제스처를 인간-기계 인터페이스에 활용하는 연구가 많아지고 있다. 그 중에서 수화 인식은 청각장애인과 일반인 사이의 원활한 의사 소통을 하게 해 주는 인터페이스로서 중요성이 날로 더해가고 있다. 하지만 기존의 수화 인식 연구는 사용자 개개인의 수화 동작의 차이를 고려하지 않고 다수 사용자를 위한 모델을 사용하기 때문에 사용자에 따라 인식이 낮아지게 된다. 이러한 점을 보완하기 위해 본 논문에서는 개개인의 수화 동작 특성을 반영하여 시스템이 사용자에 적응해 가는 과정을 다루고자 한다. 특히 점진적인 사용자 적응에 있어서 가장 문제가 되는 것은 어떻게 비관측된 상태(unobserved state)의 파라미터를 수정할 것인가 하는 것이다. 이를 위해서 본 논문에서는 동작 유사도와 적응 추이에 의한 추정을 통해 비관측된 상태의 모델 파라미터를 수정한다. 실제 청각 장애인들로부터 획득한 데이터베이스를 사용하여 제안한 방법이 기존 방법에 비해 더욱 빠르게 사용자의 특성을 시스템에 반영하고 인식을 향상시킨다는 것을 실험을 통해 보인다.

핵심어: 한국 수화 인식, 사용자 적응, 동작 유사도, 적응 추이

1. 서론

최근 들어 손 제스처를 인간-기계 인터페이스(Human Machine Interface)에 활용하는 연구가 많아지고 있다. 그 중에서 수화 인식은 청각장애인과 일반인 사이의 원활한 의사 소통을 하게 해 주는 인터페이스로서 중요성이 날로 더해가고 있다. 하지만 기존의 수화 인식 연구는 한 명의 사용자만을 대상으로 하거나 다수 사용자를 위한 모델을 사용하기 때문에 사용자 개개인의 수화 동작의 차이를 고려하지 않았다[1-13]. 전자의 경우에는 한 명의 사용자의 데이터에 기반한 것으로 이로부터 얻어진 모델을 수화자 종속 모델이라고 하고 후자의 경우에는 다수의 사용자의 데이터에 기반한 것으로 이로부터 얻어진 모델을 수화자 독립 모델이라고 한다. 현실적으로 한 명의 사용자로부터 충분한 양의 학습 데이터를 얻는 것은 힘들기 때문에 실제 상황에서 수화자 종속 모델을 사용하는 것은 어렵다. 다수의 사용자로부터 학습 데이터를 얻고 이로부터 확률에 기반한 모델을 생성하는 수화자 독립 모델은 구축이 용이한 반면 수화 동작의 개인차로 인해 인식이 상대적으로 낮다. 이러한 한계를 극복하기 위해서는 수화자 독립 모델을 기반으로 특정한 사용자의 동작 특성을 반영할 수 있도록 모델을 변형시키는 사용자 적응이 필요하다.

기존 제스처 관련 연구에서는 사용자 적응을 다루지 않았던 반면 음성인식에서는 이러한 연구가 많이 수행되어 왔다. 특히 은닉 마르코프 모델 기반의 인식을 수행하는 경우 기존

연구에서 사용된 사용자 적응 방법은 크게 두 가지로 나뉘어진다. 사전 확률이 정의되어 있는 경우와 사전 확률이 정의되어 있지 않은 경우를 나누어서 각각 MLLR (Maximum Likelihood Linear Regression)과 MAP (Maximum A Posteriori)로 알려진 두 가지 방법이 사용되었다[14,15].

사용자 적응이 시간의 경과에 따라 점진적으로 수행되는 경우 적응 데이터는 하나의 문장이 된다. 하나의 문장 안에 포함되어 있는 단어의 수는 한정되기 때문에 필연적으로 적응 데이터에서 관측되지 않은 상태(Unobserved State)가 존재하게 된다. 이 경우 비관측된 상태의 모델 파라미터를 수정할 수 없다. 이와 같이 부분적으로 적응되는 모델의 경우에는 전체 인식을 떨어뜨리게 되므로 비관측된 상태의 모델 파라미터 또한 함께 수정하는 것이 필요하다. 본 논문에서는 이처럼 수화자 독립 모델이 주어질 때 시간의 경과에 따라 제공되는 사용자의 수화 문장을 이용해서 점진적으로 적응하는 것을 다루면서 비관측된 상태의 모델 파라미터 또한 적응시키는 방법을 제안하고자 한다.

전체 논문의 구성은 다음과 같다. 먼저 2절에서는 한국 수화 인식에 대한 간략한 내용을 소개하고 3절에서는 기존의 사용자 적응에 대한 방법들을 정리한다. 4절에서는 제안하는 적응 알고리즘에 대해 서술하고 5절에서는 실제 청각 장애인들로부터 획득한 수화 데이터베이스를 사용하여 제안한 적응 알고리즘의 실험 결과를 보인다. 6절에서는 결론을 맺고 추후 연구 방향을 제시한다.

2. 한국 수화 인식 시스템

2.1 제스처 모델링

시공간상의 이중 확률 과정인 수화 동작을 은닉 마르코프 모델을 이용하여 모델링하였다. 하나의 은닉 마르코프 모델은 하나의 수화 단어에 해당하고, 인식은 사용자로부터 주어진 관측 데이터를 가장 잘 표현할 수 있는 특정 단어의 은닉 마르코프 모델을 찾는 과정이라 할 수 있다. 그리고 문장 단위의 인식의 경우에는 단어열을 찾게 된다. 은닉 마르코프 모델은 건너뛰기 존재하는 4개의 상태를 가진 왼쪽에서 오른쪽으로 진행되는 (Left-to-Right) 모델을 사용하였다. 특징 벡터는 양손으로부터 추출한 손 궤적과 손 모양의 정보로 구성하였고 손 모양의 경우 타원으로 근사화하여 추출한 특징을 이용하였다. 문장의 처음과 끝은 사용자가 정해준다고 가정하였다.

2.2 수화 데이터베이스

14살부터 19살 사이의 총 8명의 청각 장애 학생들로부터 수화 데이터를 획득하였다. 아래와 같은 17개의 수화 단어를 조합하여 만든 20개의 문장을 사용하여 문장 단위로 데이터를 수집하였으며 학생들은 각각의 문장을 10번 이상 반복하였다. 총 1891개의 문장 데이터와 12292개의 단어 데이터를 획득하였으며 각각의 문장 안에는 평균적으로 6.5개의 단어가 들어 있다. 다음은 데이터베이스에서 사용한 단어의 집합이다.

단어 집합 = {나, 우리, 어디에서, 살다, 알다, 즐겁다 만나다, 가족, 계속, 한국, 대전, 습니까, 습니다, 었, 에서, 있다, 할까요}

2.3 수화자 종속 모델과 독립 모델

인식 모델을 생성할 때 단일 사용자로부터 획득한 학습 데이터를 사용하여 생성하는 경우 수화자 종속 모델이라고 한다. 반면에 학습 데이터를 여러 명의 사용자로부터 획득하고 이를 사용하여 모델을 생성하는 경우 수화자 독립 모델이라고 한다. 수화자 종속 모델의 경우 한 명의 사용자로부터 많은 양의 데이터를 얻어야하지만 특정 사용자의 수화 특징을 잘 반영하고 있기 때문에 일반적으로 인식률이 높다. 반면에 수화자 독립 모델의 경우에는 여러 사람으로부터 데이터를 모으기 때문에 데이터 획득이 쉽다는 장점이 있지만 다수를 위한 인식 모델이므로 각 개인의 수화 특징을 반영하지 못하며, 일반적으로 수화자 종속 모델에 비해 인식률이 많이 떨어진다.

2.4 수화 인식 결과

구축한 데이터베이스를 바탕으로 인식 실험을 한 결과는 표 1과 같다. 전체 데이터의 70%를 학습을 위한 데이터로 이용하였고 나머지 30%를 실험 데이터로 이용하였다. 문장 인식의 경우에는 삽입 확률(Insertion Probability)을 모든 사용자에게 대해서 -100으로 고정하여 동일하게 사용하였다. 그리고 문장 안의 모든 단어가 올바르게 인식되는 경우에 한해서 문장 인식이 성공한다고 간주하였다.

표 1. 사용자에 따른 인식률

사용자	단어 인식률(%)		문장 인식률(%)	
	종속모델	독립모델	종속모델	종속모델
#1	96.86	82.20	66.10	28.00
#2	96.81	72.07	63.70	6.90
#3	89.00	73.37	42.11	14.04
#4	92.19	67.45	55.93	8.47
#5	89.40	59.90	42.00	5.08
#6	92.06	79.63	46.55	13.79
#7	88.43	56.68	40.38	0.00
#8	95.13	74.36	65.00	10.00

앞서 서술한 바와 같이 사용자 독립 모델을 사용하는 경우 단어 인식과 문장 인식 모두에서 낮은 인식률을 보였다. 이는 각 개인이 가지는 수화 동작의 차이가 독립 모델에는 잘 반영되어 있지 않기 때문이다. 이러한 동작의 개인적 차이는 실제 모델 파라미터 값의 차이로 나타난다. 그림 1은 동일한 수화 단어 모델의 특정 상태의 파라미터가 사용자마다 다르게 나타나고 있는 것을 보여준다.

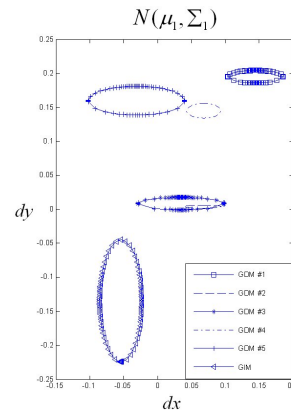


그림 1. 동일한 단어 “알다”에 대해 사용자마다 서로 다른 state 파라미터

이와 같이 파라미터의 분포가 사용자마다 다르기 때문에 수화자 독립 모델의 경우에는 인식률이 낮게 된다. 그러므로 독립 모델이 주어졌을 때 기존 모델의 파라미터를 변화시켜서 특정 사용자의 특성을 잘 반영할 수 있도록 하는 사용자 적응의 과정이 필요하다. 다음 절에서는 이러한 사용자 적응의 기존 방법과 한계점에 대해 살펴본다.

3. 기존의 사용자 적응 방법과 한계

기존에는 주로 음성인식 분야에서 사용자 적응에 관한 연구가 활발히 진행되어 왔다. 여러가지 방법 중에서 수화자 독립 모델만을 이용하여 사용자 적응을 하는 MLLR (Maximum Likelihood Linear Regression)과 MAP (Maximum A Posteriori) 방법이 널리 사용되어 왔다 [14,15]. MLLR은 사용자 적응 데이터의 관측 확률이 최대가

되도록 하는 상태 가우시안 분포 (State Gaussian Distribution)의 변환 행렬 (Transformation Matrix)을 찾아 분포의 평균과 공분산을 변화시킨다. 이 때 각각의 상태에 할당되는 적응 데이터의 개수가 충분치 못하기 때문에 MLLR에서는 평균의 모델 파라미터 공간 상에서 거리가 가까운 상태들의 집합에 대해 동일한 변환 행렬을 가진다고 가정한다. c 라고 하는 집단 (Cluster)에 R 개의 상태가 속해 있을 때 각각의 상태의 모델 파라미터 ξ_{k,m_r} 은 식 (1)과 같이 동일한 변환 행렬 $W_{k,c}$ 에 의해 적응된다. 이 때 k 는 점진적인 적응 과정에서의 시간을 나타내는 지표 (Index)이고 w 는 고려되는 단어에 대한 지표이다. ξ 는 오프셋을 포함한 확장된 평균 벡터 (Extended Mean Vector)를 의미한다.

$$\hat{\mu}_{k,w,m_r} = W_{k,c} \xi_{k,m_r}, \quad r = 1, 2, \dots, R \quad (1)$$

MAP에서는 식 (2), (3), (4)와 같이 이전 모델 파라미터 $\mu_{k,w,i}$ 와 현재 관측된 데이터로부터 얻은 파라미터 $\bar{\mu}_{k,w,i}$ 의 가중합을 통해 모델 파라미터를 얻는다.

$$\hat{\mu}_{k,w,i} = \frac{N_{k,w,i}}{N_{k,w,i} + \tau} \bar{\mu}_{k,w,i} + \frac{\tau}{N_{k,w,i} + \tau} \mu_{k,w,i} \quad (2)$$

$$\bar{\mu}_{k,w,i} = \frac{\sum_{t=1}^{T_k} \Pr(w,i | o_{k,t}) o_{k,t}}{N_{k,w,i}} \quad (3)$$

$$N_{k,w,i} = \sum_{t=1}^{T_k} \Pr(w,i | o_{k,t}) \quad (4)$$

여기서 w 는 단어의 지표, i 는 상태의 지표, k 는 점진적 적응 과정의 시간 지표를 의미한다. 그리고 $o_{k,t}$ 는 k 번째 적응 과정에서 시간 t 에서의 특징벡터를 나타낸다. 설계 파라미터인 τ 는 현재 주어진 적응 데이터로부터 추출한 파라미터와 이전 모델 파라미터 사이의 결합 강도를 조절한다. MAP에서도 MLLR과 같이 적응 데이터가 적은 경우에는 계층적 우선 순위를 이용하여 관측되지 않은 혹은 관측 데이터가 매우 적은 상태의 파라미터를 적응시킨다. MLLR과 MAP에서 모두 비관측된 상태에 대해 적응을 시키지만 이러한 방법은 모두 모델 파라미터 공간 상에서 거리가 유사한 상태는 동일한 혹은 유사한 적응 특성을 보인다는 가정에 기반을 두고 있다. 이러한 가정은 음성인식에는 대체적으로 잘 적용이 되지만 체스처 분야에 속하는 수화 인식에는 이러한 가정이 맞지 않는 것을 관찰하였다.

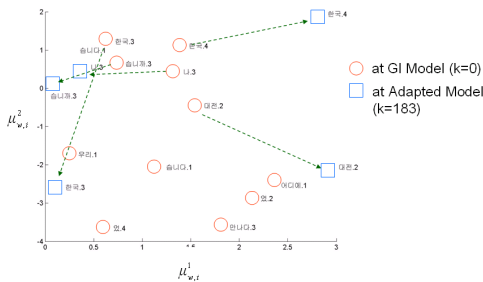


그림 2. 모델 파라미터 공간에서 유사한 위치에 있는 state들의 서로 다른 적응의 방향

그림 2는 평균 벡터 공간에서 유사한 위치에 있는 여러 상태들이 적응 전과 후에 서로 다른 적응 방향을 가진다는 것을

보인다.

점진적인 사용자 적응 과정에서 관측되지 않는 상태들은 항상 존재한다. 적응 데이터는 사용자가 인식 시스템을 사용할 때 입력되는 하나의 수화 문장이고 주어진 수화 문장에 포함되어 있는 단어는 한정적이기 때문이다. 포함된 단어에 속한 상태들만 적응의 대상으로 하는 경우에는 모든 단어에 대한 적응을 한 것이 아니기 때문에 사용자의 특성이 부분적으로만 반영이 된다. 이 경우 단어 사이의 구분도가 나빠질 수 있기 때문에 인식률의 향상이 매우 느리게 진행된다. 그림 3은 MAP을 이용하여 관측된 단어의 상태들에 대해서만 적응을 시키는 경우 인식률의 변화를 보여준다.

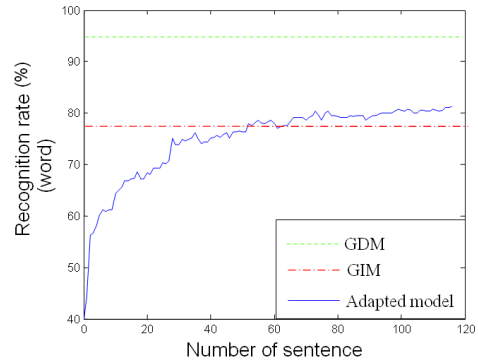


그림 3. 관측된 상태에 대해서만 적응을 수행한 경우

본 논문에서는 이와 같이 점진적인 사용자 적응 과정에서 발생하는 비관측된 상태의 모델 파라미터를 수정하는 문제를 다루고, 이를 위해 다음 절에서 동작 유사도와 변화 추이를 이용한 추정 방법에 대해서 서술한다.

4. 제안하는 사용자 적응 방법

4.1 동작 유사도에 의한 추정

수화에서 나타나는 손 운동은 직선과 원 운동과 같은 기본 운동의 조합으로 표현된다. 실제 청각 장애인들의 수화 동작을 관찰한 결과 동작이 유사한 수화의 경우 적응되는 특성도 유사하다는 것을 발견하였다. 다시 말해 동작이 유사한 경우 해당 동작의 길이와 속도 변화의 방향이 유사하다는 사실을 관찰하였다. 여기서 동작은 오른손의 움직임 궤적으로 표현하였다. 왼손은 오른손의 보조 수단으로 사용되고 고려된 단어에서 왼손만 단독으로 수화를 표현하는 경우는 없었기 때문에 제외하였다. 그림 4에서는 이러한 손 운동 궤적의 예를 보여주고 있다. 이러한 관찰을 바탕으로 관측되지 않은 상태의 파라미터는 동작 유사도가 높은 다른 관측된 상태의 데이터를 바탕으로 적응하였다. 길이가 서로 다른 동작 사이의 유사도는 Dynamic Time Warping 방법을 이용하여 측정하였다[16].

특정 상태 S 에 대해 학습 데이터에서 추출한 운동 궤적 $A_i (i=1, 2, \dots, N_1)$ 가 있고 다른 상태 S^* 에 속하는 운동 궤적 $B_j (j=1, 2, \dots, N_2)$ 이 있을 때 두 상태 간의 동작 유사도 $Sim(S, S^*)$ 는 식 (5)와 같이 모든 가능한 궤적의 쌍에 대해 거리 $DTW(A_i, B_j)$ 를 계산하여 정한다.

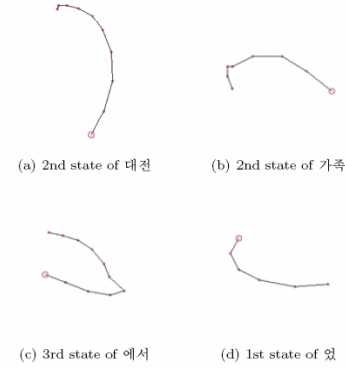


그림 4. 각 상태에 대한 손 운동 궤적의 예

$$Sim(S, S^*) = \frac{1}{\frac{1}{N} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} DTW(A_i, B_j) + 1} \quad (5)$$

모든 상태들 간의 동작 유사도를 계산한 후 특정 상태 S가 현재의 적응 데이터에서 관측되지 않는 경우, S와의 동작 유사도가 문턱값을 넘는 관측된 상태 중에서 가장 유사도가 높은 S*를 정하고 S*에 속한 적응 데이터를 이용하여 MAP을 기반으로 S의 파라미터를 적응시킨다.

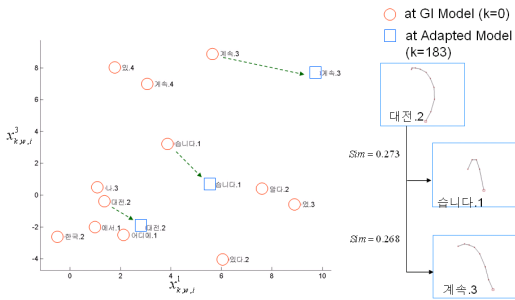


그림 5. 동작 유사도가 높은 상태들은 유사한 적응 방향을 가진다.

실험을 통해 동작 유사도가 높은 상태들은 유사한 적응 방향을 가진다는 것을 확인하였다. 그림 5에서 독립 모델에서의 상태들의 위치는 원으로, 183개의 적응 데이터를 사용하여 적응된 상태들의 위치는 사각형으로 표시하였다. 화살표는 각 상태들의 적응 방향을 나타낸다. 그림 5에서 볼 수 있듯이 “대전”의 두 번째 상태의 적응 방향은 동작 유사도가 높은 “습니다”의 첫 번째 상태와 “계속”의 세 번째 상태의 적응 방향과 유사하다.

4.2 변화 추이를 이용한 추정

일상 생활에서도 어떠한 사물의 특성이 변화하는 추이를 관찰할 수 있을 때 앞으로의 변화에 대해 추정할 수 있다. 즉 K개의 이전 값 x_1, x_2, \dots, x_K 을 알고 있을 때 K+1번째의 값 x_{K+1} 은 식 (6)과 같이 이전 값의 함수 f 로 표현될 수 있다.

$$x_{K+1} = f(x_1, x_2, \dots, x_K) \quad (6)$$

점진적인 사용자 적응 과정에서도 각각의 상태들의 적응 추이를 관찰할 수 있기 때문에 비관측된 상태에 대해서는 추정을 이용하여 모델 파라미터를 적응시킬 수 있다. 이 경우 x_k 는 k번째 시점에서의 평균이나 분산을 의미한다. 그림 6은 “나”의 두 번째 상태의 2번째 특징값에 대한 적응 추이를 보여주고 있다. 상태가 모든 시점에서 관측되는 상태인 것은 아니기 때문에 관측되는 경우는 *표시로 표현하였다. 비관측된 상태가 되는 경우에는 그림 6에서와 같이 이전 적응에서의 변화 추이를 바탕으로 현재의 모델 파라미터 값을 추정할 수 있다.

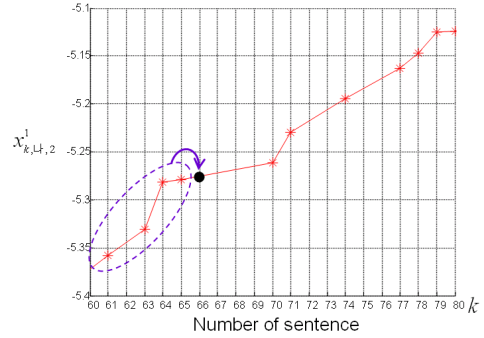


그림 6. 모델 파라미터의 변화 추이와 추정

관찰을 통해서 이러한 모델 파라미터의 변화 추이는 대체적으로 갑작스러운 변화가 없이 완만하게 변화하는 특성을 보였다. 이와 같이 일관되게 변화하는 경우에 대해 2차 랜덤 마르코프 과정 (Random Markov Process)을 이용하여 추정할 수 있다[17]. N차원의 상태 벡터 X가 있을 때 추정을 위한 2차 동적 방정식은 식 (7)과 같이 표현된다.

$$\begin{bmatrix} X_n - \bar{X} \\ X_{n+1} - \bar{X} \end{bmatrix} = \begin{bmatrix} 0 & I \\ A_0 & A_1 \end{bmatrix} \begin{bmatrix} X_{n-1} - \bar{X} \\ X_n - \bar{X} \end{bmatrix} + \begin{bmatrix} 0 \\ Bw_n \end{bmatrix} \quad (7)$$

$$\bar{X} = \frac{\sum_{i=1}^N X_i}{N}$$

여기서 w_n 은 브라운 운동 (Brownian Motion)을 나타낸다. 위 식을 정리하면 다음 식 (8)과 같다.

$$X_{n+1} = A_0 X_{n-1} + A_1 X_n + (I - A_0 - A_1) \bar{X} + Bw_n \quad (8)$$

[18]에서는 MLE (Maximum Likelihood Estimation)을 이용하여 식 (9)와 같이 A_0, A_1 을 얻었다.

$$\begin{aligned} S_{20} - A_0 S_{00} - A_1 S_{10} &= 0 \\ S_{21} - A_0 S_{01} - A_1 S_{11} &= 0 \\ S_{ij} &= \sum_{n=1}^{K-2} X_{(n-1)+i} X_{n-1+j}^T \quad i, j = 0, 1, 2 \end{aligned} \quad (9)$$

파라미터의 변화 추이가 일관된 경우에는 그림 7과 같이 추정의 정확도가 높은 반면, 변화 추이가 일관되지 않은 경우에는 그림 8에서와 같이 오차를 수반하여 추정의 정확도가 낮아진다. 이와 같이 오차가 크게 발생하는 경우 모델 파라미터의 적응이 부정확하게 이루어지기 때문에 매우 위험하다. 이러한 오차를 반영하는 것을 막기 위해 추정에 대한 신뢰도를 측정하였다.

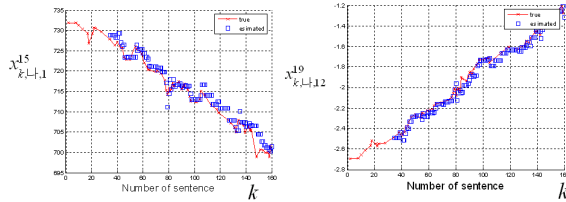


그림 7 변화 추이가 일관된 경우

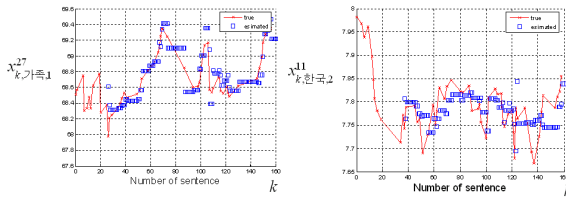


그림 8. 변화 추이가 일관되지 않은 경우

현재의 추정에 대한 신뢰도는 가장 최근에 수행했던 추정 정확도에 근거하여 결정하였다. 다음 식 (10)과 같이 가장 최근의 N개의 추정에 대한 평균 오차를 구한 후 k시점에서의 추정의 신뢰도는 식 (11)과 같이 지수 함수로 표현하였다.

$$e_{k,w,i}^{j-} = \frac{\sum_{i=1}^N (x_{k-n,w,i}^{j,true} - x_{k-n,w,i}^{j,predicted})}{N} \quad (10)$$

$$\beta_{k,w,i}^j = e^{-\rho_j e_{k,w,i}^{j-}} \quad (11)$$

감쇄 파라미터인 ρ_j 를 결정하기 위해 모든 데이터에 대한 추정을 수행한 후 발생한 오차를 계산하였다. 오차의 분포가 가우시안이라는 가정 하에 10%의 특이값 (Outlier)을 제외한 최대 오차 e_j^* 를 정하였다. 추정 오차가 e_j^* 가 되었을 때 신뢰도가 0.1이 되도록 하는 ρ_j 를 정하였다.

4.3 제안한 두 가지 방법의 결합

앞 절에서 비관측된 상태의 모델 파라미터를 추정하기 위해 제안한 두 가지 방법은 각각 장단점을 가지고 있다. 동작 유사도에 의한 추정은 적응 데이터가 적은 적응의 초기 단계에서도 이용할 수 있다는 장점이 있는 반면 때에 따라서는 동작 유사도가 높은 모든 상태가 관측되지 않아 이용할 수 없다는 단점이 있다. 적응 추이에 의한 추정의 경우에는 추이가 일관되지 않은 경우에는 신뢰도가 낮아진다는 단점을 가지고 있다. 따라서 각각의 방법을 단독으로 사용하기 보다는 양자를 결합한 방법으로 서로의 단점을 보완하는 것이 필요하다. 이를 고려한 방법의 절차는 다음과 같다.

단계 1. 모든 관측된 상태의 모델 파라미터를 MAP을 이용하여

여 적응시킨다.

단계 2. 모든 비관측된 상태에 대해

단계 2-1. 동작 유사도에 의한 추정치 $x_{k,w,i}^{j,motion}$ 를 구한다.

단계 2-2. 적응 추이에 의한 추정치 $x_{k,w,i}^{j,history}$ 를 구한다.

단계 2-3. 적응 추이에 의한 추정치의 신뢰도를 계산한 후 이를 이용하여 가중합을 한다.

$$x_{k,w,i}^{j,estimation} = (1 - \beta_{k,w,i}^j) x_{k,w,i}^{j,motion} + \beta_{k,w,i}^j x_{k,w,i}^{j,history}$$

단계 3. 위에서 계산한 가중합된 추정치를 이용하여 비관측된 상태의 모델 파라미터를 적응한다.

$$x_{k+1,w,i}^j = \alpha x_{k,w,i}^j + (1 - \alpha) x_{k,w,i}^{j,estimation}$$

이 때 단계 3에서 k번째의 모델 파라미터와 가중합된 추정치를 결합하는 파라미터 α 는 실험을 통해 인식률의 증가가 가장 최대가 되는 0.65로 설정하였다.

5. 실험 결과

제안한 방법으로 구축한 수화 데이터베이스에 대해 실험을 수행하였다. 70%를 적응 데이터로 사용하고 나머지 30%를 테스트 데이터로 사용하였다. 평균적인 성능을 구하기 위해 동일한 과정을 30번 반복하였다. 인식률은 단어 단위와 문장 단위 모두 측정하였고 문장 인식에 있어서는 문법과 같은 다른 정보는 이용하지 않았다. 비관측된 상태의 파라미터를 적응시키지 않는 MAP방법과 동작 유사도와 적응 추이를 이용하여 비관측된 상태의 파라미터를 적응하는 제안한 방법을 비교하였다.

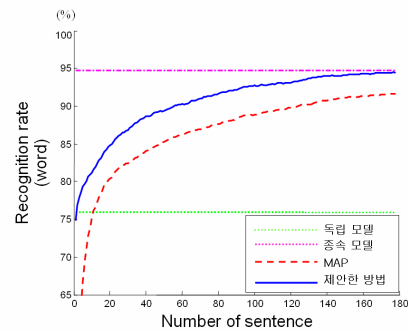


그림 9. 사용자 #1에 대한 적응

그림 9는 사용자 #1에 대한 적응의 결과를 보인다. 먼저 독립모델을 사용했을 때의 인식률과 종속 모델을 사용했을 때의 인식률이 기준이 된다. 적응 방법의 우수성은 독립 모델의 인식률에서 출발하여 얼마나 빠르게 종속 모델의 인식률까지 도달할 수 있는가로 이해할 수 있다. 비관측된 상태를 다루지 않는 MAP의 경우에는 초반에 오히려 인식률이 독립 모델보다 낮기도 하고 인식률의 증가도 느리다. 반면에 제안한 방법은 비관측된 상태의 모델 파라미터를 추정에 의해 적응시키기 때문에 인식률의 증가가 MAP에 비해 빠르다는 사실을 알 수 있다. 그림 10~그림 16은 다른 사용자에 대한 적응 결과를 보여준다.

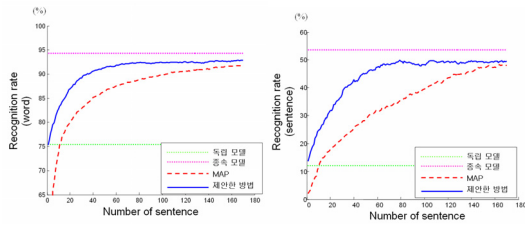


그림 10. 사용자 #2에 대한 적응

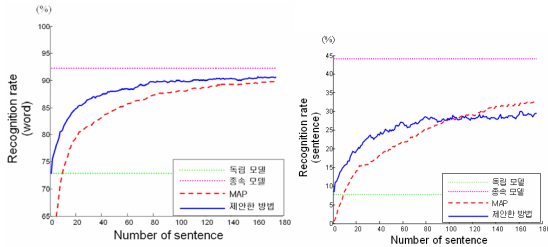


그림 11. 사용자 #3에 대한 적응

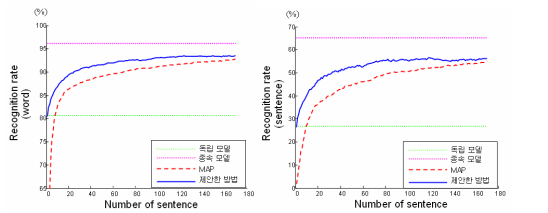


그림 12. 사용자 #4에 대한 적응

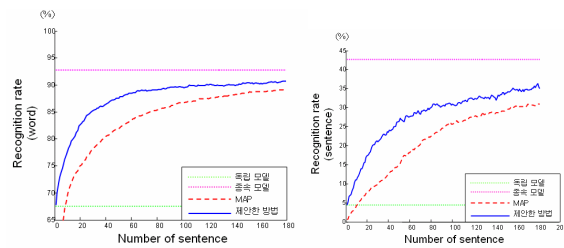


그림 13. 사용자 #5에 대한 적응

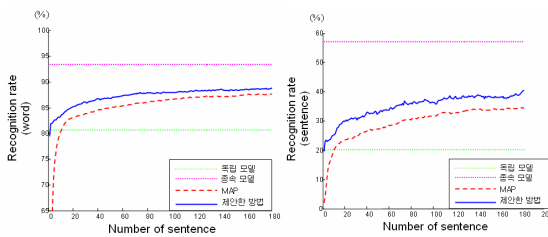


그림 14. 사용자 #6에 대한 적응

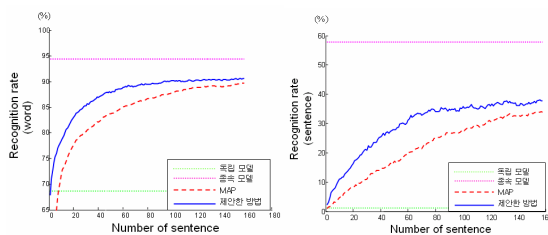


그림 15. 사용자 #7에 대한 적응

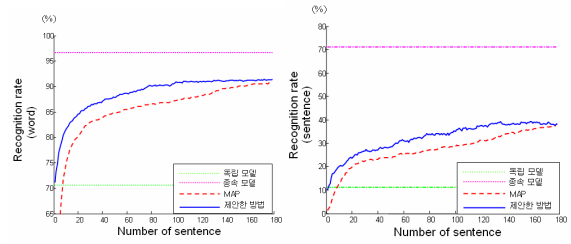


그림 16. 사용자 #8에 대한 적응

6. 결론 및 추후 과제

본 논문에서는 기존의 수화 인식에서 다루지 않았던 사용자 적응 문제를 다루었다. 점진적으로 이루어지는 사용자 적응에서 발생하는 비관측된 상태를 효과적으로 적응시키기 위해 동작 유사도에 의한 추정과 적응 추이에 의한 추정 방법을 제안하였고 이들을 결합한 형태 또한 제안하였다. 실제 청각 장애인들로부터 획득한 수화 데이터베이스를 사용하여 제안한 방법이 비관측된 상태를 다루지 않는 기존의 방법보다 빠르게 인식률을 향상시키는 것을 보였다. 앞으로의 추후 과제로는 제안한 방법에서 사용한 여러 개의 설계 파라미터를 효과적으로 정하는 것과 좀더 확장된 수화 데이터베이스를 이용하여 제안한 방법의 효과를 검증하는 것이 필요하다.

Acknowledgement

본 연구는 과학기술부/한국과학재단 우수연구센터육성사업의 지원으로 수행되었음 (R11-1999-008)

참고문헌

- [1] T.Yamaguchi, M.Yoshihara, M.Akiba, M.Kuga, N.Kanazawa and K.Kamata, "Japanese Sign Language Recognition System Using Information Infrastructure," Proceedings of IEEE International Conference on Fuzzy System, pp.65-66, March 1995
- [2] W.Kadous, "Recognition of Australian Sign Language Using Instrumented Gloves," Bachelor's thesis, University of New South Wales, Oct 1995.
- [3] Jong-Sung Kim, Won Jang and Zeungnam Bien, "A Dynamic Gesture Recognition System for the Korean Sign Language (KSL)," IEEE Transactions on System, Man, and Cybernetics, Part B, vol.26, no.2, pp.354-359, April 1996
- [4] Chan-Su Lee, Zeungnam Bien, Gyu-Tae Park, Won Jang, Jong-Sung Kim and Sung-Kwon Kim, "Real-time Recognition System of Korean Sign Language Based on Elementary Components," Proceedings of the Sixth IEEE International Conference on Fuzzy Systems, vol.3, pp.1463-14168, 1997.

- [5] Thad Starner and Alex Petland, "Real-time American Sign Language Recognition from Video Using Hidden Markov Models," Proceedings of International Symposium on Computer Vision, pp.265-270, 1995.
- [8] Rung-Huei Liang, "Continuous Gesture Recognition System for Taiwanese Sign Language," Ph.D Thesis, National Taiwan University, Taiwan 1997.
- [9] Mu-Chun Su, Yu-Xiang Zhao, Hai Huang and Hsuan-Fan Chen, "A Fuzzy Rule-based Approach to Spatio-Temporal Hand Gesture Recognition," IEEE Transactions on System, Man, and Cybernetics-Part C: Applications and Reviews, vol.30, no.2, pp.276-281, May 2000.
- [10] Hirohiko Sagawa and Masaru Takeuchi, "A Method for Recognizing a Sequence of Sign Language Sentence," Proceedings of Fourth IEEE International Conference on Automatic Face and Gesture Recognition, pp.29-37, March 2000.
- [11] Britta Bauer and Hermann Hienz, "Relevant Features for Video-based Continuous Sign Language Recognition," Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition, pp.440-445, 2000.
- [12] Gaolin Fang and WenGao "A SRN/HMM System for Signer-independent Continuous Sign Language Recognition," Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition, pp.297-302, May 2002.
- [13] K. J. Bae, A Study on the Continuous Hand Gesture Recognition System for the Korean Sign Language, PhD Thesis, KAIST, 2004.
- [14] J. Gauvain and C. H. Lee. "Maximum a Posteriori Estimation for Multivariate Gaussian Mixture Observation of Markov Chains," Proceedings of IEEE Transaction on Speech and Audio, vol.2, pp.291-298, April,2004.
- [15] C. Leggetter and P. Woodland. "Maximum Likelihood Linear Regression for Speaker Adaptation of Continuous Density HMMs," Proceedings of Communication and Technology, vol.2. pp. 171-185, 1995.
- [16] Keogh, E.J. and Pazzani, M.J., "Derivative Dynamic Time Warping," Proceedings of the first International Conference on Data Mining, 2001.
- [17] Leonid Sigal, Stan Scalroff and Vassilis Athitsos, "Skin Color-Based Video Segmentation under Time-Varying Illumination," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.26, no.7, pp.862-877, 2004.