

연속된 수화 인식을 위한 자동화된 Coarticulation 검출

(Automatic Coarticulation Detection for Continuous Sign Language Recognition)

양 희 덕 [†] 이 성 환 ^{**}
(Hee-Deok Yang) (Seong-Whan Lee)

요 약 수화 적출은 연속된 손 동작에서 의미 있는 수화 단어를 검출 및 인식하는 것을 말한다. 수화는 손의 움직임과 모양의 변화가 다양하기 때문에 수화 문장에서 수화를 적출하는 것은 쉬운 문제가 아니다. 특히, 자연스러운 수화 문장에는 의미 있는 수화, 수화가 아닌 손동작이 무작위로 발생한다. 본 논문에서는 CRF(Conditional Random Field)에 기반한 적응적 임계치 모델을 제안한다. 제한된 모델은 수화 어휘집에 정의된 수화 손동작과 수화가 아닌 손동작을 구별하기 위한 적응적 임계치 역할을 수행한다. 또한, 수화 적출 및 인식의 성능 향상을 위해 손 모양 기반 수화 인증기, 짧은 수화 적출기, 부사인(subsign) 추론기를 제안된 시스템에 적용하였다. 실험 결과, 제안된 방법은 연속된 수화 동작 데이터에서 88%의 적출률, 사전에 적출된 수화 동작 데이터에서 94%의 인식률을 보였으며, 적응적 임계치 모델, 짧은 수화 적출기, 손 모양 기반 수화 인증기, 부사인 추론기를 사용하지 않은 CRF 모델은 연속된 수화 동작 데이터에서 74%의 적출률, 사전에 적출된 수화 동작 데이터에서 90%의 인식률을 보였다.

키워드 : 수화 인식, 수화 적출, Conditional Random Field, 임계치 모델

Abstract Sign language spotting is the task of detecting and recognizing the signs in a signed utterance. The difficulty of sign language spotting is that the occurrences of signs vary in both motion and shape. Moreover, the signs appear within a continuous gesture stream, interspersed with transitional movements between signs in a vocabulary and non-sign patterns(which include out-of-vocabulary signs, epentheses, and other movements that do not correspond to signs). In this paper, a novel method for designing a threshold model in a conditional random field(CRF) model is proposed. The proposed model performs an adaptive threshold for distinguishing between signs in the vocabulary and non-sign patterns. A hand appearance-based sign verification method, a short-sign detector, and a subsign reasoning method are included to further improve sign language spotting accuracy. Experimental results show that the proposed method can detect signs from continuous data with an 88% spotting rate and can recognize signs from isolated data with a 94% recognition rate, versus 74% and 90% respectively for CRFs without a threshold model, short-sign detector, subsign reasoning, and hand appearance-based sign verification.

Key words : Sign language recognition, sign language spotting, conditional random field, threshold model

- 이 논문은 2008년도 한국과학재단의 지원을 받아 수행된 연구임
(No. R01-2008-000-12034-0)
· 이 연구에 참여한 연구자는 '2단계 BK21 사업'의 지원을 받았음

[†] 정 회 원 : 조선대학교 컴퓨터공학부 교수
heedeok_yang@chosun.ac.kr

^{**} 종신회원 : 고려대학교 컴퓨터 통신공학부 교수
swlee@image.korea.ac.kr

논문접수 : 2008년 10월 6일
심사완료 : 2008년 11월 7일

Copyright©2009 한국정보과학회 : 개인 목적이나 교육 목적인 경우, 이 저작물의 전체 또는 일부에 대한 복사본 혹은 디지털 사본의 제작을 허가합니다. 이때, 사본은 상업적 수단으로 사용할 수 없으며 첫 페이지에 본 문구와 출처를 반드시 명시해야 합니다. 이외의 목적으로 복제, 배포, 출판, 전송 등 모든 유형의 사용행위를 하는 경우에 대하여는 사전에 허가를 얻고 비용을 지불해야 합니다.
정보과학회논문지 : 소프트웨어 및 응용 제36권 제1호(2009.1)

1. 서 론

수화 인식 시스템은 사용자가 이용하기 편리해야 한다[1-3]. 특히, 수화 인식 시스템은 사용자가 칼라 장갑, 데이터 글러브 등의 특별한 입력 장치의 이용 없이 사용할 수 있도록 구성되어야 한다. 하지만, 기존의 많은 연구는 사용자에게 칼라 장갑과 같은 특수한 입력 장치의 사용을 요구하고 있다. 이러한 장치들을 이용하면 수화 인식 시스템은 손의 모양과 움직임 정보를 쉽고, 정확하게 추출할 수 있는 장점은 있지만, 사용자에게 수화를 수행하는 동안 불편함을 주는 단점을 갖게 된다.

그림 1에서와 같이 수화 인식 시스템에서 어려운 문제는 연속된 손동작에서 수화의 시작점과 끝점을 검출하는 것과 검출된 손동작을 정의된 수화 어휘집에서 인식하는 것으로 나누어진다. 수화는 손의 움직임 정보와 모양 정보의 조합으로 표현되기 때문에 연속된 수화 문장에서 수화를 정확히 적출하기는 쉽지가 않다.

기존 연구에서는 HMM(Hidden Markov Model) 또는 CRF를 적용하여 손의 움직임 정보를 성공적으로 분석하였다. HMM이나 CRF를 이용한 방법은 수화와 비수화(수화와 수화 사이에 발생하는 수화가 아닌 손동작,

인식할 수 있는 수화 어휘집에 정의되지 않은 수화)를 구별하기 위해서 고정된 임계치를 사용하였다. 그림 2(a)에서와 같이 임계치 T_a 가 선택되면 수화 “AND”와 “BIG”이 모두 검출되지 않고, 임계치 T_b 가 선택되면 수화 “AND”와 “BIG”이 모두 검출된다. 예에서 볼 수 있듯이 수화를 적출하기 위한 최적의 임계치를 결정하는 것은 쉬운 문제가 아니다. 이러한 문제를 해결하기 위해서 음성인식 연구 분야에서는 HMM에 기반한 필러 모델(Filler Model) 또는 가비지 모델(Garbage Model)을 이용하였다. 하지만, 비수화 동작은 무제한이기 때문에,

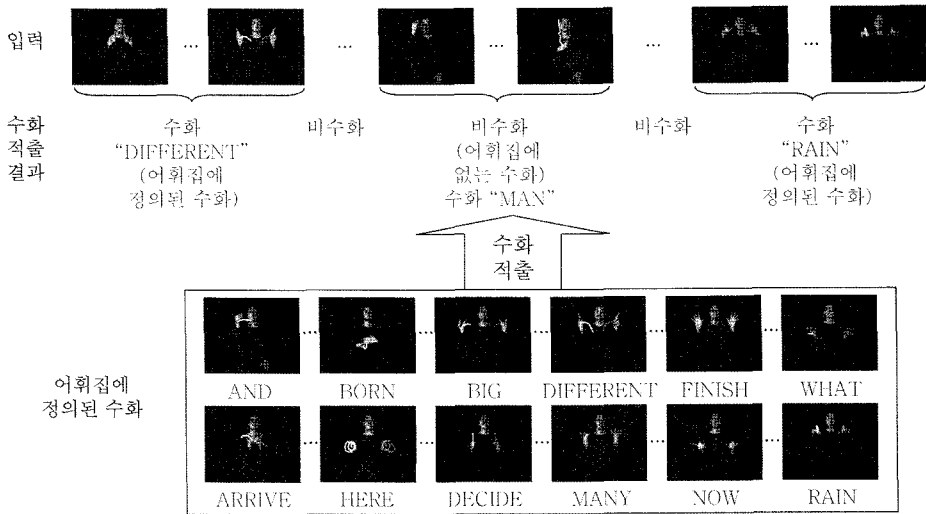
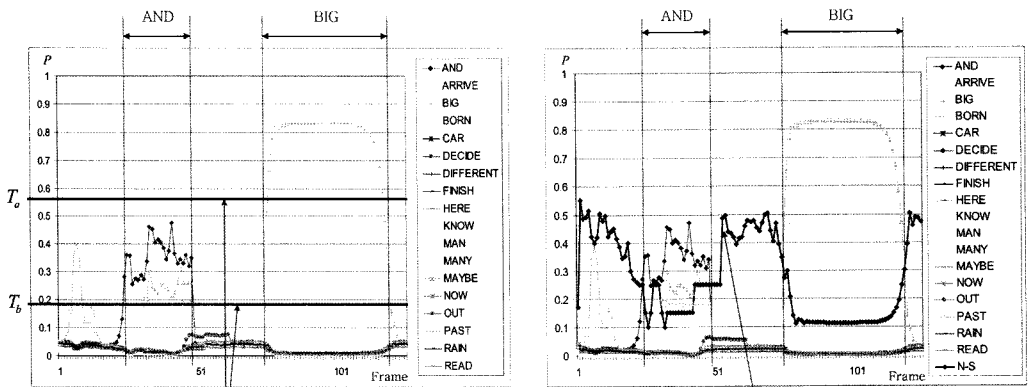


그림 1 수화 적출 시스템의 예: 수화 적출은 수화의 시작점과 끝점을 검출하는 것과 찾아진 손 동작을 정의된 수화 어휘집에서 인식하는 것으로 구성됨



(a) 고정된 임계치 사용

(b) 적응적 임계치 사용

그림 2 CRF를 이용한 적응적 임계치 모델 생성: 비수화(non-sign, N-S) 레이블을 추가하고 이의 확률값을 임계치값으로 이용

필러 모델 또는 가비지 모델을 학습하기 위한 비수화 동작을 수집하는 것은 불가능하다. 필러 모델의 단점을 극복하기 위해서 HMM 기반의 적응적 임계치 모델이 제안되었다[4,5]. 본 논문에서는 CRF 모델에 비수화를 대표할 수 있는 레이블을 추가하여 CRF 기반의 적응적 임계치 모델을 제안한다. 제안된 CRF 기반의 적응적 임계치 모델은 그림 2(b)에서와 같이 추가된 비수화 레이블의 확률값이 임계치 역할을 수행하여 수화와 비수화를 구별할 수 있다. CRF 기반의 적응적 임계치 모델 생성을 위해서, 일차적으로 전통적인 CRF 모델을 이용하여 비수화 레이블이 없이 수화 인식을 위한 모델을 생성한다. 생성된 CRF 모델에 비수화 레이블을 추가하고, 추가된 비수화 레이블의 전이, 상태 특징 함수의 가중치값을 비수화 레이블 없이 생성된 전통적인 CRF 모델의 전이, 상태 특징 함수의 가중치값을 이용하여 생성한다. 이렇게 비수화 레이블을 위한 전이, 상태 특징 함수의 가중치값을 비수화 레이블을 위한 학습 데이터가 없이 생성할 수 있기 때문에, 제안된 방법은 수화 어휘집에 정의된 수화와 수화 어휘집에 정의되지 않은 모든 손동작을 구별할 수 있다.

수화 인식에서 손의 움직임 정보는 전역적 특징, 손의 모양 정보는 지역적 특징을 표현한다[6]. 위에서 언급한 방법들은 손의 움직임 정보, 즉 전역적 정보만을 분석할 수 있는 방법이다. 하지만, 비슷한 손의 움직임 정보를 갖고 있으면서 손의 모양이 다른 수화가 많이 존재하기 때문에, 본 논문에서는 손 모양 인식기를 동시에 적용한다.

또한, 수화 인식기의 성능 향상을 위해서, 짧은 수화 적출기와 부사인 추론기를 제안된 시스템에 적용한다. 짧은 수화 적출기는 평균 수화 수행 시간보다 짧은 수화의 인식 성능을 향상시킨다. 짧은 수화는 긴 수화에

비해 수화로 인식될 수 있는 시간이 상대적으로 적기 때문에 긴 수화에 비해 인식하기가 어렵다[7]. 마지막으로 서로 다른 수화끼리 비슷한 움직임을 공유할 때, 다른 수화로 인식되는 것을 방지하기 위해서 부사인 추론기를 적용한다[7]. Alon 등이 DTW(Dynamic Time Warping)에 적용한 방법을 본 연구에서는 2계층 CRF를 이용하여 적용한다[7].

그림 3은 제안된 수화 적출 방법의 구성도를 보여준다. 입력 영상으로부터 손과 얼굴의 위치를 검출하고, 검출된 손과 얼굴의 위치 정보로부터 특징을 추출한다. 마지막으로 제안된 임계치 CRF 모델, 짧은 수화 적출기, 손 모양 인식기, 부사인 추론기를 이용하여 수화를 적출한다.

2. 관련 연구

패턴 적출은 입력된 신호에서 의미 있는 부분을 검출하고 정의된 데이터 집합에서 검출된 데이터를 인식하는 것이다. 수화 적출은 패턴 적출의 하나의 예이다[7].

HMM은 시-공간적으로 변화가 많은 자료를 처리하기 위해서 널리 이용되고 있는 방법이다. 최근에는 DNA 서열 분석 등 많은 분야에 적용이 되고 있다. McCallum 등은 MEMM(Maximum Entropy Markov Model)을 제안하였다[8]. MEMM은 HMM에서 나누어져 있던 관측치 확률과 전이 확률을 하나로 결합하였다[8]. 따라서, MEMM에서 관측치는 현재와 이전의 상태에 의존적이다. 하지만, MEMM은 레이블 편향 문제가 발생한다. 이는 MEMM이 지역적 정규화 작업을 수행하기 때문이다. 이를 해결하기 위해서 Lafferty 등은 CRF 모델을 제안하였다[9]. Morcency 등은 CRF 모델에 은닉 변수를 추가하여 CRF의 장점과 HMM의 장점

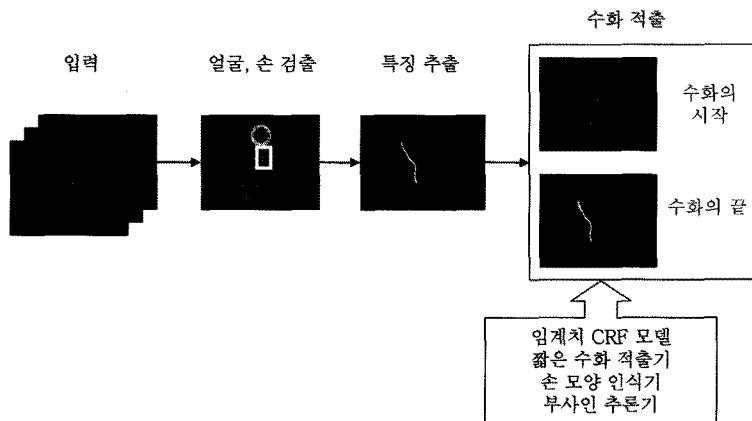


그림 3 제안된 수화 적출 방법의 구성도

을 갖는 HCRF(Hidden Conditional Random Field) 모델을 제안하였다[10,11].

수화 인식 방법은 데이터 입력 방법에 따라 마커, 데이터 글러브 등을 사용하는 장치 기반 방법과 카메라를 이용하는 비전 기반 방법으로 나누어진다[1,12].

2.1 장치 기반 수화 인식 방법

Braffort는 HMM을 이용하여 프랑스 수화를 인식할 수 있는 시스템을 제안하였다[2]. 데이터 글러브를 이용하여 손의 모양과 위치 정보를 추출하였다. 7 단어로 구성된 수화 어휘집을 이용하여 수화 인식을 수행하였으며, 96%의 수화 인식률을 얻었다. 실험은 연속된 손동작에서 수행되었다.

Gao 등은 DTW를 이용하여 중국어 수화를 인식할 수 있는 시스템을 제안하였다[13]. 데이터 글러브와 3개의 위치 추적기를 이용하여 손의 위치 정보를 추출하였고, 비수화 부분을 모델링 하였다. 5,113개의 단어로 구성된 수화 어휘집을 이용하여 수화 인식을 수행하였으며, 90.8%의 수화 적출률을 보였다. 실험은 연속된 손동작에서 수화 적출 및 인식을 동시에 수행하였다.

Volger와 Metaxas는 HMM을 이용하여 미국 수화(American Sign Language, ASL)를 인식할 수 있는 방법을 제안하였다[14]. 손의 위치 정보를 추출하기 위해서 전자 장치를 사용하였다. 22개의 수화 단어를 이용하여 생성한 99개의 문장을 이용하여 실험을 수행하였다. 단어 단위의 인식에서는 94.5%의 인식률을 문장 단위 인식에서는 84.5%의 수화 인식률을 보였다.

2.2 비전 기반 수화 인식 방법

Starner 등은 HMM 기반의 ASL 인식 시스템을 개발하였다[15]. 사용자가 착용한 모자에 카메라를 부착하여 영상을 획득한 방법과 책상에 장착된 카메라를 이용하여 영상을 획득한 방법으로 실험을 수행하였다. 실험 결과, 각각 92%, 98%의 수화 적출률을 보였다.

Holden 등은 HMM 기반의 ASL 인식 방법을 제안하였다[16]. 43개의 단어로 구성된 수화 어휘집을 이용하여 97%의 인식 성능을 보였다. 실험은 연속된 손동작에서 수화 적출 및 인식을 동시에 수행하였다.

Yang 등은 신경망을 이용하여 ASL 인식 기술을 제안하였다[17]. 40개의 단어로 구성된 수화 어휘집에서 96.2%의 수화 인식률을 보였다. 이 실험에서는 사전에 적출된 수화를 인식하였다.

Yang 등은 CRF 기반 ASL 인식 기술을 제안하였다[18]. 학습 단계에서는 수화 문장으로부터 핵심 프레임을 추출하였고, 추출된 핵심 프레임을 수화와 비수화로 분류하였다. 인식 단계에서는 핵심 프레임을 추출하고 이를 수화, 비수화로 인식하였다. 39개의 단어로 구성된 수화 어휘집에서 85% 이상의 수화 적출률을 보였다.

3. Conditional Random Field 기반 임계치 모델

CRF는 조건부 확률 방법을 이용하여 순차적 데이터의 레이블링을 수행하기 위해서 사용된다[9,19]. CRF는 관측열 x 에 대한 레이블열 y 의 확률을 전이, 상태 특정 함수들의 곱으로 계산한다. 각각의 함수의 곱은 아래 수식으로 표현된다[9,19].

$$\exp\left(\sum_v \lambda_v t_v(y_{i-1}, y_i, x, i) + \sum_m \mu_m s_m(y_i, x, i)\right) \quad (1)$$

여기서, $t_v(y_{i-1}, y_i, x, i)$ 는 관측열 x 의 $i-1$ 과 i 위치에서의 전이 특정 함수, $s_m(y_i, x, i)$ 는 관측열 x 의 i 위치에서의 상태 특정 함수, y_{i-1}, y_i 는 관측열 x 의 $i-1$ 과 i 위치에서의 레이블, 그리고, λ_v, μ_m 는 각각 전이, 상태 특정 함수의 가중치를 나타낸다.

상태 특정 함수는 특징의 특정 레이블에서의 발생 여부를 결정하고, 전이 특정 함수는 두 레이블 사이에서의 특징 발생 여부를 결정한다.

수식 (1)로부터 관측열 x 에 대한 레이블열 y 의 확률은 아래와 같다.

$$p_\theta(y|x) = \frac{1}{Z_\theta(x)} \prod_{i=1}^n F_\theta(y_{i-1}, y_i, x, i) \quad (2)$$

여기서, 파라미터 $\theta = (\lambda_1, \lambda_2, \dots, \lambda_{N_t}; \mu_1, \mu_2, \dots, \mu_{N_s})$, N_t 는 전이 특정 함수의 개수, N_s 는 상태 특정 함수의 개수, n 은 관측열 x 의 길이, $Z_\theta(x)$ 는 정규화값

$$Z_\theta(x) = \sum_y \exp\left(\sum_{i=1}^n F_\theta(y_{i-1}, y_i, x, i)\right) \quad (3)$$

여기서,

$$F_\theta(y_{i-1}, y_i, x, i) = \sum_j \lambda_j t_j(y_{i-1}, y_i, x, i) + \sum_k \mu_k s_k(y_i, x, i) \quad (4)$$

CRF는 파라미터 학습을 위해서 최대 엔트로피 방법을 이용하고 있다. 최대 엔트로피 방법은 학습 데이터의 로그 우도 함수가 최대가 되는 함수를 선택한다. CRF를 위한 로그 우도 함수는 아래와 같다.

$$L(\theta) = \sum_{i=1}^T \left(\sum_{i=1}^n F_\theta(y_{i-1}, y_i, x', i) - \log Z_\theta(x') \right) \quad (5)$$

여기서, T 는 학습 데이터의 개수이고, x' 는 학습 데이터의 관측열, y' 는 x' 에 대응되는 레이블열이다.

일차적으로, 위의 수식을 이용하여 비수화에 대한 정보가 없는 학습 데이터를 이용하여 전통적인 CRF 모델을 생성한다. 이렇게 생성된 전통적인 CRF 모델에 비수화를 위한 레이블을 추가하고 비수화 레이블의 전이,

상태 특징 함수의 가중치값을 아래의 방법을 이용하여 생성한다.

3.1 임계치 모델의 상태 특징 함수 가중치

CRF 모델에서 상태 특징 함수의 가중치는 상태 특징 함수가 여러 레이블에서 발생하면, 이를 나타내는 가중치의 분산은 작고, 반면에 상태 특징 함수가 하나의 레이블에 편중되면 가중치의 분산이 크다. 비수화를 위한 레이블의 상태 특징 함수의 가중치를 비수화 레이블 없이 생성된 CRF 모델의 가중치의 분산값을 기반으로 생성한다.

비수화를 위한 레이블 G의 상태 특징 함수 가중치는 Dugad 등의 방법에 의해서 아래와 같이 계산된다[20].

$$\mu_m(G) = \bar{\mu}_m(G) + T_d \sqrt{\sigma_{\mu_m}} \quad (6)$$

여기서,

$$\bar{\mu}_m = \left(\sum_{k=1}^l \mu_m(Y_k) \right) / l \quad (7)$$

여기서, l 은 수화 어휘집에 정의된 수화의 개수이고, Y_k 는 레이블이다.

3.2 임계치 모델의 전이 특징 함수 가중치

수화 문장에서 수화 단어보다 비수화 단어의 빈도수는 상대적으로 높다. 따라서, 비수화를 위한 레이블 G의 전이 특징 함수 가중치는 다른 레이블의 전이 특징 함수 가중치보다 크다. 비수화를 위한 레이블 G의 전이 특징 함수 가중치는 아래와 같다.

$$\lambda_v(G, G) = \arg \max_{k=1, \dots, l} \lambda_v(Y_k, Y_k) + \frac{C}{\bar{N}_{ff}} \quad (8)$$

여기서, $C = \sum_{k=1}^l \sum_{m=1}^{N_s} \mu_m(Y_k)$ 이고, \bar{N}_{ff} 는 상태 특징 함수의 가중치 값이 0보다 큰 특징 함수의 평균값이다.

특정 수화에서 비수화로 전이될 때는 한 프레임만 사용된다. 따라서, 수화 어휘집에 정의된 수화에서 비수화로 전이되는 전이 특징 함수 가중치는 아래와 같이 계산된다.

$$\forall_{k \in \{1, \dots, l\}} \lambda_v(Y_k, G) = \frac{\lambda_v(Y_k, Y_k)}{l} \quad (9)$$

또한, 비수화에서 특정 수화로 전이될 때도 한 프레임만 사용된다. 따라서, 비수화에서 수화 어휘집에 정의된 수화로 전이되는 전이 특징 함수 가중치는 아래와 같이 계산된다.

$$\forall_{k \in \{1, \dots, l\}} \lambda_v(G, Y_k) = \frac{\lambda_v(G, G)}{l} \quad (10)$$

4. 수화 적출 시스템

그림 4는 제안된 수화 적출 시스템의 흐름도를 보여

준다. 그림에서 볼 수 있듯이 제안된 시스템은 2계층의 CRF 모델로 구성된다. 1단계에서는 3장에서 제안된 CRF 기반의 적응적 임계치 모델을 이용하여 수화와 비수화를 구분하고, 2단계에서는 전통적인 CRF 모델을 이용하여 부사인 추론을 수행한다. 그리고, 2계층의 CRF 모델에서 검출된 수화의 손 모양 인식을 수행한다. 마지막으로, 한손 수화인지, 양손 수화 인지를 판단한다.

1단계의 CFR 임계치 모델과 2단계의 CRF 모델은 손의 움직임 정보를 분석한다. 손의 움직임 정보 분석을 위해서 그림 5와 같이 손과 얼굴을 검출하고, 손과 얼굴의 방향과 거리를 기준으로 표 1에 보이는 좌와 같이 6개의 위치 및 움직임 기반 특징을 추출한다. 추출된 6개의 특징 정보를 이용하여 1단계의 CFR 임계치 모델과 2단계의 CRF 모델을 학습한다. CRF 학습을 위해서 윈도우 크기는 2를 사용하였으며, 10개의 상태 특징 함수를 이용하였다.

양손의 수직 평형 여부 판단은 아래와 같이 수행한다.

$$S_{TH} = \begin{cases} 1, & |d_{HL} - d_{HR}| < T_h \text{ and } d_v < T_v \\ 0, & \text{otherwise} \end{cases} \quad (11)$$

여기서, T_h 와 T_v 는 임계치값이다.

양손의 겹침 여부는 아래와 같이 판단한다.

$$O_{TH} = \begin{cases} 1, & \min\left(\frac{R_o}{H_L}, \frac{R_o}{H_R}\right) > T_o \\ 0, & \text{otherwise} \end{cases} \quad (12)$$

여기서, H_L 은 왼쪽손 영역, H_R 은 오른쪽손 영역, R_o 는 양손의 겹침 영역, 그리고, T_o 는 임계치값이다.

C_{LH} , C_{RH} 8 방향 코드북과 1개의 더미 코드로 표현된다. 더미 코드는 두 위치간의 움직임이 미세한 경우를 표현한다.

하나의 손 모양을 인식하기 위해서 하나의 SVM (Support Vector Machine)을 생성한다[21]. 제안된 시스템에서는 손의 시작점에서만 손 모양을 인식한다. 손 모양 인식을 위해서 HOG(Histograms of Gradient) 특징을 사용한다. 손 모양 인식 결과를 반영하기 위해서 아래와 같이 일정 시간 t_a 동안 인식 결과를 누적한다.

$$v = \sum_{i=t-t_a/2}^{i=t+t_a/2} C^g(y_i, L_i) \quad (13)$$

여기서, t 는 현재 프레임, t_a 는 윈도우 크기, 그리고 $C^g(y_i, L_i)$ 는

$$C^g(y_i, L_i) = \begin{cases} 1, & y_i = L_i \\ 0, & \text{otherwise} \end{cases} \quad (14)$$

여기서, y_i 는 i 번째 프레임에서 CRF를 이용한 움직임 정보 분석 결과, L_i 는 i 번째 프레임에서 SVM 인식 결과이다.

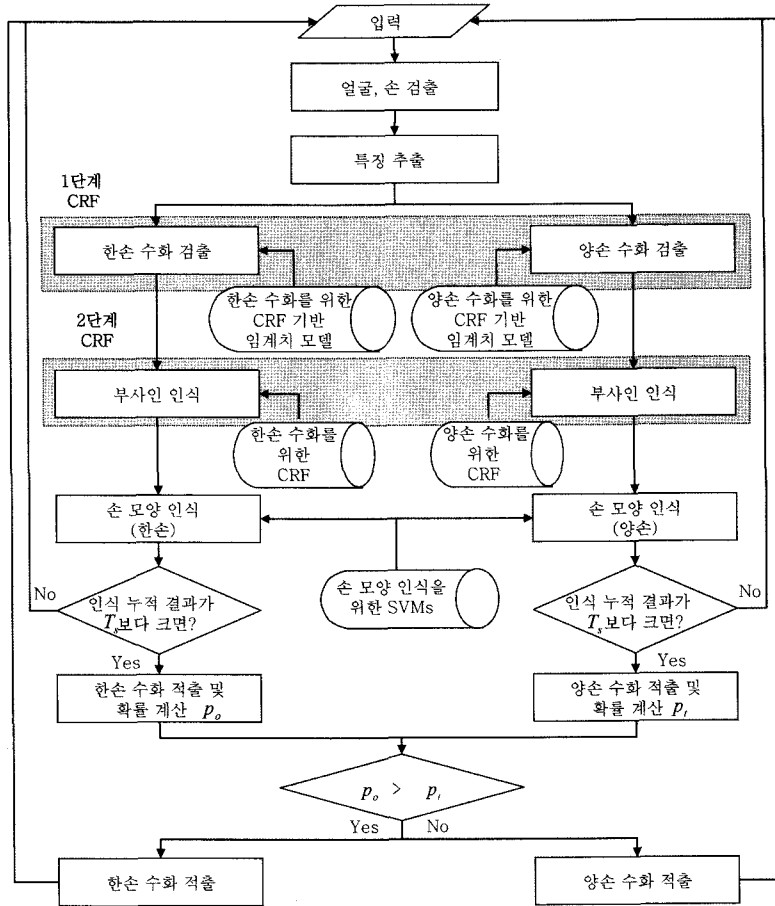


그림 4 제안된 수화 적출 시스템의 흐름도

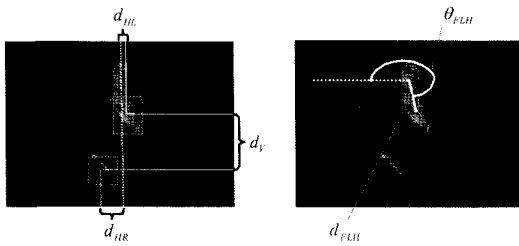


그림 5 손의 위치 정보와 각도 정보

표 1 2개의 움직임 기반 특징과 4개의 위치 기반 특징

특징	의미
P_{LH}	왼쪽손의 위치
P_{RH}	오른쪽손의 위치
C_{LH}	왼쪽손의 이동 방향
C_{RH}	오른쪽손의 이동 방향
S_{TH}	양손의 수직 평형 여부
O_{TH}	양손의 겹침 여부

5. 실험 및 결과 분석

5.1 실험 환경

스튜디오 환경에서 수집된 ASL 데이터베이스를 이용하였다. 데이터베이스는 48단어의 수화로 구성되었으며, 청각장애인이 수화 문장을 수행하였다. 수화 수행자는 학습 데이터를 위해서는 칼라 글러브를 착용하였다. 테스트 데이터를 위해서는 두 가지 형태의 데이터를 수집하였다. 하나는 긴 소매의 옷을 입고 칼라 글러브를 착용하지 않고 수집하였고, 다른 하나는 짧은 팔의 옷을 입고 수화 문장의 수행 속도 등, 사용자에게 제약을 두지 않고 수집하였다. 48개의 수화 단어 중 18개는 한손을 이용하는 수화이고, 30개는 양손을 이용하는 수화이다. 표 2는 48개의 수화 단어를 보여주고 있고, 그림 6은 6개의 수화 단어 예를 보여주고 있다.

제안된 방법의 성능을 평가하기 위해서 SER(Sign Error Rate)를 식 (15)와 같이 계산하였다.

표 2 실험에 사용된 48개의 수화 단어

한손 수화	AND, BLACK, BOY, FARM, FUNNY, GOOD, HOT, KNOW, LIBRARY, LIE, LIKE, MAN, OUT, PAST, SORRY, STRANGE, TELL, YESTERDAY
양손 수화	ARRIVE, BICYCLE, BIG, BORN, BUTTER, CAR, COLD, DAY, DECIDE, DIFFERENT, FINISH, HERE, INFORM, INTERPRET, MAGAZINE, MANY, MAYBE, NAME, NIGHT, NOW, RAIN, READ, SHOES, SIT, TAKE-OFF, TOGETHER, WANT, WHAT, WORK, WOW

$$SER(\%) = \frac{S + I + D}{N} \times 100 \quad (15)$$

여기서, N 은 테스트 수화의 수, S 대체 오류의 수, I 는 삽입 오류의 수, 그리고 D 는 삭제 오류의 수이다[5].

또한, CDR(Correct Detection Rate)를 아래와 같이 계산하였다.

$$R(\%) = \frac{C}{N} \times 100 \quad (16)$$

여기서, C 는 올바르게 검출된 수화의 수이다.

5.2 실험 결과

3장에서 제안된 CRF 기반 임계치 모델(T-CRF)의 성능과 4장에서 제안된 부사인 추론기를 포함하는 2계층의 수화 적출 모델(Two-layer CRF)의 성능을 평가하기 위해서 HMM과 CRF[22]와 성능을 비교 평가하였다. 5개의 상태를 갖는 이산 HMM을 각 수화 단어별로 생성하였다. 그리고, HMM, CRF를 위해서는 학습 데이터에서 CDR값을 최대로 갖는 고정된 임계치값을 선택하였다.

제안된 기술의 성능을 평가하기 위해서 수화 문장에서 수화 부분을 찾고 인식을 수행하는 수화 적출률과, Ground Truth 정보를 이용하여 사전에 수화 부분만을 적출하고 이를 인식하는 인식률을 계산하였다.

첫 번째로, 긴 소매의 옷을 입고 칼라 글러브를 착용하지 않고 수집한 수화 문장에서 수화를 적출하였다. 표 3

표 3 긴 옷을 입고 수행한 수화 문장을 이용한 수화 적출 결과

Models	C	S	I	D	SER	CDR
HMM ^s	325	78	242	77	82.7	67.7
CRF ^{s,s}	323	31	253	126	85.4	67.2
T-CRF ^{s,s}	389	35	84	56	36.4	81.0
Two-layer CRF ^{s,s}	420	29	11	31	14.7	87.5

a : 손 모양 기반 수화 인식기 사용
 G : 고정된 임계치값 사용
 s : 짧은 수화 적출기 사용
 N : 테스트 수화의 개수, 480
 CRF와 T-CRF의 윈도우 크기 : 2

에서 볼 수 있듯이 제안된 T-CRF^{s,s}, Two-layer CRF^{s,s}의 CDR이 HMM이나 CRF보다 높고, SER은 낮았다. 특히, Two-layer CRF^{s,s}는 T-CRF^{s,s}보다 SER을 22% 줄였다.

그림 7은 수화 문장에서 검출된 6개의 수화 단어 예를 보여주고 있다. 수화 “MANY”의 경우 수화 단어의 끝점에서 정확히 검출이 안된 것을 볼 수 있다. 이는 수화 “MANY”의 끝점에서 손동작과 이후의 손동작이 비슷해서 일부가 오검출된 것이다.

그림 8은 제안된 T-CRF를 이용하여 수화를 적출하는 과정에서 각 레이블의 확률값 변화를 보여주고 있다. 그림에서 볼 수 있듯이 비수화를 위한 레이블의 확률값과 어휘집에 정의된 수화들의 확률값을 비교하여 최대값을 선택하여 수화와 비수화를 구별하였다.

두 번째로, 긴 소매의 옷을 입고 칼라 글러브를 착용하지 않고 수집한 수화 문장에서 수화 부분을 Ground Truth 정보를 이용하여 사전에 적출하고, 이를 인식하는 실험을 수행하였다. 표 4는 인식 결과를 보여주고 있다.

표에서 볼 수 있듯이 제안된 T-CRF^{s,s}, Two-layer CRF^{s,s}의 CDR이 HMM이나 CRF보다 높고, SER은 낮았다.

세 번째로, 짧은 팔의 옷을 입고, 수화 문장의 수행 속도 등 사용자에게 제약을 두지 않고 수집한 수화 문

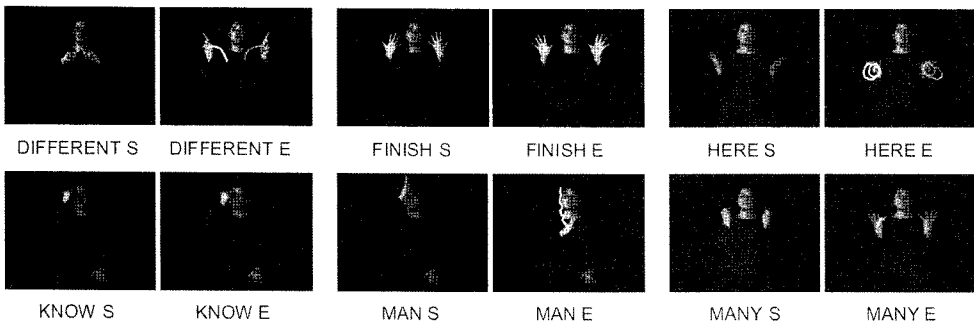


그림 6 6개의 ASL 수화 예: S는 수화 단어의 시작점, E는 수화 단어의 끝점



그림 7개의 수화 적출 결과 예: S는 수화 단어의 시작점, E는 수화 단어의 끝점

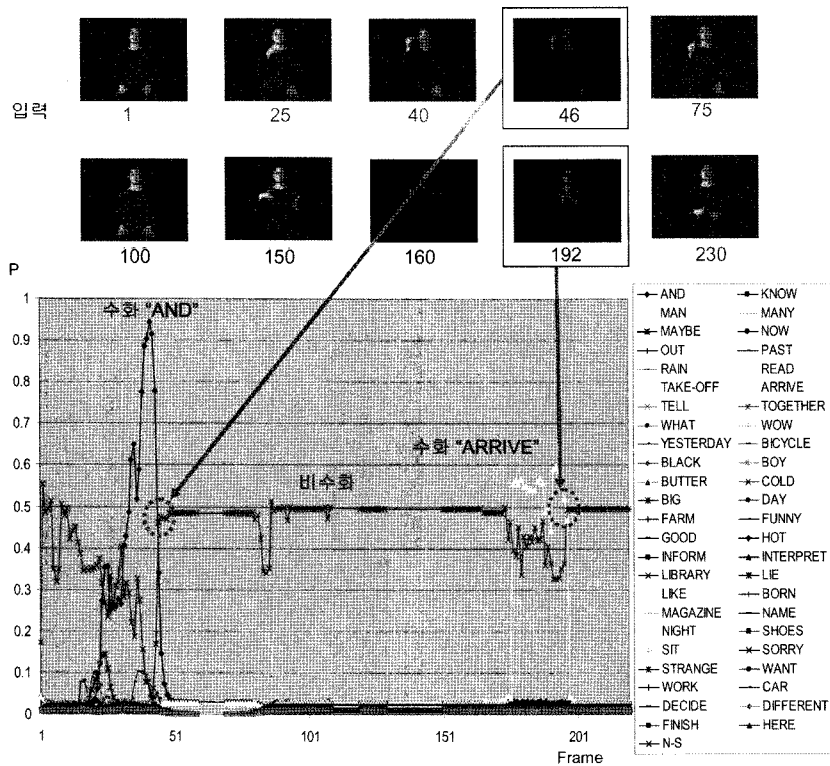


그림 8 수화와 비수화의 확률 변화 그래프: 원으로 표시된 지점부터 비수화의 확률값이 수화들의 확률값보다 크게 나타남

표 4 적출된 수화 단어를 이용한 수화 인식 결과

Models	C	S	I	D	SER	CDR
HMM ⁸	362	118	0	0	25.0	75.4
CRF ^{8,9}	410	70	0	0	14.6	85.4
HCRF ⁸	436	44	0	0	10.0	90.8
T-CRF ^{8,5}	444	36	0	0	7.5	92.5
Two-layer CRF ^{8,5}	452	28	0	0	5.8	94.2

N : 테스트 수화의 개수, 480

장을 이용하여 실험을 수행하였다. 수화 문장 수행 속도, 얼굴의 움직임 등 제약 조건이 없는 환경이기 때문에 표 5에서 볼 수 있듯이 표 3, 4에서의 결과보다 SER은 높고, CDR은 낮았다.

그림 9는 짧은 옷을 입고 수행한 수화 문장에서 검출된 수화 단어 예를 보여주고 있다.

6. 결론

본 논문에서 CRF에 기반한 입계치 모델을 제안하였

표 5 짧은 팔의 옷을 입고 수행한 수화 문장을 이용한 수화 적출 결과

Models	N	C	S	I	D	SER	CDR
CRF ^{a,g}	237	133	24	118	80	93.6	56.1
Two-layer CRF ^{a,s}	237	178	21	112	38	72.1	75.1

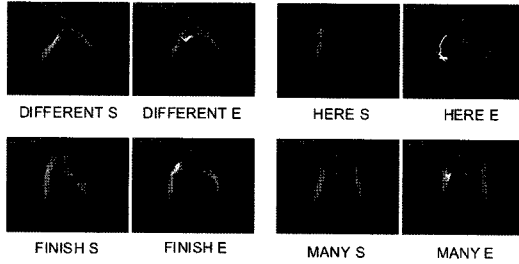


그림 9 4개의 수화 적출 결과 예: S는 수화 단어의 시작점, E는 수화 단어의 끝점

다. 제안된 CRF기반 임계치 모델은 정의된 수화와 비수화 손 동작을 구별하기 위해서 적응적 임계치 역할을 수행한다. 제안된 방법은 어휘집에 정의된 수화 동작과 이외의 동작을 우수하게 구별할 수 있는 방법임을 실험 결과 보여주고 있다.

또한, 수화 적출 및 인식의 성능 향상을 위해 손 모양 기반 수화 인증기, 짧은 수화 적출기, 부사인 추론기를 추가하였다. 손 모양 기반 수화 인증기는 손의 움직임 정보가 비슷한 수화의 오검출을 줄였고, 짧은 수화 적출기는 평균 수화 수행 시간보다 짧은 수화의 인식 성능을 높였고, 부사인 추론기는 서로 다른 수화끼리 비슷한 움직임을 공유할 때, 다른 수화로 오인식하는 것을 방지하였다.

실험 결과, 제안된 방법은 연속된 수화 동작 데이터에서 88%, 사전에 적출된 수화 동작 데이터에서 94%의 인식률을 보였으며, 임계치 모델, 짧은 수화 적출기, 손 모양 기반 수화 인증기, 부사인 추론기를 사용하지 않은 CRF 모델은 연속된 수화 동작 데이터에서 74%, 사전에 적출된 수화 동작 데이터에서 90%의 인식률을 보였다.

참고 문헌

[1] R. Bowden, D. Windridge, T. Kadir, A. Zisserman, and M. Brady, "A Linguistic Feature Vector for the Visual Interpretation of Sign Language," Proc. of European Conference on Computer Vision, Prague, Czech Republic, pp. 390-401, 2004.

[2] A. Braffort, "Argo: An Architecture for Sign Language Recognition and Interpretation," Proc. of Int. Gesture Workshop on Progress in Gestural Interaction, London, UK, pp. 17-30, 1996.

[3] R.D. Yang, S. Sarkar, and B. Loeding, "Enhanced Level Building Algorithm for the Movement Epenthesis Problem in Sign Language Recognition," Proc. of IEEE Conf. on Computer Vision and Pattern Recognition, Minnesota, USA, pp. 1-8, Aug. 2007.

[4] H.-K. Lee and J.-H. Kim, "An HMM-based Threshold Model Approach for Gesture Recognition," IEEE Trans. on Pattern Analysis and Machine Recognition, Vol. 21, No. 10, pp. 961-973, 1999.

[5] H.-D. Yang, A.-Y. Park, and S.-W. Lee, "Gesture Spotting and Recognition for Human-Robot Interaction," IEEE Trans. on Robotics, Vol. 23, No. 2, pp. 256-270, 2007.

[6] W.C. Stokoe, Sign Language Structure: An Outline of the Visual Communication Systems of the American Deaf, Studies in Linguistics: Occasional Papers 8, Linstok Press, 1960.

[7] J. Alon, V. Athitsos, and S. Sclaroff, "Accurate And Efficient Gesture Spotting via Pruning and Subgesture Reasoning," Proc. of ICCV-HCI, Beijing, China, pp. 199-207, Oct. 2005.

[8] A. McCallum, D. Freitag, and F. Pereira, "Maximum Entropy Markov Models for Information Extraction and Segmentation," Proc. of Int. Conf. on Machine Learning, Stanford, USA, pp. 591-598, 2000.

[9] J. Lafferty, A. McCallum, and F. Pereira, "Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data," Proc. of Int. Conf. on Machine Learning, Williamstown, USA, pp. 282-289, Jun. 2001.

[10] L.-P. Morency, A. Quattoni, and T. Darrell, "Latent-dynamic Discriminative Models for Continuous Gesture Recognition," Proc. of IEEE Conf. on Computer Vision and Pattern Recognition, Minneapolis, USA, 2007, pp. 1-8, <http://sourceforge.net/projects/crf>.

[11] S. Wang, A. Quattoni, L.P. Morency, D. Demirdjian, and T. Darrell, "Hidden Conditional Random Fields for Gesture Recognition," Proc. of IEEE Conf. on Computer Vision and Pattern Recognition, New York, USA, pp. 1521-1527, Jun. 2006.

[12] C.W. Ong and S. Ranganath, "Automatic Sign Language Analysis: A Survey and the Future beyond Lexical Meaning," IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 27, No. 6, pp. 873-891, 2005.

[13] W. Gao, G. Fang, D. Zhao, and Y. Chen, "Transition Movement Models for Large Vocabulary Continuous Sign Language Recognition," Proc. of Int. Conf. on Automatic Face and Gesture Recognition, Seoul, Korea, pp. 553-558, May 2004.

[14] C. Vogler and D. Metaxas, "A Framework for

- Recognizing the Simultaneous Aspects of American Sign Language," *Computer Vision and Image Understanding*, Vol. 81, No. 3, pp. 358-384, 2001.
- [15] T. Starner, J. Weaver, and A. Pentland, "Real-Time American Sign Language Recognition Using Desk and Wearable Computer Based Video," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 20, No. 12, pp. 1371-1375, 1998.
- [16] E.-J. Holden, G. Lee, and R. Owens, "Australian Sign Language Recognition," *Machine Vision and Applications*, Vol. 1, No. 5, pp. 312-320, 2005.
- [17] M. Yang, N. Ahuja, and M. Tabb, "Extraction of 2D Motion Trajectories and Its Application to Hand Gesture Recognition," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 24, No. 8, pp. 1061-1074, 2002.
- [18] R.D. Yang and S. Sarkar, "Detecting Coarticulation in Sign Language Using Conditional Random Fields," *Proc. of Int. Conf. on Pattern Recognition*, Hong Kong, China, pp. 108-112, Aug. 2006.
- [19] H. M. Wallach, "Conditional Random Fields: An Introduction," *Technical Report MS-CIS-04-21*, University of Pennsylvania, 2004.
- [20] R. Kasturi and R. Jain, *Computer Vision: Principles*, IEEE Computer Society Press, 1991.
- [21] C.-C. Chang and C.-J. Lin, *LIBSVM: A Library for Support Vector Machine*, 2001, <http://www.csie.ntu.edu.tw/~cjlin/libsvmtools/>.
- [22] T. Kudo, *CRF++: Yet Another CRF Toolkit*, 2005, <http://chasen.org/taku/software/CRF++/>.



양 희 덕

1998년 충남대학교 컴퓨터과학과(학사)
2003년 고려대학교 컴퓨터학과(석사). 2008
년 고려대학교 컴퓨터학과(박사). 2005
년~2006년 Boston University 방문연
구원. 2008년~현재 조선대학교 전자정
보공과대학 컴퓨터공학부 조교수. 관심분

야는 영상처리, 패턴인식, 뇌공학 등



이 성 환

1984년 서울대학교 계산통계학과(학사)
1986년 KAIST 전산학과(석사). 1989년
KAIST 전산학과(박사). 1989년~1995년
충북대학교 컴퓨터과학과 조교수. 2001
년~2002년 MIT Brain and Cognitive
Sciences 학과 방문교수. 1995년~현재

고려대학교 정보통신대학 정교수. 1997년~현재 고려대학교
인공지능연구소 소장. 관심분야는 패턴인식, 컴퓨터 시각,
뇌공학 등