

세미-마르코프 조건 랜덤 필드 기반의 수화 적출

(Sign Language Spotting Based on
Semi-Markov Conditional Random Field)

조 성 식 [†] 이 성 환 ^{**}
(Seong-Sik Cho) (Seong-Wan Lee)

요 약 수화 적출이란 연속된 영상에서 수화의 시작과 끝점을 찾고, 이를 사전에 정의된 수화 단어로 인식하는 방법을 말한다. 수화는 매우 다양한 손의 움직임과 모양으로 구성되어 있고, 그 변화가 다양하여 적출에 많은 어려움이 있다. 특히, 다양한 길이의 궤적 정보로 구성된 수화는 길이가 긴 수화에 대해 짧은 길이를 갖는 수화가 인식에 필요한 정보를 추출하기 어려운 문제점 있다. 본 논문에서는 다양한 길이를 갖는 입력 데이터의 특징을 반영할 수 있는 Semi-Markov Conditional Random Field에 기반하여 다양한 수화의 길이에 강인하게 수화를 적출하는 방법을 제안한다. 성능 평가를 위해 미국 수화와 한국 수화 데이터베이스를 사용하여 연속된 수화 영상에서의 수화 적출 성능을 평가하였고, 실험 결과 기존의 Hidden Markov Model과 Conditional Random Field보다 뛰어난 성능을 보였다.

키워드 : 수화인식, 수화적출, CRF(Conditional Random Field), Semi-CRF(Semi-Markov Conditional Random Field)

Abstract Sign language spotting is the task of detecting the start and end points of signs from continuous data and recognizing the detected signs in the predefined vocabulary. The difficulty with sign language spotting is that instances of signs vary in both motion and shape. Moreover, signs have variable motion in terms of both trajectory and length. Especially, variable sign lengths result in problems with spotting signs in a video sequence, because short signs involve less information and fewer changes than long signs. In this paper, we propose a method for spotting variable lengths signs based on semi-CRF (semi-Markov Conditional Random Field). We performed experiments with ASL (American Sign Language) and KSL (Korean Sign Language) dataset of continuous sign sentences to demonstrate the efficiency of the proposed method. Experimental results show that the proposed method outperforms both HMM and CRF.

Key words : Sign language recognition, sign language spotting, CRF(conditional random field), semi-CRF(semi-Markov conditional random field)

· 이 논문은 2009년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. 2009-0060113). 이 연구에 참여한 연구자는 '2단계 BK21사업'의 지원을 받았다음

[†] 비 회 원 : 고려대학교 컴퓨터학과
sscho@image.korea.ac.kr

^{**} 중 심 회 원 : 고려대학교 컴퓨터·통신공학부 교수
swlee@image.korea.ac.kr

논문접수 : 2009년 7월 22일
심사완료 : 2009년 9월 30일

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

정보과학회논문지 : 소프트웨어 및 응용 제 36권 제 12호(2009.12)

1. 서 론

수화 인식에 대한 중요성이 증가함에 따라, 많은 연구자들이 수화 인식에 대한 연구를 진행해 왔다[1-4]. 이러한 수화 인식에 있어서 핵심적인 문제는 연속된 수화 영상에서 수화의 시작점과 끝점을 찾아 사전에 정의된 수화 단어로 인식하는 수화 적출 문제이다. 수화는 매우 다양한 손의 움직임과 모양으로 구성되어 있고, 그 변화가 다양하여 적출에 많은 어려움이 있다. 특히, 다양한 길이의 궤적정보로 구성된 수화는 길이가 긴 수화에 비해 짧은 길이를 갖는 수화가 인식에 필요한 정보를 추출하기 어려운 문제점이 있다. 이러한 문제점은 Brasher

등의 논문과 Alon의 연구에 자세히 명시되어 있다[1-3]. 기존의 수화 인식 연구에서는 손의 움직임 정보를 표현하기 위해 DTW(Dynamic Time Warping), HMM(Hidden Markov Model) 그리고 CRF(Conditional Random Field)와 같은 모델들이 사용되어 졌다. 최근 가장 많이 사용된 HMM과 CRF를 비교해보면, HMM은 다양하고 복잡한 구조를 모델링할 수 있는 장점이 있지만 모든 관측치 간에 독립이라는 가정을 한다는 단점이 있다. 반면, CRF는 HMM에서 가장했던 관측치 간에 독립이라는 가정을 완화시킬 수 있는 장점을 보이고 있지만, 다양한 길이의 데이터에 대한 특징을 표현하는 것에 부족한 점이 있다.

본 논문에서는 Semi-Markov Conditional Random Field(Semi-CRF)에 기반하여 추론된 확률에 임계값을 적용하여 다양한 수화의 길이에 강인하게 수화를 적출하는 방법을 제안한다. Semi-CRF는 CRF의 장점을 유지하면서 다양한 길이를 갖는 입력 데이터의 특징을 반영할 수 있는 장점이 있다. 성능 평가를 위해 미국 수화와 한국 수화 데이터베이스를 사용하여 연속된 수화 영상에서의 수화 적출을 수행한 실험 결과, 기존의 HMM과 CRF보다 뛰어난 성능을 보였다.

2. 관련 연구

제스처 및 수화 인식에 대한 중요성이 증가함에 따라, 많은 연구자들이 수화 인식에 대한 연구를 진행해 왔다. 수화 인식에서의 중요한 이슈들은 손 검출 및 추적, 손의 3차원 형태 정보 인식, 연속된 영상에서의 수화 적출 등이 있다[4].

Vogler[5] 등은 대용량의 수화를 인식하기 위해 수화를 구성하는 기본 요소들로 특징을 추출하여 이를 parallel HMM으로 인식하는 방법을 제안하고 제안된 방법이 기존의 HMM보다 대용량의 수화 인식에 적합함을 보였다. 다양한 길이의 수화에서의 문제점은 Brasher 등과 Alon의 논문에 자세히 명시되어 있다[1-3]. Brasher[2] 등은 모바일 기기에 적용 가능한 HMM 기반의 수화 인식 알고리즘에 대한 이전 연구에서, 다양한 길이의 수화의 정확한 인식을 위한 HMM 구조를 제안하였으나, 제안된 HMM 모델이 주어진 학습 데이터의 길이에 의존적이라는 단점이 있었다. Alon[1,3]은 DTW 기반의 수화 적출 연구에서 짧은 길이를 갖는 수화의 특성을 모델로 표현하는 것은 매우 어렵기 때문에, 적출률이 매우 낮은 결과를 보인다고 분석하였다.

3. 수화 적출 알고리즘

3.1 손의 궤적과 위치 정보에 기반한 특징 추출

본 논문에서는 기존의 수화 인식 연구[1,3]와 같이 결

러 정보와 움직임 정보 기반의 얼굴 및 손 검출 방법을 사용하였다. 얼굴은 Haar-like 얼굴 검출기를 사용하여 검출하였으며, 손의 후보 영역을 추출하기 위한 피부색 영역 검출은 얼굴 검출기를 통해 획득한 얼굴 영역의 피부색 분포를 기반으로 검출하였다. 주어진 영상에서 얼굴과 손이 검출되면 표 1에서 보이는 바와 같은 6가지의 특징을 계산하여 수화 적출 및 인식에 이용한다. P_{LH} 는 그림 1에서 보이는 바와 같이 얼굴의 중심과 왼손의 중심 사이의 각도로 표현되는 특징 벡터 θ_{FLH} 와 얼굴의 중심과 왼손의 중점 사이의 거리로 표현되는 특징 벡터 d_{FLS} 를 GMM(Gaussian Mixture Model)을 이용하여 클러스터링한 결과로 추출된다. 양손의 상대적 위치를 나타내는 S_{TH} 는 다음과 같이 계산된다.

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

d_{HL} 과 d_{HR} 은 얼굴의 중심과 오른손 및 왼손의 중점과의 수평 거리를 각각 나타낸다. d_v 는 양손의 중점간의 수직 거리를, T_h 와 T_v 는 양손의 수직 및 수평거리에 대한 임계값을 나타낸다.

표 1 손의 움직임에 기반한 특징들

특징	의미
P_{LH}	왼손의 현재 위치
P_{RH}	오른손의 현재 위치
S_{TH}	양손의 상대적 위치
O_{TH}	양손의 겹침 유무
C_{LH}	왼손의 이동 방향을 표현하는 체인코드
C_{RH}	오른손의 이동 방향을 표현하는 체인코드

양손의 겹침은 양손의 후보영역이 겹쳐진 비율에 의해 다음과 같이 계산된다.

$$O = \begin{cases} 1, & \min\left(\frac{R_b}{H_l}, \frac{R_b}{H_r}\right) > T_o \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

H_l 은 왼손 영역, H_r 은 오른손 영역, R_b 는 양손의 겹친 영역을 그리고 T_o 는 양손의 겹침을 검출하기 위한 임계값으로 사용되었다(실험을 통해 $T_o = 0.3$ 으로 사용).

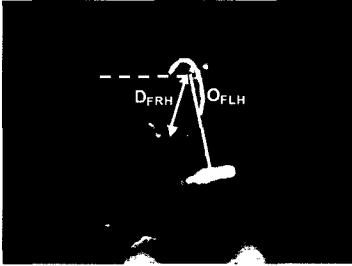
손의 움직임에 대한 체인코드 C_{LH} 와 C_{RH} 는 이전 영상에서 현재 영상으로 손이 이동한 방향을 표현하는 8개의 방향 코드와 움직임이 없는 경우를 표현하는 1개의 더미 코드로 표현된다[6].

3.2 Semi-CRF 기반의 수화 적출

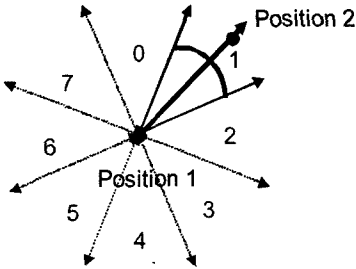
Semi-CRF는 CRF의 확장된 모델로서, 입력 데이터의 길이의 특징을 표현할 수 있다는 장점을 갖는다. $s = s_1, s_2, \dots, s_n$ 이 연속된 관측치 x 의 부분열일 때, 연속

된 관측치 x 에 대한 수화 레이블 y 의 확률은 다음과 같이 계산된다[7].

$$P_{\theta}(s|x) = \frac{1}{Z_{\theta}} \exp(G(s,x)) \quad (3)$$



(a) 양손 및 얼굴에 대한 특징



(b) 손의 움직임에 대한 체인 코드의 예시

그림 1 손의 움직임과 위치 정보에 기반한 특징 추출의 예시

Z_{θ} 는 정규화 상수, 파라미터 $\theta = \lambda_1, \lambda_2, \dots, \lambda_n; \mu_1, \mu_2, \dots, \mu_n$
 $G(s,x) = \sum_j |s|g(j,x,s)$ 이며, 부분열 특징 함수 $g(j,x,s)$
 는 다음과 같이 정의된다.

$$g(j,x,s) = \lambda_v t_v(y_{i-1}, y_i, x, t_j, u_j) + \mu_m s_m(y_{i-1}, y_i, x, t_j, u_j) \quad (4)$$

여기서 $t_v(y_{i-1}, y_i, x, t_j, u_j)$ 는 시간 i 와 $i-1$ 상태에 대한 전이 특징 함수(transition feature function), λ_v 는 전이 특징 함수의 가중치를 나타내며, $s_m(y_{i-1}, y_i, x, t_j, u_j)$ 는 시간 i , $i-1$ 상태와 관측치 x 에 대한 상태 특징 함수(state feature function)을 의미하고 μ_m 은 상태 특징 함수의 가중치를 의미한다. 전이 특징 함수와 상태 특징 함수의 t_j 와 u_j 는 각각 부분열의 시작점과 끝점을 의미하며, 이는 $|t_j - u_j|$ 길이의 부분열을 의미한다.

Semi-CRF의 각 상태는 학습된 하나의 레이블을 의미하며, 이는 각 수화 어휘에 해당한다. CRF와는 다르게 Semi-CRF는 하나의 영상에 대한 인식 결과를 보이

는 것이 아니라 입력 영상의 부분열에 하나의 인식 결과를 할당한다. 이는 서로 다른 수화가 비슷한 궤적정보를 보일 경우에 발생하는 오인식에 강인하다는 장점을 갖게 되며, 특히 위의 부분열 특징 함수에서 설명한 바와 같이 $|t_j - u_j|$ 길이의 입력에 대한 특징을 함께 고려하여 다양한 길이의 수화에도 강인하게 적용할 수 있는 장점 역시 갖게 된다.

본 논문에서는 Semi-CRF 모델을 기반으로 추론된 확률에 임계값을 적용하여 수화를 적용하였다. 연속된 수화 영상이 주어진 경우 Semi-CRF 기반의 확률 추론은 다음과 같이 계산된다[7].

$$\begin{aligned} \operatorname{argmax}_{\theta} P(s|x, \theta) &= \operatorname{argmax}_{\theta} \theta \cdot G(s,x) \\ &= \operatorname{argmax}_{\theta} \theta \cdot \sum_j g(y_j, y_{j-1}, x, t_j, \mu_j) \end{aligned} \quad (5)$$

4. 실험 및 결과 분석

4.1 실험환경

본 논문에서는 미국 수화와 한국 수화 두 가지 종류의 데이터베이스를 사용하여 실험을 수행하였다. 미국 수화 데이터베이스는 실생활에서 수화를 구사하는 한 명의 미국인 여성이 구사하는 24개 단어의 수화 단어로 구성되어 있다. 미국 수화의 길이에 따라 8개를 짧은 수화, 나머지 16개를 보통 길이의 수화로 분류하였다. 수화 단어는 학습 및 테스트를 위해 장갑을 착용한 경우와 맨손인 경우 각각 10회씩 반복하여 촬영되었으며 7개의 수화는 한손을 사용한 수화, 17개의 수화는 양손을 사용한 수화로 구성되어 있다. 한국 수화 데이터베이스는 실생활에서 수화를 구사하는 한 명의 한국인 남성이 구사하는 60개 단어의 수화 단어로 구성되어 있다. 한국 수화의 길이에 따라 13개를 짧은 수화, 나머지 47개를 보통 길이의 수화로 분류하였다. 각 수화 단어는 학습 및 테스트를 위해 장갑을 착용한 경우와 맨손인 경우 각각 10회씩 반복하여 촬영되었으며 28개의 수화는 한손을 사용한 수화, 32개의 수화는 양손을 사용한 수화로 구성되어 있다.

제한된 수화 적출 알고리즘의 정확도를 평가하기 위하여 삽입 오류(insertion error), 삭제 오류(deletion error), 교체 오류(substitution error)의 3가지 종류의 오류 형태를 정의하여 사용하였다. 삽입 오류는 수화가 아닌 의미 없는 손동작에서 수화로 적출하는 경우를 말하며, 삭제 오류는 이와는 반대로 수화인 손동작에서 수화를 적출하지 못한 경우를 의미한다. 마지막으로 교체 오류는 수화를 오인식하여 적출한 경우를 의미한다. 총 N 개의 수화가 있을 경우 I 를 삽입 오류, D 를 삭제 오류, S 를 교체 오류라 하면 수화 오류율(SER: Sign Error Rate)은 다음과 같이 계산된다[1,3].

$$SER = \frac{S+I+D}{N} \times 100 \quad (6)$$

또한, 수화를 쉽게 적출한 경우를 C 라 했을 경우 수화 적출율은 다음과 같다.

$$P_d = \frac{C}{N} \times 100 \quad (7)$$

4.2 실험 결과

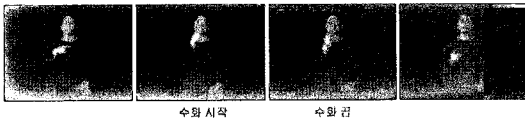
본 논문에서는 제안된 Semi-CRF 기반의 수화 적출 인식 방법의 성능을 평가하기 위해 HMM과 CRF와 비교 평가를 수행한 실험 결과는 표 2, 3과 같다. Semi-CRF 기반의 수화 적출 방법은 기존의 HMM, CRF 기반의 수화 적출방법보다 뛰어난 성능을 보이고 있다. 특히 그림 2에서 보이는 것처럼 입력 데이터의 길이 특징을 반영 및 표현 가능한 Semi-CRF 기반의 수화 적출 방법은 짧은 수화에서도 정확한 적출 성능을 보이고 있다.

표 2 미국 수화 적출 결과

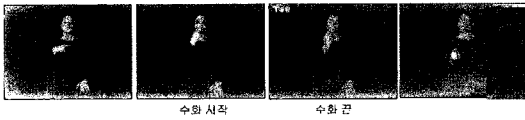
Models	N	C	S	I	D	SER	P_d
HMM	240	162	32	62	46	58.3	67.5
CRF	240	172	31	43	37	46.3	71.7
Semi-CRF	240	194	24	28	22	30.8	80.8

표 3 한국 수화 적출 결과

Models	N	C	S	I	D	SER	P_d
HMM	600	375	118	159	107	64.0	62.5
CRF	600	410	98	157	92	57.8	68.3
Semi-CRF	600	448	75	114	78	44.3	74.7



(a) CRF 기반 수화 적출 결과: TELL



(b) semi-CRF 기반 수화 적출 결과: TELL

그림 2 짧은 수화 Tell에 대한 적출 결과 비교

5. 결론

본 논문에서는 Semi-CRF에 기반하여 다양한 수화의 길이에 강인하게 수화를 적출하는 방법을 제안하였다. Semi-CRF는 CRF의 장점을 유지하면서 다양한 길이를 갖는 입력 데이터의 특징을 반영할 수 있는 장점이 있는 모델로, 연속된 수화 영상에서의 수화 적출 및 인식

을 수행한 실험 결과, 기존의 HMM과 CRF보다 뛰어난 성능을 보였다. 차후 연구로, 수화 적출 및 인식에서 많은 문제가 되고 있는 급격한 궤적 변화와 수화 구사자에 따른 인식을 저하 문제를 해결하는 방향으로 연구가 필요할 것이다.

참고 문헌

- [1] J. Alon, V. Athitsos, Q. Yuan, and S. Sclaroff, "A Unified Framework for Gesture Recognition and Spatiotemporal Gesture Segmentation," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol.31, 2009.
- [2] H. Brashear, T. Starner, P. Lukowicz, and H. Junker, "Using Multiple Sensors for Mobile Sign Language Recognition," *Proc. IEEE International Symposium on Wearable Computers, Florida, USA*, pp.45-52, October 2003.
- [3] J. Alon, Spatiotemporal Gesture Segmentation, PhD Dissertation, Department of Computer Science, Boston University, 2006.
- [4] C. Ong and S. Ranganath, "Automatic Sign Language Analysis: A Surevey and the Future beyond Lexical Meaning," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol.27, no.10, pp.1848-1852, 2007.
- [5] 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.
- [6] H.-K. Lee and J.-H. Kim, "An HMM-based Threshold Model Approach for Gesture Recognition," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol.21, no.10, pp.961-973, 1999.
- [7] S. Sarawagi and W. Cohen, "Semi-Markov Conditional Random Fields for Information Extraction," *Proc. Annual Conference on Neural Information Processing Systems, Vancouver, Canada*, pp. 1185-1192, December 2004.

조 성 식

정보과학회논문지 : 소프트웨어 및 응용
제 36 권 제 1 호 참조

이 성 환

정보과학회논문지 : 소프트웨어 및 응용
제 36 권 제 1 호 참조