

손동작 인식 시스템을 위한 동적 학습 알고리즘

심재록*, 박호식*, 김태우*, 나상동**, 배철수*

*관동대학교 전자통신공학과

**조선대학교 컴퓨터공학과

e-mail : jaeroks@empal.com

Dynamic Training Algorithm for Hand Gesture Recognition System

Jae-Rok Shim*, Ho-Sik Park*, Tae-Woo Kim*, Sang-Dong Ra**, Cheol-Soo Bae*

*Dept. of Electronic Communication Eng, Kwandong University

**Dept. of Computer Eng, Chosun University

요 약

본 논문에서는 카메라-투영 시스템에서 비전에 기반을 둔 손동작 인식을 위한 새로운 알고리즘을 제안하고 있다. 제안된 인식방법은 정적인 손동작 분류를 위하여 푸리에 변환을 사용하였다. 손분할은 개선된 배경 제거 방법을 사용하였다. 대부분의 인식방법들이 같은 피검자에 의해 학습과 실험이 이루어지고 상호작용에 이전에 학습단계가 필요하다. 그러나 학습되지 않은 다양한 상황에 대해서도 상호작용을 위해 동작 인식이 요구된다. 그러므로 본 논문에서는 인식 작업 중에 검출된 불완전한 동작들을 정정하여 적용하였다. 그 결과 사용자와 독립되게 동작을 인식함으로써 새로운 사용자에게 신속하게 온라인 적용이 가능하였다.

1. 서론

본 논문에서 효과적인 인간과 컴퓨터 사이의 인터페이스를 위하여 카메라-투영시스템을 이용한 가상의 마우스 시스템을 제안하고 있다. 비디오투영은 멀티미디어 프레젠테이션에서 넓게 사용되고, 일반적으로 키보드, 마우스와 같은 표준장치들을 이용하여 컴퓨터와 대화한다. 이 경우 프레젠테이션은 컴퓨터 주변으로 제한된다. 그러므로 만약 사용자들이 하드웨어의 제약 없이 정확하게 디스플레이 시킬 수 있다면 좀 더 유용하고 효과적일 것이다.

본 논문에서는 카메라-투영 시스템에서 손동작에 의해 투영된 프레젠테이션과 상호작용하는 방법을 제안한다. 상호작용을 위해 투영된 영상의 정면에서 프레젠테이션 동작을 카메라로 측정하여 저장된 영상을 사용하였다.

동작 인식 시스템들은 일반적으로 사용자와 독립적으로 동작을 인식할 수 있어야 한다. 그러므로 본 논문에서는 상호 작용하는 지시에 의해 학습과 인식을 하도록 하였다. 기존의 상호작용에 의한 동작 매개변수는 초기의 인식 매개변수를 적용하고, 실제의 사용

자는 지시 여부와 관계없이 계속해서 이 매개변수를 변경한다. 지시 받은 학습은 사용자가 잘못된 동작에 대해서는 재학습을 의미한다. 반면에 지시 받지 않은 학습은 인식하는 동안 계속해서 새로운 매개변수가 발견됨을 의미한다. 제안된 방법의 장점은 불완전한 동작들을 재학습하고, 새로운 사용자들에게 추가 학습 없이 적용이 가능하다.

2. 시스템의 개요 및 보정

본 논문에서는 9 가지의 손동작에 의한 단어 표현과 복잡한 배경에 대한 투영을 다루고 있다. 카메라는 단지 투영된 표면에서부터 투영된 배경만을 획득한다.

카메라로부터 획득한 영상은 카메라와 투영된 화면 사이의 객체와 배경에 투영된 영상에서 기인한다. 카메라 영상의 사물은 3 차원 환경 하에서 2 차원으로 투영되므로 키스토프 같은 원근 왜곡이 나타난다. 따라서 이 시스템은 투영된 이미지와 카메라에 의해 왜곡된 영상 사이의 픽셀의 좌표 기록이 필요하다.

이러한 원근 왜곡은 카메라와 투영된 영상 사이의

좌표의 왜곡을 식 (1)과 같은 2 차 다항식에 의해 모델화 된다.

$$\begin{aligned} x' &= a_0 + a_1 \cdot x + a_2 \cdot y + a_3 \cdot x^2 + a_4 \cdot xy + a_5 \cdot y^2 \\ y' &= b_0 + b_1 \cdot x + b_2 \cdot y + b_3 \cdot x^2 + b_4 \cdot xy + b_5 \cdot y^2 \end{aligned} \quad (1)$$

여기서 (a_i, b_i) 에서 기하학적 왜곡에 대한 가중 계수이고, (x, y) 의 원래의 (x', y') 는 새로운 변환 위치이다. 이러한 입출력의 견본 지점은 패턴 영상의 특별 보정으로 결정된다. 카메라 영상에서 추출된 좌표의 측정치와 (x', y') 좌표 사이의 평균자승 오류를 최소로 하는 값을 가중 계수로 사용하였다.

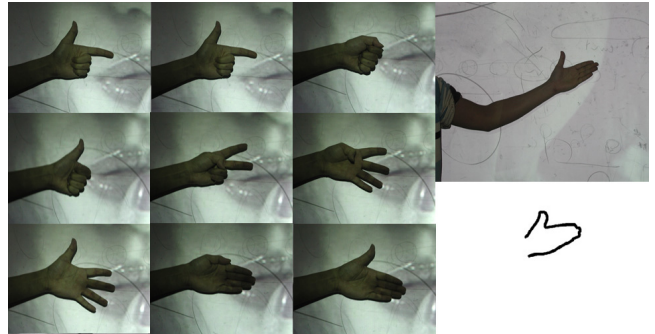
3. 팔 분할

카메라에 포착되는 영상은 투영된 영역의 일부 영상이다. 투사기와 화면 사이의 객체들은 빔을 반사함으로써 손의 구성과 색상은 투영된 영상과 객체의 위치에 따라 연속적으로 변하게 된다. 그러므로 색분할 또는 분할을 위한 영역 확장이 어렵게 된다. 손가락 추적법[5]으로 분할 할 수 있으나 나타낼 수 동작의 구분이 어려워 표현할 수 있는 단어의 제한이 있다. 본 논문에서 배경 제거 방법과 배경 변화를 다루는 방법을 확장시켜 분할을 시도하였다. 투영시 투영된 화면의 반사율은 100%에 가깝지만 인간의 피부는 부분적으로 빛을 흡수하여 광학 필터[7]로 처리 동작하기에 최대 70%의 반사율을 나타낸다. 제안된 방법은 문턱치 값에 의해 분류된 차이 영상에 의해 각 영상 채널과 전경 객체와의 차이를 나타낸다. 만약 투영기의 광선 강도가 손의 위치에서 약하다면 즉, 예를 들어 투영된 배경이 검은색이라면 손과 배경 반사 사이의 차이는 작고 잡음이 있을 것이다. 그러므로 최소한의 투영기 조명은 투영되었던 막대그래프의 변환에 의해 최소한 빛의 강도가 약 20%정도 증가되도록 하였다.

전완은 어떠한 중요 정보를 포함하고 있지 않으므로, 손바닥과 전완을 완벽하게 분할하는 것은 중요하다. 본 논문에서는 순간 영상에서 팔의 주된 방향을 계산하는 방법을 사용하였다. 손의 방향을 고려하여 손목의 너비와 전완의 너비를 측정한다. 전완에서 손바닥 사이 손목 지점에서 너비 값이 증가하기 때문에 전완의 너비 매개변수들을 분석하고 손의 해부학적 구조를 이용하여 손목의 위치를 결정한다.

또한 배경 분할은 투영된 배경의 변화에 민감하다. 그러나 투영되는 배경 영상은 인공적으로 만든 배경임으로 중요한 문제로 대두되지는 않는다. 가장 큰 문제점은 카메라에 잡힌 영상은 원근 투영에 의한 기하학적 왜곡과 카메라와 투사기의 색상 변환 기능에 의한 색상 변화이다. 색상 보정을 위해 입력 영상과 획득된 영상의 색상으로 LUT(look up table)를 생성한다. LUT 생성을 위해 5 차 다항식을 견본 값에 맞춘다. 배경을 변화 시키었을 때 시스템은 기하학적 왜곡 평형에 의해 입력 영상을 왜곡시킨다. 그리고 영상 미분을 위해 계산된 LUT 에 의거하여 정확한 배경이미지를 생성한다. 상호작용하는 동안 이 초기 배경 영상은

정확한 분할을 위해 원본 카메라 영상에 의해 새롭게 갱신된다.



(그림 1) 동작단어와 분류결과

4. 윤곽선 분류

본 논문에서는 분류를 위해 경계선에 근거한 방법을 사용하였다. 푸리에 변환은 형태 묘사에 넓게 사용된다. 푸리에 변환인식은 신경네트워크 분류에 근거를 두고 있으며 6 가지 동작을 측정된 결과 90~91% 인식률을 가진다. 이 방법에서 동작 윤곽선은 최근린 법과 MFD(modified Fourier descriptor)[10]에 근거한 측정법으로 분류되어진다. 이 측정법은 모양의 회전, 변형, 비례에 대하여 불변한다. 검사된 형태는 벡터특징에 의해 정의 내려지고, 그것은 푸리에 시리즈로 발전하여 주기적이 된다. 제안된 방법은 형태 경계에 따라 두 손목 지점 사이에 연속적인 특징을 생성한다. 이러한 방법은 좀 더 명백한 특징을 나타낼 수 있다. 예를 들자면 집게나 엄지손가락만 보여 질 때는 손바닥의 형태 윤곽은 매우 유사하다. 반면에 손목 지점 사이의 윤곽은 명확히 나타난다. 정의 내려진 경계의 결과는 경계 지점 x, y 좌표의 복합 시퀀스로 구성되어 진다. 복합 시퀀스의 DFT(discrete Fourier transform)를 계산하고, 회전 불변의 DFT 계수의 확대 값을 적용한다. 대칭거리 계산으로 얻어진 MFD 방법을 확장했다. F_n^1 과 F_n^2 의 곡선을 비교한 DFT 계수표시는 σ 로 표준 편차를 표시한다. 두 곡선 사이의 거리 측정법은 식(2)와 같다.

$$Dist(F_n^1, F_n^2) = \sigma \left(\frac{F_n^1}{F_n^2} \right) + \sigma \left(\frac{F_n^2}{F_n^1} \right) \quad (2)$$

제안방법은 DC 구성요소를 제외한 첫 번째 6 개의 계수들로 계산한다. 그러므로 이것은 형태 경계의 불규칙한 잡음에 대하여 건설하다.

제안된 방법을 실험을 위하여 그림 1 과 같은 9 개의 동작에 대하여 인식실험을 하였다. 초기 학습 집단의 수는 매우 작아 일반적으로 1 개를 사용하였다. 이 특징은 만약에 온라인 학습을 위해서는 매우 중요하다. 표 1 은 일부의 학습자와 피검자와 함께 제안된 자세 분류방법인식률을 나타낸 것이다. 학습자들은 중

으로 피검자는 열로 표시하였다. 총 400 명의 피검자로 실험을 진행하였다.

표 1. 자제 분류 결과

Test User	Recognition results [%] Trainer users			
	User A	User B	User C	User D
User A	99.8	96.1	86.2	94.6
User B	90.4	97.6	90.1	93.9
User C	94.5	92.7	99.6	98.9
User D	95.5	95.0	96.7	99.1

만약 학습자와 피검자가 같은 사람인 경우의 인식률은 97%를 상회한다. 다른 경우에는 최소 86%를 나타내었다. 실험 결과 모든 동작의 효율이 의미 있게 감소하지 않음을 알 수 있었다. 그러므로 만약 시스템이 사용자와 상호작용에 의해 불완전하게 관찰되었던 동작들을 배울 수만 있다면 좀 더 효율적으로 동작할 것으로 사료된다.

5. 지시 학습에 의한 동적 학습

본 논문의 주된 목적은 새로운 사용자들에 대해 모든 동작의 재학습을 배제하는 것이다. 시스템의 인식 효율 향상을 위해 단지 잘못 검출된 동작만을 수정하도록 하였다. 이 학습방법은 지시 받지 않은 학습과 지시 받은 학습 모두를 포함한다. 사용자들은 인식의 결과를 따르고, 만약 그 결과가 정확하지 않다면 특별한 동적인 동작에 의해 사용자 피드백을 생성한다.

만약 그 결정이 인식 면에서 바른다면, 그 확인된 동작 파라미터들은 지금의 동작파라미터에 의해 지속적으로 갱신 될 것이다. 지속적으로 새롭게 되는 저장된 동작 파라미터들에 의해 이 시스템은 무의식적인 손동작(지시 받지 않은 학습)의 작은 변화에 적응하게 된다. 이 학습은 시스템 파라미터와 실제동작 파라미터 사이에 실행되고 있는 평균적인 계산 값을 제공한다. 예를 들어 사용자가 피곤하여 표준 동작을 할 수 없을 때 이 시스템은 새로운 동작으로 학습할 것이다.

만약 부정확한 결정일 시에는 사용자는 계획된 사용자 인터페이스 위에 상황을 알아차릴 수 있다. 그리고 피드백 동작 실행에 의해 지시 받은 학습의 시작을 가리킨다. 이 피드백 동작은 악수하는 것처럼 빠른 손의 동요하는 동작을 의미한다. 이 피드백신호는 사용자-독립으로 단지 손바닥 중심의 속도를 사용한다. 그리고 발견 되어진 다른 결과를 무시한다. 상호작용하는 동안 이 빠른 손 흔들림 동작은 드물게 관찰되었다. 손바닥 위치가 변화가 처음 주어진 시간 동안보다 클 때 이 시스템은 피드백 신호를 인식했다. 만약 손이 빠르게 되면 이 요약된 기준은 미리 정의 내려진 기준에 비해 더 클 것이다. 그리고 피드백신호는

검출 되고 문턱치 값이 조정된다.

이 지시된 학습은 다음과 같은 규칙을 따른다

1. 동작하는 동안 학습된 사람들은 바르게 표시되었던 동작을 실행할 시간을 필요 한다. 그 시스템은 만약에 실행 된 동작의 확인이 시간 안에 일관되어 질 때 단지 학습을 위해 견본을 기록한다. 따라서 MDF 에 의해 측정 된 거리는 연속적인 손 윤곽 사이에서 3 초 동안 인정된다.
2. 학습하는 동안 시스템은 선택된 동작의 통계그래프를 표시한다. 그리고 완벽한 분할과 학습을 위해 동일한 흰색배경을 투사한다. 이 통계그래프는 마지막 학습과정으로부터 분할 된다. 그리고 이것은 인식에 관해 투영기 결과로 생기는 연속의 피드백에 의해 카메라의 관점 밖에서 표시 된다.
3. 시스템은 학습을 위해 동작견본을 받아들인 후에 다음에 가상의 동작구분에 기인하기 위해 통보되고 표시된다.
4. 만약 불안정하게 확인된 클래스가 시스템에서 보고 되었고 사용자가 그것을 정정했다면 그 학습은 피드백신호에 의해 끝날 수 있다. 만약 제시한 동작이 적절한 동작이 아니면 사용자는 정확한 동작이 표시 될 때까지 학습해야 한다. 최악의 경우 모든 동작이 재학습 된다

표 2. 동적 학습을 적용한 인식 결과.

user order	Without dynamic correction[%]	After dynamic gesture correction[%]	Number of correction steps
User A	99.5	-	-
User B	89.8	96.1	4
User C	94.7	97.0	2
User D	99.1	-	-
User B	93.0	98.8	1
User C	96.8	99.2	2

본 논문에서는 9 가지 모든 동작을 위해 몇 개의 작업을 정의하였다. 사용자들은 임의의 명령에서 연속해서 시스템을 시험해보고 동적인 학습 여부에 인식 효율을 실험하였다. 첫 번째 사용자는 시스템의 처음 동작 파라미터들을 학습한다. 그리고 다른 사용자들은 예비학습을 하지 않는다. 표 2 에서 동적인 동작 교정 후와 상호작용 없이 일반인식 결과의 시험결과를 나타내었다. 마지막으로 만약 사용자가 동작교정을 지적했을 때 교정단계들 나타내었다.

6. 결론

본 논문에서는 동작인식에 있어 사용자 친화적이고 독립적인 동적 학습 알고리즘을 제안하였다. 제안된 방법은 한정된 수의 학습 집단으로 학습이 가능하였고, 지시 학습은 검출된 불완전한 동작을 정정이 가능하도록 하였다. 몇몇 피검자를 대상으로 지시 학습 시스템을 실험한 결과 인식 성능의 향상을 확인할 수 있었다. 실험 결과 96%를 이상의 인식률을 얻을 수 있어, 푸리에에 근거한 다른 방법보다 보다 효율적이라는 것을 입증하였다.

참고문헌

- [1] A. Licsar, T. Sziranyi, "Hand Gesture Based Film Restoration", Proc. of PRIS'02, Alicante, 2002, pp. 95-103.
- [2] N.A. Streitz, J. Geissler, J.M. Haake, and J. Hol, "DOLPHIN: Integrated Meeting Support across Liveboards, Local and Remote Desktop Environments", Proc. of the ACM CSCW, 1994, pp. 345-358.
- [3] Q.S. Fraser, and P. Robinson, "BrightBoard: A Video-Augmented Environment", Proc. of CHI'96, 1996, pp. 134-141.
- [4] P. Wellner, "Interacting with Paper on the DigitalDesk", Communications of the ACM, 1993.
- [5] C. Hardenberg, and F. Berard, "Bare-Hand Human Computer Interaction", Proc. of ACM PUI, Orlando, 2001.
- [6] W.K. Pratt, Digital Image Processing, Wiley-Interscience, New York, 2001.
- [7] M. Storring, H. J. Andersen, and E. Granum, "Skin colour detection under changing lighting conditions", 7th Symposium on Intelligent Robotics Systems, Coimbra, Portugal, pp. 20-23.
- [8] K. Imagawa, R. Taniguchi, D. Arita, H. Matsuo, S. Lu, S. Igi, "Appearance-based Recognition of Hand Shapes for Sign Language in Low Resolution Image", Proc. of 4th ACCV, 2000, pp. 943-948.
- [9] E.S. Koh, Pose Recognition System. BE Thesis, National University of Singapore, 1996.
- [10] Y. Rui, A. She, T.S. Huang, "A Modified Fourier Descriptor for Shape Matching in MARS", Image Databases and Multimedia Search, 1998, pp. 165-180.
- [11] C.W. Ng, and S. Ranganath, "Real-time gesture recognition system and application", Image and Vision Computing 20, 2002.