

손의 외곽선 추출에 의한 실시간 제스처 인식

이인호*, 박찬중

한국전자통신연구원 컴퓨터소프트웨어기술연구소 가상현실연구개발센터

Real-Time Gesture Recognition Using Boundary of Human Hands from Sequence Images

LEE, Inho and PARK, Chanjong

VR Research Center, Computer & Software Technology Lab., ETRI

요약

제스처 인식은 직관적일 뿐 아니라, 몇 가지의 기본 구성요소에 의하여 코드화(code)가 용이하여, 인간과 컴퓨터의 상호작용(HCI, Human-Computer Interaction)에 있어서 폭넓게 사용되고 있다. 본 논문에서는 손의 모양이나 크기와 같은 개인차 및 조명의 변화나 배율과 같은 입력환경의 영향을 최소화하여, 특별한 초기화 과정이나 모델의 준비과정 없이도 제스처를 인식할 수 있고, 적은 계산량으로 실시간 인식이 가능한 제스처 인식 시스템의 개발을 목표로 한다. 본 논문에서는 손에 부착하는 센서나 마커 없이, CCD 카메라에 의하여 입력된 컬러영상에서, 컬러정보 및 동작정보를 이용하여 손영역을 추출하고, 추출된 손의 경계선 정보를 이용하여 경계선-중심 거리 함수를 생성했다. 그리고, 손가락의 끝 부분에서는 경계선-중심 거리가 극대점을 이룬다는 원리를 이용하여 생성된 함수의 주파수를 분석하여 극대점을 구함으로써 각각의 손가락 끝 위치를 찾고, 손의 자세를 인식하여 제스처를 인식했다. 또한 본 논문에서 제안된 제스처 인식 방법은 PC상에서 구현되어 그 유용성과 실효성이 증명되었다.

1. 서론

최근 컴퓨터의 급속한 발전과 정보화 사회로의 진입에 힘입어, 키보드나 마우스 또는 조이스틱과 같은 기존의 장치들을 이용한 인터페이스에서 벗어나 좀더 자유롭고 편리한 새로운 사용자 인터페이스(HCI, Human Computer Interaction)의 요구가 증대되고 있다¹⁾²⁾. 제스처 인식을 이용한 사용자 인터페이스는 HCI의 대표적 분야로, 물체의 조작이나 지시에 있어서 직관적일 뿐 아니라, 손의 공간상의 위치, 움직임 및 손의 형상 등의 몇 가지의 기본 구성요소에 의한 코드화(code)가 상대적으로 용이하여, 시각

장애자를 위한 수화인식, 원격지 로봇의 조정, 가상 현실 환경에서 물체의 조작 등 다양한 분야에서 폭넓게 이용되고 있다³⁾⁴⁾⁵⁾.

일반적으로 제스처를 인식하기 위해서는, 먼저 특징을 추출하고(특징추출), 미리 만들어둔 패턴(모델)과 특징을 매칭(인식)하는 방법이 필요하다. 이때, 특징을 쉽고 정확하게 추출하기 위해서 손에 센서를 부착하던가, 구별이 손쉬운 표식(marker)을 부착하거나 표식이 있는 장갑을 사용하여 촬영된 영상을 이용하는 경우가 많다. 그러나 이 방법은 접촉형의 센서나 마커를 신체에 부착하는 불편함과, 센서나

마커가 신체의 움직임에 구속하게되어 동작이 부자연스럽게 되는 등 몇 가지의 문제점이 있다.

또한, 센서나 마커를 이용하지 않고 실세계의 자연스러운 영상만을 이용하여 제스처를 인식하는 방법은, 본질적으로 편하고 자연스러운 상호작용이 가능하다는 장점이 있지만, 영상정보의 특징상 처리하여야 할 정보의 양이 매우 많아 속도가 늦고, 배경, 개인차, 조명조건 등의 환경의 변화에 민감하여 많은 제약이 필요하다⁽⁶⁾⁽⁷⁾⁽⁸⁾.

그러나 컴퓨터와의 보다 자연스러운 인터페이스를 위해서는 이 제약을 최소화하여 환경조건에 영향을 받지 않으며, 인식에 필요한 계산량이 적어 실시간 처리가 가능한 인식방법이 필수적이다.

따라서, 본 논문에서는 처리할 정보의 양을 획기적으로 줄여 저속의 연산장치를 이용하여도 실시간으로 처리 가능하며, 손의 색깔, 크기, 모양과 같은 개인차나 배율, 조명등과 같은 영상 입력환경에 영향을 받지 않아 범용성이 있는 제스처의 인식을 목표로 한다.

본 논문에서는 3D LUT를 이용하여, CCD 카메라에 의하여 입력된 2차원 컬러영상으로부터 손영역을 추출하고, 추출된 손영역에서 손의 경계선 및 중심을 구하고, 경계선과 중심간의 거리 정보만으로 구성된 1차원 정보의 주파수를 분석하는 것으로 각 손가락 끝의 위치를 인식하여 제스처를 인식한다(그림 1 참조).

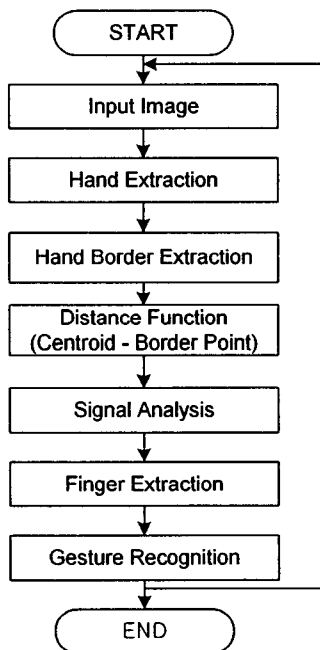


그림 1 시스템의 흐름도

2. 손 영역의 검출

제스처를 인식하기 위해서는 먼저, 입력된 영상으로부터 손영역을 추출하여야 한다. 컬러영상에서의 손영역의 추출은 피부색에 해당되는 색상마스크를 작성하고, 색상마스크와 일치하는 부분을 영상으로부터 추출하는 방법이 이용되고 있다⁹⁾. 그러나 이러한 수법은 임계값(threshold)에 따라 불완전한 마스크가 작성되어, 손영역 이외의 영역이 추출되거나, 손영역의 일부가 추출되지 못하여, 제스처 인식에 지장을 주는 경우가 빈번히 발생한다.

본 논문에서는, 손영역을 손의 내부와 외부의 색상히스토그램의 빈도(頻度)로부터 구하는 색상영역 추출법을 이용하여 추출한다. 또한, 손영역은 배경에 대하여 상대적인 움직임이 있으므로, 배경과의 차분 영상에 본 논문에서 제안하는 색상영역추출법을 적용하여 보다 정확하고 정밀하게 손영역을 추출하였다.

일반적으로 제스처동작 촬영시간 내에, 조명등의 변화는 거의 없다고 가정할 수 있다. 그러므로 먼저, 대상영상과 입력영상의 차분을 구하면 그림 2에서와 같이 대상영상에서의 손영역 이외의 대부분의 배경은 제거가 되지만 일부 제거가 되지 않는 부분이 존재하게 된다.

다음, 색상공간을 표현하는 3차원 룩업테이블(3D LUT)을 만들어두고, 제스처영상의 제1프레임에 후술하는 방법을 적용한 결과 얻어진 값을 LUT에 설정하여, 영상의 각 프레임으로부터 손영역을 추출한다. 또한, 색상공간에 대해서는 처리의 간편성과 실시간처리를 고려하여 RGB공간을 이용한다.

먼저, 입력영상의 제 1프레임에서 손의 영역을 분할하고, 손영역과 배경영역의 색상히스토그램을 계산한다. 색상히스토그램의 계산은, RGB공간의 각축을 각각 Q 분할한 3차원 히스토그램을 작성한다. 손영역의 히스토그램을 $M_j(j=1, \dots, N)$, $N=Q^3$, 배경영역의 히스토그램을 $B_j(j=1, \dots, N)$, $N=Q^3$ 으로 둔다. 다음, 식 1로 정의되는 통합히스토그램 C_j 및 식 2와 같이 정의되는 3D LUT를 구한다.

$$C_j = w_m \cdot M_j - w_b \cdot B_j \quad (1)$$

$$L_j = \max\left(\frac{C_j D}{C_{\max}}, 0\right) \quad (2)$$

이때, w_m w_b 는 가중치이며, C_{\max} 는 C_j 의 최대값이며, D 는 출력영상의 각 화소의 최대값이다.

식 1, 2에 의하여 영상상의 각 화소에 대하여 3D LUT를 참조하여 변환한다.

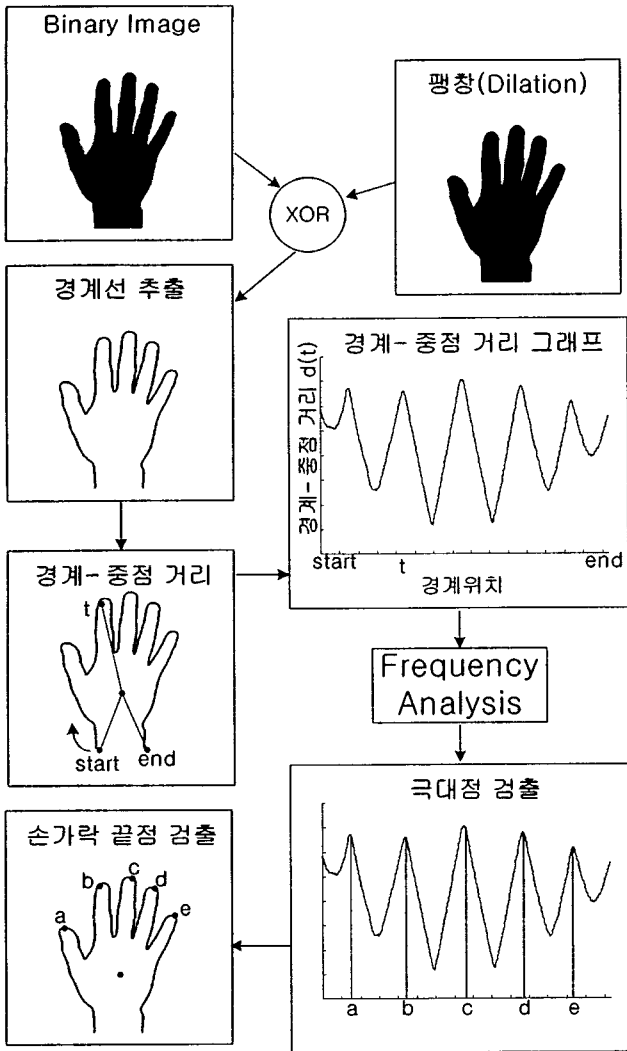


그림 2 손가락 끝점 검출 방법

이상의 방법으로 얻어진 손영역 추출결과를 그림2에 나타내었다.

3. 경계선에 의한 제스처 인식

그림 3은 본 논문에서 제안하는 손가락 끝점의 인식 방법이다. 먼저, 손영역이 추출된 영상에서 손영역 내부를 흑화소(black pixel), 외부를 백화소(white pixel)로 하는 이진화 영상을 구하고, 이진화된 영상에 있어서 각 흑화소를 1화소 팽창(dilation)시키고, 이진화 영상과 팽창영상의 배타적 논리합(XOR)을 구하면, 영상에 있어서 손의 경계선 부분만이 1의 값을 가지고 나머지는 모두 0의 값을 가지는 경계선 영상을 얻을 수 있다.

다음, 경계선과 중점간의 거리를 나타내는 함수 경계선-중점 함수를 식 3, 4를 이용하여 구한다.

$$X_c = \frac{1}{N_b} \sum_{i=1}^{N_b} X_i, Y_c = \frac{1}{N_b} \sum_{i=1}^{N_b} Y_i \quad (3)$$

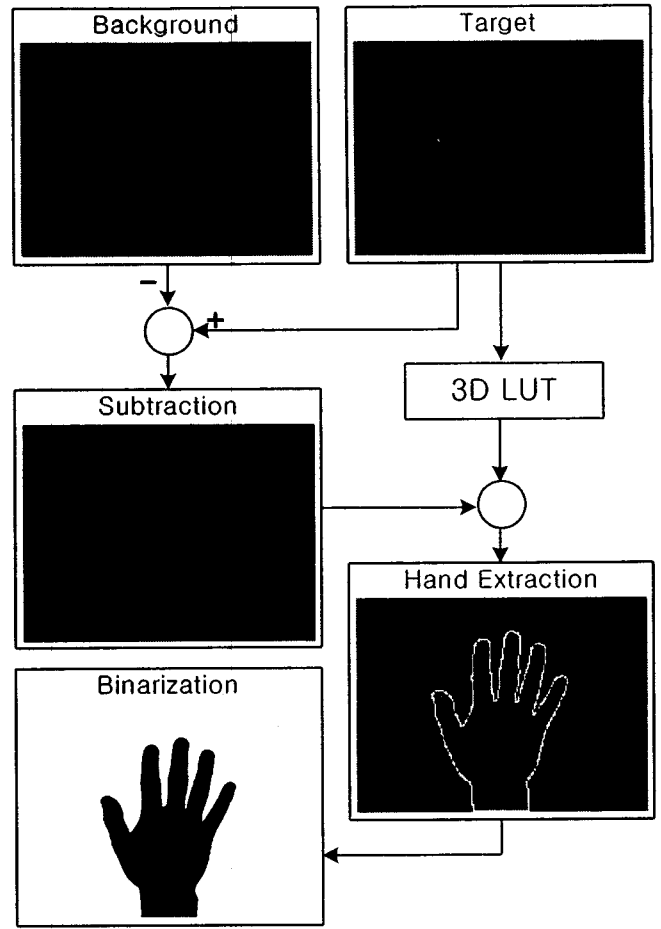


그림 3 손영역 추출

$$D_i = \sqrt{(X_i - X_c)^2 + (Y_i - Y_c)^2} \quad (4)$$

식 3에서 X_c, Y_c 는 손의 중점의 X, Y 좌표이고, X_i, Y_i 는 손의 경계선상의 임의의 화소의 X, Y 좌표이며, N_b 는 경계선상에 존재하는 화소의 수이다. 식 4에서 D_i 는 경계선-중심간의 거리이며, 식 4에서 얻어진 D_i 의 값을 이용해서 $D(i)=D_i$ 의 함수를 만들 수 있으며 이 함수를 그래프로 나타내면 경계선-중점 거리 그래프이다.

경계선-중점 거리 그래프 $D(i)$ 는 DFT(Discrete Fourier Transformation)이나 DCT(Discrete Cosine Transformation)등의 주파수 특성 분석 알고리즘을 이용한 해석을 통하여 그 극대점을 구할 수 있고, 극대점의 위치는 원래의 대상 영상에 맵핑(mapping)하면, 손가락의 끝점에 해당된다.

이상과 같은 방법으로 대상영상에서의 손가락 끝점을 구할 수 있고, 손가락끝점의 위치에 의하여 손의 위치나 자세는 결정되므로 손의 위치나 자세를 알 수 있으며, 손의 위치나 자세가 인식되면 손의 위치나 자세의 시간적 변화에 의하여 제스처를 인식할 수 있다.

4. 실험 및 고찰

본 논문에서 제안한 제스처 인식 시스템은 Pentium II PC상에서 C/C++언어를 이용하여 구현하였다. 영상입력 장치로는 Pulnix사의 TMC-7RGB CCD카메라를 사용하였으며, Matrox사의 MeteorII 보드를 이용하여 320×240의 해상도로 샘플링하여 사용하였다. 본 논문에서 제안한 제스처 인식 방법은 초당 30프레임 이상의 고속 인식이 가능하였고, 10가지 이상의 다양한 정적 상태인식이 가능하므로, 응용프로그램이 제공하는 동작들에 각각의 정적 제스처를 할당함으로써 사용자 인터페이스를 구현할 수 있다.

그러나, 정적 제스처만으로 인터페이스를 구성하는 것은 많은 기능을 제공하기 힘들고 확장이 어렵다는 단점이 있다. 따라서, 기본적인 정적 제스처와 간단한 구문론적 방법을 결합한 동적 제스처를 활용함으로써 보다 효율적이고 다양한 기능의 인터페이스를 제고할 수 있게 될 것이다⁹⁾¹⁰⁾.

5. 결론

본 논문에서는, 손의 모양이나 크기와 같은 개인 차나, 입력 환경에 영향을 받지 않고 실시간으로 손 동작을 인식 할 수 있는 손동작 인식방법을 제안하였다. 먼저, 손에 부착하는 센서나 마커 없이, CCD 카메라에 의하여 입력된 컬러영상에서, 컬러정보 및 동작정보를 이용하여 손영역을 추출하고, 추출된 손의 경계선 정보를 이용하여 경계선-중심 거리 함수를 생성하고, 생성된 함수의 주파수를 해석하여 각각의 손가락 끝 위치를 구하고, 손의 자세를 인식하여 제스처를 인식했다. 즉, 손의 중심과 경계선과의 거리를 특징으로 이용하여 손가락의 끝 부분에서는 상기 중심과 경계선과의 거리가 극대점을 이룬다는 원리를 이용함으로써 계산 시간을 획기적으로 단축할 뿐만 아니라 개개인의 손의 모델을 입력하지 않고도 손가락의 위치를 측정함으로써, 기존의 키보드나 마우스를 대신하는 새로운 사용자 인터페이스의 활용을 기대할 수 있다.

본 논문에서 제안하는 방법은 손동작 인식뿐 아니라 인간의 전체적인 자세의 인식에도 이용될 수 있을 것이다.

【참고문헌】

[1] V. I. Pavlovic, R. Sharma, and T. S. Huang,

"Visual Interpretation of Hand Gestures for Human-Computer Interaction : A Review", IEEE Trans. On PAMI., Vol. 19, No. 7, pp. 677-695, 1997.

[2] D. M. Gavrila, "The Visual Analysis of Human Movement: A Survey", Computer Vision and Image Understanding, vol. 73, No. 1, pp. 82-98, 1999.

[3] W. Stokoe, D. C. Casterline, and C. G. Groneberg, "A Dictionary of American Sign Language on Linguistic Principles", Linstok Press, London, 1976.

[4] K. Kanda, "A Computer Dictionary of Jpanese Sign Language: General Idea and Its Present Stage", The 5th International Symposium of Sign Language Research, pp.409-419, 1992.

[5] V. I. Pavlovic, R. Sharma, and T. S. Huang, "Gestural Interface to a Visual Computing Environment for Molecular Biologists", Proc. 2nd Conference on Automatic Face and Gesture Recognition, pp.30-35, 1996.

[6] K. Grobel and H. Hienz, "Video-Based Handshape Recognition using a Handshape structure Model in Real Time", Proc. International Conference on Pattern Recognition, pp.446-450, 1996.

[7] Y. Cui, D. L. Swets, and J. J. Weng, "Learning-Based Hand Sign Recognition using SHOSLIF-M", Proc. International Confrence on Computer Vision, Massachusetts, pp.631-636, 1995.

[8] A. F. Bobick and A. D. Wilson, "A State-Based Technique for the Summarization and Recognition of Gesture", Proc. International Conference on Computer Vision, Massachusetts, pp.382-388, 1995.

[9] 조오영, 김형곤, 고성제, 안상철, "인터랙티브 가상 환경을 위한 손 제스처 인식 시스템", 전자공학회논문지, Vol. 36, No. 4, pp.70-82, 1999.

[10] J. Park and I. W. Sandberg, "Universal Approximation Using Radial Basis Function Networks", Neural Computation, Vol. 3, pp.246-257, 1991.

참고 문헌

- 김영아, 김진관, 박수경, 오경자. (1997). 정서관련 어휘 분석을 통한 내적상태의 차원 연구. '97 한국감성과학회 연차학술대회 논문집, 209-214.
- 반세범, 한재현, 정찬섭. (1997). 표정-내적상태 상관모형을 위한 얼굴 DB 구축. '97 한국감성과학회 연차학술대회 논문집, 209-214.
- 한재현. (1997). 얼굴표정에 위한 내적상태 추정. 연세대학교 석사학위 청구논문.
- Andorfer, J.C. (1984). Affective Pattern Recognition and Schizophrenia. *Journal of Clinical Psychology*, 40:2, 403-409.
- Borod, J.C., Martin, C.C., Alpert, M., Brozgold, A. & Welkowitz, J. (1993). Perception of Facial Emotion in Schizophrenic and Right Brain-Damaged Patients. *The Journal of Nervous and Mental Disease*, 181:8, 494-502.
- Burch, J.W. (1995). Typicality Range Deficit in Schizophrenics' Recognition of Emotion in Faces. *Journal of Clinical Psychology*, 51:2, 140-152.
- Burch, J.W., & Pishkin, V. (1984). Family resemblance-category structure of joy and shame. *Journal of Clinical Psychology*, 40, 1136-1143.
- Dougherty, F.E., Bartlett, E., & Izard, C. (1974). Responses of schizophrenics to expressions of the fundamental emotions. *Journal of Clinical Psychology*, 30, 243-246.
- Gaebel, W., & Wolwer, W. (1992). Facial expression and emotional face recognition in schizophrenia and depression. *European Archives of Psychiatry and Clinical Psychology*, 28, 19-29.
- Heimberg, C., Gur, R.E., Erwin, R.J., Shatsel, D.L., & Gur, R.C. (1992). Facial Emotion Discrimination: III. Behavioral Findings in Schizophrenia. *Psychiatry Research*, 42:253-265.
- Kerr, S.L. & Neale, J.M. (1993). Emotion Perception in Schizophrenia: Specific Deficit or Further Evidence of Generalized Poor Performance? *Journal of Abnormal Psychology*, 102:2, 312-318.
- Morrison, R.L., Bellack, A.S. & Mueser, K.T. (1988). Deficits in Facial-Affect Recognition and Schizophrenia. *Schizophrenia Bulletin*, 14:1, 67-83.
- Rosch, E. (1978). Principles of categorization. In E. Rosch & B.B. Lloyd(Eds.), *Cognition and categorization* (27-48). Hillsdale, NJ: Erlbaum.
- Russell, J. A. (1994). Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies. *Psychological Bulletin*, Vol. 115, No. 1, 012-141.
- Salem, E.J., Kring, A.M., & Kerr, S.L. (1996). More Evidence for Generalized Poor Performance in Facial Emotion Perception in Schizophrenia. *Journal of Abnormal Psychology*, 105:3, 480-483.
- Walker, E., Marwit, S.J. & Emory, E. (1980). A Cross-Sectional Study of Emotion Recognition in Schizophrenics. *Journal of Abnormal Psychology*, 89:3, 428-436.
- Walker, E., McGuire, M., & Bettles, B. (1984). Recognition and identification of facial stimuli by schizophrenics and patients with affective disorders. *British Journal of Clinical Psychology*, 23:37-44.
- Wolwer, W., Streit, M., Polzer, U., & Gaebel, W. (1995). Facial Affect Recognition in the Course of Schizophrenia. *European Archive of Neuroscience Psychiatry*, 165-170.