

동작인식을 위한 배경 분할 및 특징점 추출 방법*

유휘중*, 김태영**

서경대학교 컴퓨터공학과

youhj86@naver.com, tykim@skuniv.ac.kr

A Background Segmentation and Feature Point Extraction Method of Human Motion Recognition

Hwi-Jong You*, Tae-Young Kim**

Dept. of Computer Engineering, SeoKyeong University

요 약

본 논문에서는 동작인식 위한 정확한 배경 분할 및 특징점 추출 방법을 제안한다. 배경 분할 과정에서는 먼저, HSV 입력 이미지를 RGB 색상 공간에서 HSV 색상 공간으로 변환한 뒤, H와 S 값에 대한 두 개의 임계치를 사용하여 살색 영역을 분할, 프레임간의 차영상을 이용하여 움직임이 있는 영역을 추출한다. 차영상에서 발생하는 잔상 영역을 제거하기 위하여 헤시안 어파인 영역 검출기를 적용하고, 잡음이 제거된 차 영상과 살색 영역의 이진화 영상을 이용하여 사람의 동작이 나타나는 영역을 분할한다. 특징점 추출 과정은 전체 영상을 블록 단위로 나눠서 각 블록 안에서 분할된 영상에 포함되는 픽셀들의 중점을 구하여 특징점을 추출한다. 실험 결과 복잡한 환경에서도 정확한 배경 분할과 사용자 동작을 대표하는 특징점 추출이 약 12 fps로 가능함을 알 수 있었다.

ABSTRACT

In this paper, we propose a novel background segmentation and feature point extraction method of a human motion for the augmented reality game. First, our method transforms input image from RGB color space to HSV color space, then segments a skin colored area using double threshold of H, S value. And it also segments a moving area using the time difference images and then removes the noise of the area using the Hessian affine region detector. The skin colored area with the moving area is segmented as a human motion. Next, the feature points for the human motion are extracted by calculating the center point for each block in the previously obtained image. The experiments on various input images show that our method is capable of correct background segmentation and feature points extraction 12 frames per second.

Keywords : Motion Recognition, Background Segmentation, Feature Point Extraction, Hessian Affine Region Detector, Augmented Reality Game

접수일자 : 2011년 01월 17일 일차수정 : 2011년 02월 24일 심사완료 : 2011년 03월 10일

교신저자(Corresponding Author) : 김태영

* 본 논문은 문화체육관광부 및 한국콘텐츠진흥원의 2010년도 콘텐츠산업기술지원사업(지정공모, 2-10-7602-001-10749-08-001) 사업의 연구결과로 수행되었음.

1. 서 론

동작인식은 사용자의 신체적인 움직임에 따라 감각적으로 상호작용할 수 있는 게임 기술에 적용될 수 있는 기술이다. 최근 동작인식 게임의 잠재력을 보여주는 여러 게임이 출시되고 있다. 그 예로 테이블 상에서의 자동차 경주 게임이나 볼링 게임 등[1]이 있다. Sony사의 PS3용 카드 게임인 Eye of Judgement는 2008년 출시된 게임으로 특정 카드를 인식해서 테이블 보드 상에서 3차원적으로 가상 캐릭터들을 보여준다. 또한 제한된 실내 공간에서 사용자의 움직임을 인식하는 게임[2]도 있다. 이 게임은 천장에 많은 마커들을 부착하여 사용자 손의 제스처를 인식하여 상호 작용한다. 보다 넓은 실외 공간에서의 게임의 한 예로는 Human PacMan[3]이라는 실외용 체험 게임이 있다. 이 게임은 gps와 관성 센서 등을 사용하여 사용자의 동작을 인식한다. 이와 같이 키보드나 조이스틱처럼 입력 기기가 아닌 신체적 움직임에 의한 입력 방법을 이용한 동작 기반의 인터페이스를 제공하여 사용자가 가상의 게임 환경에서 좀 더 몰입감 있는 게임을 제공하기 위한 여러 시도가 이뤄지고 있다.

이와 같은 동작 인식이 가능하려면 입력받은 영상에서 인식해야 되는 객체와 배경의 분할이 필요하고 분할된 객체의 대표점을 구해야 하는 문제점이 발생하게 된다. 배경 분할을 수행하기 위한 방법은 색 기반의 영상 추출[4], Saliency Map[5], Motion History Image[6], Boundary Energy[7]를 이용한 방법 등이 있다. 색 기반의 영상추출은 영상에서 사람의 특징을 나타낼 수 있는 영역을 색에 의하여 분할하는 방법으로 구현이 간편한 장점이 있으나 분할 정확도가 낮고, Saliency Map은 영상에서 가장 두드러지는 부분영역을 추출하는데 복잡한 영상에서는 원하는 영역의 추출이 어렵다는 단점이 있다. Motion History Image은 각 영상 프레임에 대하여 사람이 움직인 영역들을 누적시킴으로서 움직인 부분들만 찾아내는 방법으로 다른

움직이는 물체가 생기게 되면 인식의 문제점을 가지게 된다. Boundary Energy는 edge를 기반으로 edge의 복잡도가 높아지는 부분을 찾는 방법으로 손인식과 같이 영상에서 차지하는 부분이 많을 때 주로 사용한다. 하지만 동작인식과 같이 영상에서 차지하는 영역이 비교적 작은 영상에서는 동작인식 영역외의 부분에서도 edge의 복잡도가 높은 경우가 발생할 수 있으므로 동작인식에는 부적합한 방법이다.

본 논문에서는 기존 기법들의 문제점을 개선하여 복잡한 배경에서도 사람의 동작을 인식할 수 있는 배경 분할 및 특징점 추출 기법을 제안한다. 본 논문의 구성은 2장에서 배경 분할 방법, 3장에서 특징점 추출 방법을 설명하고 4장에서 제안하는 방법의 실험 결과를 기술한다. 마지막 5장에서는 본 논문의 결론과 향후 연구방향에 대하여 서술하도록 한다.

2. 배경 분할 방법

본 논문이 제안하는 배경 분할의 과정은 먼저, HSV 컬러 기반의 임계값을 이용하여 살색 영역을 분할한 뒤 이진화를 수행한다. 그 후 프레임간의 차 영상[8]을 이용하여 움직임이 있는 영역을 추출하고, 차 영상에서 발생하는 잔상 영역을 제거하기 위하여 헤시안 어파인 영역 검출기(Hessian Affine Region Detector)[9]를 적용한 후, 잡음이 제거된 차 영상과 살색 영역의 이진화 영상을 이용하여 사람의 동작이 나타나는 영역을 분리한다. [그림 1]은 본 논문에서 제안하는 알고리즘의 전체 과정을 도식화한 것이다.



[그림 1]

2.1 살색 검출을 이용한 분할

영상에서 받은 색은 일반적으로 RGB 모델로 표현된다. 그 값 중 R과 G의 값은 빛에 영향을 많이 받는 단점이 있어 본 연구에서는 RGB 영상을 HSV(Hue, Saturation, Value) 영상으로 변환하여 사용한다. H는 빛의 파장의 자체적인 시각적 특성을 나타내고 물체에 반사되어 나오는 파장의 색상 관점에서 나타낸 것이다. S는 색이 얼마나 순수한지를 나타내며, 흰색과 순수한 컬러와의 혼합비율에 의해 0 ~ 100%를 가진다. V는 빛의 밝기를 나타내며, 색에 나타난 빛의 반영 비율이 나타난다. 그러므로 실제 빛의 밝기에 민감하지 않은 H와 S 값을 사용하여 살색에 해당하는 영역을 검출하는데, 이때 이진화한 영상을 사용한다.

$$\begin{aligned}
 V &\leftarrow \max(R, G, B) \\
 S &\leftarrow \begin{cases} \frac{V - \min(R, G, B)}{V} & \text{if } \neq 0 \\ 0 & \text{otherwise} \end{cases} \\
 H &\leftarrow \begin{cases} 60(G - B)/S & \text{if } V = R \\ 120 + 60(B - R)/S & \text{if } V = G \\ 240 + 60(R - G)/S & \text{if } V = B \end{cases}
 \end{aligned}$$

(식 1) RGB를 HSV로 변환

2.2 차 영상을 이용한 분할

동작인식에서 필요한 영역은 사람이 움직이는 영역을 말한다. 그러므로 본 논문에서는 현재 영상과 이전영상의 차를 이용하여 움직이는 영역을 추출한다. 이를 차 영상(difference image)이라고 하며 차 영상을 얻는 식은 다음 (식 2)와 같다. (식 2)에서 픽셀의 움직임이 이전 영상과 다를 경우를 1이라고 표현하고, 다르지 않을 경우 0이라고 표현한다. 표현된 수식을 이용하여 영상을 다시 그리게 되면 실루엣 영상이 생성된다.

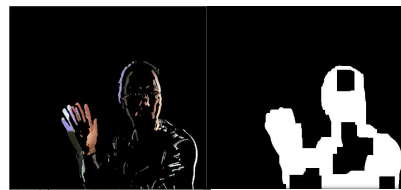
$$d(x,y) \begin{cases} = 1, & \text{when } |f_1(x,y) - f_2(x,y)| \geq T \\ = 0, & \text{otherwise} \end{cases}$$

(식 2) 차 영상

이진화된 실루엣 영역만을 가지고 영상에서 대표점을 추출하기는 어렵다. 따라서 영상에서 나타난 실루엣을 이진화 시킨 후 모폴로지의 닫기 연산을 수행한다. 우선 영상을 적절한 크기로 팽창을 시킨 후, 정확한 형태를 얻기 위해 침식을 시킨다. 이러한 연산을 반복하여 수행하면 보다 정확한 사람의 실루엣 영상을 얻을 수 있다.

$$\begin{aligned}
 (f \ominus b)(x) &= \min_{z-x \in D_f, z \in D_b} f(z-x) - b(z) \\
 (f \oplus b)(x) &= \max_{z-x \in D_f, z \in D_b} f(z-x) + b(z)
 \end{aligned}$$

(식 3) 모폴로지



(a) (b)

[그림 2] 차영상(a)과 모폴로지 영상(b)

2.3 헤시안 어파인 영역 검출기 기반 잡음

인식

현재 영상과 이전 영상을 비교하는 차 영상을 이용하여 배경을 제거 하고, 움직인 부분만 남게 되면 영상은 잔상이 생기며, 이러한 잔상이 남은 부분들을 잡음으로 인식한다. 본 논문에서는 차영상의 잡음을 제거하기 위해 SURF[9]에서 사용된 헤시안 어파인 영역 검출기를 이용한다.

$$H(x, \sigma) = \begin{bmatrix} L_{xx}(x, \sigma) & L_{xy}(x, \sigma) \\ L_{xy}(x, \sigma) & L_{yy}(x, \sigma) \end{bmatrix}$$

(식 4) Hessian Matrix

먼저 입력 영상에서 헤시안 어파인 영역 검출기를 수행하여, 기술자 $I(x, y)$ 를 찾고 검출된 기술자 집합을 (식 5)의 w_i 에 누적한다. 이때 첨자 i 는 프레임의 숫자를 나타낸다.

$$w_i = \begin{cases} = 1, & \text{if}(I_{i-1} = I_i) \\ = 0, & \text{otherwise} \end{cases}$$

(식 5) 기술자 비교

(식 5)에서 누적된 기술자 성분인 w_i 에서 이전 영상의 같은 위치에 같은 기술자가 나타나게 되면 1로 표기하고 기술자가 나타나지 않으면 0으로 표기한다. 표기된 기술자의 누적 숫자가 일정수치를 넘게 되면 고정된 배경인 잔상으로 인식하고 그 부분을 차영상에서 제외시킨다.

$$w_0 + w_1 + w_2 + w_3 + \dots > 20$$

(식 6) 기술자 성분 누적

본 논문에서는 (식 6)과 같이 20 프레임 이상 같은 기술자를 가지는 특징점의 영역들은 영상에서 제외시킨다. 여기서 임계값을 20 프레임으로 정한 이유는 사람이 실시간으로 인식하기 최적화된 프레임 수이기 때문이다.

3. 특징점 추출 방법

본 장에서는 2장에서 제시된 배경분할 방법에 의하여 생성된 세 가지 영상으로 부터 특징점을 구한다. 먼저, 구하고자 하는 객체의 크기에 비례하여 블록을 분할한다. 객체보다 큰 영역을 블록의 크기로 잡게 되면 동작인식과 같은 두 가지 이상의 영역에서 특징점을 구할 때 겹치는 부분이 발생하여 인식의 어려움이 발생하게 된다. 또한, 인식할 영역보다 작은 블록을 잡게 되면 필요이상의 특징점이 추출되는 어려움이 있다.

(식 7,8)과 같이 분할된 블록안에서 각각 살색 검출 영상($S_i(x,y)$), 모폴로지가 된 차 영상($D_i(x,y)$), 고정된 기술점($H_i(x,y)$)의 AND 연산의 영상($M_i(x,y)$)를 구하고 $M_i(x,y)$ 의 중점($C_i(x,y)$)을 구한다.

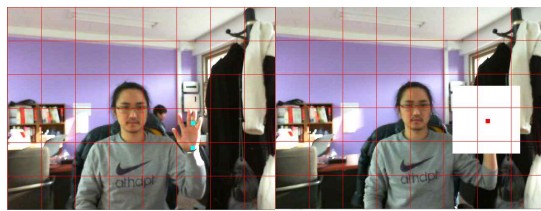
$$M_i = S_i \cdot D_i \cdot H_i$$

(식 7) 특징 영역 추출

중점은 각 블록 별로 3x3 크기의 영역을 비교하게 되는데 이때, 3x3 크기를 갖는 블록의 영역 안에 여러 개의 중점이 있게 되면 중점들 간의 특징점 N 을 구할 수 있다.

$$N = (C_0 + C_1 + C_2 + \dots + C_8) / 9$$

(식 8) 중점 간의 특징점 추출



[그림 3] 영상블록화를 이용한 특징점추출

4. 실험 및 결과

본 논문의 실험 환경은 Window7에서 Visual Studio 2005와 OpenCV를 사용하였고 Intel(R) Core(TM) i5 CPU 760 @ 2.8GHz, 4.00GB RAM의 하드웨어에서 실험하였다. 실험 결과 표1에서 보는 바와 같이 배경이 복잡한 영상에서도 높은 인식률로 약 12fps의 속도로 배경 분할을 통한 특징점이 추출됨을 알 수 있었다.

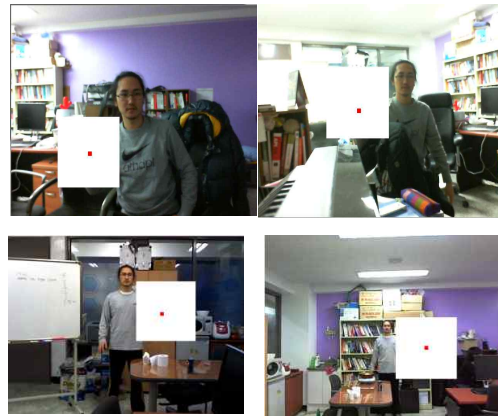
[표 1] 각 단계별 평균 수행 시간

총 시간	배경 분할			특징점 추출
	살색 이진화 (a)	차 영상의 모폴로지 (b)	헤시안 어파인 영역 검출기 (c)	
				
96ms	9ms	9ms	76ms	2ms미만

[표 1]은 본 논문에서 제시하는 알고리즘의 단계별 평균 수행시간을 보여주는 표이다. (a)는 입력받은 영상을 HSV로 변환한 뒤 S와 V를 이용하여 살색부분만을 추출한 후, 이진화한 영상이고 (b)는 차영상을 이용하여 실루엣을 추출하고 그 후, 추출된 실루엣에 모폴로지를 적용하여 영역을 확장한 영상이다. (c)헤시안 어파인 영역 검출기를 이용하여 움직이지 않는 영상을 제거한 영상이다. (d)는 (a),(b),(c)의 영상을 이용하여 영상을 블록별로 나눠서 특징점을 추출한 영상이다. 일반적인 헤시안 어파인 영역 검출기를 사용하게 되면 속도의 저하 문제를 야기시킨다. 이러한 문제를 해결하기 위하여 본 논문에서는 가우시안 블러를 적용하여 영상에 블러를 발생시켜 기술자의 개수를 줄여서 속도를 향상시켰다. 본 논문에서는 (a)영상만을 사용했을 경우 손뿐만 아니라 뒤에 나타나는 살색들

도 인식됨을 알 수 있다. 이러한 문제를 제거하기 위해 차영상을 사용하였고 차 영상에서 나타나는 정확한 손의 위치를 얻기 위해 헤시안 어파인 영역 검출기를 사용하였다.

[그림 4]는 여러 가지 복잡한 배경에서 촬영한 사진이다. 다른 위치, 배경, 조명에서 다른 영역을 움직이지 않는 상태에서 손을 움직이고 있는 그림이다. 객체의 크기가 다르지만 블록 안에 객체가 들어가기 때문에 특징점이 올바르게 추출됨을 확인할 수 있다.



[그림 4] 복잡한 배경에 대한 결과 영상

[그림 5]는 추출된 특징점을 이용하여 구현된 스켈레톤의 얼굴과 손과 다리에 매칭을 시킨 영상이다. 스켈레톤을 얼굴과 손과 다리에 매칭을 한 뒤, 손이 아래서 위로 움직이게 되면 매칭된 스켈레톤이 손의 움직임을 트래킹하는 모습을 담은 영상이다.



[그림 5] 추출된 특징점을 이용하여 스켈레톤과 매칭된 영상

5. 결론 및 연구 방향

본 논문에서는 인체의 살색과 유사한 색이 배경에 존재하는 경우, 인체의 모션을 인식할 수 있도록 하는 배경분할 및 특징점을 추출 방법에 대하여 제안하였다. 제안된 방법은 배경에서 차 영상을 이용하여 고정된 물체를 제거 하였고, 헤시안 어파인 영역 검출기를 사용하여 차 영상에서 잔상이 남는 문제를 해결하였다. 잡음이 제거된 차 영상과 살색 영역의 이진화 영상을 이용하여 사람의 동작이 나타나는 영역을 분할한다. 특징점 추출 과정은 전체 영상을 블록 단위로 나눠서 각 블록 안에서 분할된 영상에 포함되는 픽셀들의 중점을 구하여 특징점을 추출한다.

향후 연구로 살색검출의 동적임계값 적용 방법과 특징점 추출을 위한 블록크기의 동적 계산방법에 대한 연구를 수행하고자 한다.

참고문헌

[1] <http://www.eyefofjudgment.com/>

[2] 김기영, 이민경, 박영민, 이종원, 우운택, "ARPushPush: 실내 환경 증강 현실 게임," KHCI, pp.354-359, 2005.

[3] Cheok, A.D., Fong, S.W., Goh, K.H., Yang, X., Liu, W., Farzbiz, F., Li, Y., "Human Pacman: Amobile entertainment system with ubiquitous computing and tangible interaction over a wide outdoor area," Mobile HCI 2003, pp. 209 - 224, 2003.

[4] Vezhnevets V, Sazonov V, Andreeva A, "A Survey on Pixel-Based Skin Color Detection Techniques", GraphiCon, 2003.

[5] Fritz, Seifert, Paletta, Bischof, "Entropy based saliency maps for object recognition", ECOVISION, 2004

[6] Takehito OGATA, Joo Kooi TAN, "High-Speed Human Motion Recogniton Based on a Motion History Image and an Eigenspace", IEICE, VOL.E89-D, 2006.

[7] 박상윤, 이용주 "복잡한 영상에 강인한 손동작 인

식 방법", 멀티미디어 학회 논문지, 제13권, 제7호, 한국연구재단 등재 학술지, pp. 1000~1015, 2010년 7월.

[8] Cao Xin-yan, Liu Hong-fei, Zou Ying-yong, "Gesture Segmentation Based on Monocular Vision Using Skin Color and Motion Cues", IASP, 2010

[9] Herbert Bay, Andreas Ess, Tinne Tuytelaars, Luc Van Gool "SURF: Speeded Up Robust Features", Computer Vision and Image Understanding (CVIU), Vol. 110, No. 3, pp. 346-359, 2008.



유 휘 중 (You, Hwi Jong)

2010년-현 재 서경대학교 컴퓨터 공학과 학사 과정

관심분야 : 네트워크, 보안, 컴퓨터그래픽스, 영상처리



김 태 영 (Kim, Tae Young)

1991년 2월 이화여자 대학교 전자계산학과 학사
1993년 2월 이화여자 대학교 전자계산학과 석사
1993년 3월-2002년 2월 한국통신 멀티미디어 연구소
선임 연구원

2001년 8월 서울대학교 컴퓨터 공학부 박사
2002년 3월-현 재 서경대학교 컴퓨터공학과 부교수

관심분야 : 실시간 렌더링, 모바일3D, 증강현실, 볼륨
그래픽스