

감정 표현이 가능한 실시간 반응형 그림자 아바타

임양미[†], 홍의석^{**}, 이재원^{***}

요 약

본 논문에서는 사용자의 동작에 대한 반응으로 표정을 변화하여 감정을 표현할 수 있는 실시간 반응형 그림자 아바타인 RISA(Real-time Interactive Shadow Avatar)를 제안한다. 아바타 형태는 사용자의 실사로 부터 실시간으로 추출한 가상의 그림자를 사용하며, 손동작의 유형에 따라 변화하는 표정 애니메이션이 그림자의 얼굴 위치에 겹쳐지도록 하였다. 가상 그림자의 추출을 위해서는 배경 차분화 기법을 사용하며, 머리 위치 및 손동작의 추적 및 유형 탐지를 위해 단순화된 영역 단위 추적 기법을 사용하였다. 또한 표정의 자연스러운 변화를 표현하기 위해 표정 애니메이션은 기존의 동작 이모티콘보다 많은 수의 애니메이션 프레임들을 사용하는 변형된 모핑 기법을 적용하였다. RISA는 인터페이스 미디어 아트 분야에 직접 응용될 수 있을 것이며, RISA에 적용된 탐지 기법은 향후 입력 장치의 간결성이 요구되는 DMB나 카메라폰 등을 위한 대체 인터페이스에도 활용될 수 있을 것이다.

A Real-time Interactive Shadow Avatar with Facial Emotions

Yangmi Lim[†], Euuseok Hong^{**}, Jaewon Lee^{***}

ABSTRACT

In this paper, we propose a Real-time Interactive Shadow Avatar(RISA) which can express facial emotions changing as response of user's gestures. The avatar's shape is a virtual Shadow constructed from the real-time sampled picture of user's shape. Several predefined facial animations overlap on the face area of the virtual Shadow, according to the types of hand gestures. We use the background subtraction method to separate the virtual Shadow, and a simplified region-based tracking method is adopted for tracking hand positions and detecting hand gestures. In order to express smooth change of emotions, we use a refined morphing method which uses many more frames in contrast with traditional dynamic emoticons. RISA can be directly applied to the area of interface media arts and we expect the detecting scheme of RISA would be utilized as an alternative media interface for DMB and camera phones which need simple input devices, in the near future.

Key words: Interactive Shadow Avata(반응형 그림자 아바타), Detecting Scheme(탐지 기법)

1. 서 론

웹을 통한 사용자간의 커뮤니케이션이 일반화됨에 따라 무수히 많은 인터넷 사용자들이 텍스트나

파일 또는 오디오나 비디오 콘텐츠들을 채팅 도구 등을 이용하여 주고받고 있다. 또한 웹캠이 PC 환경을 구성하는 중요한 요소 중 하나가 됨에 따라 카메라에 기반한 상호 커뮤니케이션 기술들도 여러 방법

※ 교신저자(Corresponding Author): 홍의석, 주소: 서울시 성북구 동선동3가 성신여자대학교(136-742), 전화: 02)920-7369, FAX: 02)920-7157, E-mail: hes@sungshin.ac.kr

접수일: 2006년 11월 9일, 완료일: 2007년 2월 26일

[†] 중앙대학교 첨단영상대학원 박사과정

(E-mail: yosimi@chol.com)

^{**} 정회원, 성신여자대학교 컴퓨터정보학부 교수

^{***} 성신여자대학교 컴퓨터정보학부 교수

(E-mail: jwlee@sungshin.ac.kr)

※ 이 논문은 2006년도 성신여자대학교 학술연구조성비 지원에 의하여 연구되었음

으로 발전되고 세분화 되고 있다. 이중 하나로 사용자의 여러 몸동작에 기반한 커뮤니케이션도 관심이 높아지고 있다. 몸동작에 기반한 시스템은 카메라를 통해 사용자의 몸동작이나 얼굴 표정, 걸음걸이 등을 인식하여 적당한 응용 분야에 적용하는 시스템으로 감시 카메라를 통하여 범죄자를 자동 인식하는 스마트 감시 분야, 사용자의 손동작이나 몸동작을 해석하여 컴퓨터 입력으로 인식하는 고급 사용자 인터페이스 분야[1], 몸동작을 이용한 커뮤니케이션 분야로 나눌 수 있다. 몸동작에 기반한 커뮤니케이션의 예로는 채팅룸, 게임, 캐릭터 애니메이션, 원격 회의 등을 구현한 가상현실 시스템을 들 수 있으며 언어 장애자들을 위한 동작기반 기호언어를 통한 커뮤니케이션도 이에 해당된다. 몸동작에 기반한 상호작용에 관한 연구가 많이 이루어졌지만 아직 분할(segmentation), 모델링, 폐쇄 처리(occlusion handling)와 같은 많은 문제들에 대한 연구가 필요한 상태이다[2].

본 논문에서는 간단하지만 인터넷 상에서 몸동작 기반 상호작용의 도구로 매우 유용하게 사용할 수 있는 실시간 반응형 그림자 아바타인 RISA (Real-time Interactive Shadow Avatar)를 제안한다. RISA는 기존의 아바타들과는 다음과 같은 면에서 차이가 있다. 첫째 RISA의 움직임은 웹캠 앞에 있는 사용자의 실제 동작으로부터 얻어낸 가상 그림자이다. 그러므로 실제 그림자를 얻기 위한 스팟 라이트 조명이라든가, 빛과 공간에 대한 일정한 제약 조건 등이 필요 없다. 둘째 RISA의 얼굴 표정은 사용자의 손동작에 따라 실시간으로 변화한다. 이의 구현을 위해 사용자의 동작에 따라, 실시간 애니메이션 기법에 의해 미리 만들어진 표정 이미지들을 그림자와 결합하였다. 셋째 사용자의 움직임에 따라 생성된 RISA 정보들은 네트워크를 통해 커뮤니케이션하는 상대방에게 키보드나 마우스 같은 입력장치 없이 자연스럽게 보내진다는 것이다. 이로써 보다 자연스러운 커뮤니케이션이 가능해진다.

RISA의 구현은 크게 두 단계로 나누어진다. 가상 그림자를 얻는 첫 번째 단계와 사용자의 얼굴 위치와 손동작을 추적 판별하여 적당한 표정 애니메이션을 가상 그림자에 결합하는 두 번째 단계이다. 가상 그림자를 얻는 첫 번째 단계를 위해서는 정적 배경에서 동작 분할 기법으로 많이 사용하는 배경 차분화 기법

(background subtraction method)을 사용하였다. 이 기법에서는 동작이 일어난 정확한 영역을 추출하기 위해 현재 이미지와 배경 이미지의 차이를 픽셀 단위로 구한다. 이 과정에서 빛의 산란과 실제 그림자로 인해 노이즈가 발생할 수 있는데 이들을 해결하기 위해 본 논문에서는 네가지의 새로운 이미지 처리 기법을 적용하였다.

RISA 구현의 두 번째 단계를 위해서 머리와 손의 위치를 추적하고 손동작의 유형을 판별하기 위한 휴리스틱 알고리즘을 고안하여 적용한다. 이 알고리즘에서는 머리와 손 등 동작을 이루는 요소들이 불록 형태라는 점에 착안하여 몸을 이루는 각 픽셀들에 대해 인접 픽셀들을 검사하여 머리와 손의 중심 위치를 추적하는 방식을 택한다. 즉 사람의 몸을 큰 덩어리로 보고 돌출 부분(블록 부분)을 머리와 팔로 구분하여 추적한다. 이로 인해 사람의 몸을 머리, 몸통, 팔, 다리 등으로 분리한 후 각 부분을 추적하는 기존의 알고리즘들에 비해 분리 과정을 없앴으로써 빠른 속도로 위치를 추적할 수 있으며, 또한 위치의 추적과 동시에 손의 방향이 탐지됨으로써 동작 탐지를 위한 별도의 알고리즘을 요구하지 않는다.

본 연구와 관련된 연구로는 물리적 장치를 인터페이스 컴포넌트로 사용하는 프레임워크에 대한 연구 [3,4], 배경이미지와 전경이미지를 분리하기 위한 알고리즘 연구[5-7], 몸동작 탐지에 관한 연구[8-10] 등을 들 수 있다. 하지만 RISA와 같이 감정 표현이 가능한 동적 아바타에 대한 연구는 없는 실정이다. RISA의 사용은 사무실이나 가정의 방과 같은 곳에서 웹캠을 사용하는 일상적인 환경에서 이루어지므로 [3,4]에서와 같이 물리적인 장치를 이용하지 않는다. 또한 일상적인 환경의 배경 모델은 매우 정적이라는 점에서 특별한 배경 모델을 가정한 [5-7]의 연구와도 차이가 있으며 이들이 RISA에서 요구하는 그림자 노이즈를 제거하는 데 문제점이 있으므로 본 연구에서는 네가지의 노이즈 제거 알고리즘을 사용하였다. 기존의 몸동작 탐지 연구는 몸의 모든 영역을 탐지하며 분리 과정의 포함이라는 문제가 있기 때문에 본 연구에서는 앞에서 기술한 바와 같이 새로운 탐지 알고리즘을 제안하였다.

RISA는 아바타의 용도로 설계되었으나, 가상공간에 투영된 그림자 및 표정 애니메이션과 사용자 자신이 상호작용하는 놀이 도구로 사용될 수 있으며, 사

이러한 공간 상에서 통신 상대자와의 감정 교환의 도구로 사용될 수도 있다. 이를 이용한 채팅은 사용자의 실제 형태를 이용함으로써 기존의 아바타보다 훨씬 동적이고 친숙하게 느껴지며, 실제 영상을 추상화한 동적 아바타의 형태를 띠므로써 실제 영상에 없는 감정이나 아트적인 요소가 가미되어 있다는 장점이 있다. 그러므로 RISA는 기존의 인터넷 채팅 프로그램에 감정 채팅의 기법으로 적용될 수 있으며, 연구에서 제안한 탐지 알고리즘들은 실시간 미디어 아트 분야에 적용될 수 있다.

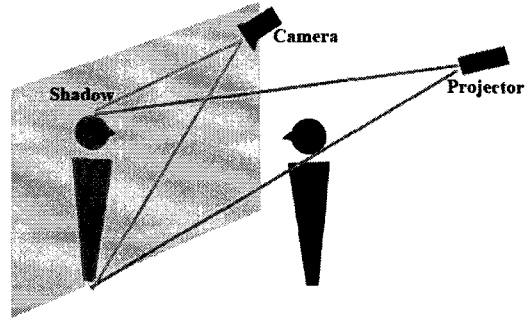
2장에서는 그림자를 얻기 위한 기존 환경과 본 연구에서 사용한 환경을 비교하고 가정된 환경에서 사용자의 윤곽을 추출하여 가상 그림자를 만드는 기법을 설명한다. 3장에서는 머리와 손의 중심 위치 탐지와 손 동작 판별 알고리즘을 언급하고, 4장에서는 손 동작에 따른 얼굴 표정 애니메이션 맵핑에 대해 기술한다. 5장은 알고리즘의 성능에 영향을 미치는 파라미터들을 변형하면서 구현된 RISA를 실험한 결과를 보이고, 6장에서는 결론과 향후 연구과제에 대해 기술한다.

2. 가상 그림자의 추출

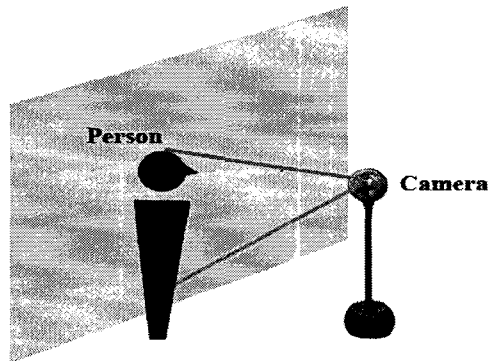
2.1 공간 및 조명 환경

일반적인 그림자 생성 기법은 그림 1의 (a)와 같이 프로젝터나 할로젠 조명을 사용하여 빛을 피사체에 쏘아서 벽면에 생긴 그림자를 위에서 촬영을 하는 방법을 사용한다[11]. 아트작가들에 의해 많이 사용되는 이러한 기법의 장점은 그림자가 2D 오브젝트이므로 빛의 반사가 일정하다는 것이다. 2D로 촬영된 그림자 영상은 테두리가 선명하고 깨끗하게 컴퓨터 데이터로 들어와 영상 노이즈가 거의 없는 것이 특징이다. 단점은 그림자를 투영해서 찍어야 함으로 공간이 많이 필요하고 조명이 항상 어두워야 하는 조건이 필요하다는 것이다.

RISA의 사용 환경은 일상적인 PC 사용 환경을 전제로 하므로 본 연구에서는 기존 방법의 제약점을 해결하기 위해 그림 1의 (b)와 같이 3D 오브젝트 영상을 직접 카메라에 담아 영상을 이진화 시켜 오브젝트를 2D 그림자로 추출하였다. 이 같은 기법은 작은 공간에서도 가능하며 특별한 조명 없이 형광등이나 백열등을 사용하여 가능하지만 산란광이나 반사광



(a) 일반 그림자 추출 환경



(b) 가상 그림자 추출 환경

그림 1. 그림자 추출 환경

에 의한 노이즈가 발생할 수 있다. RISA의 가상 그림자 생성 시 이러한 노이즈는 몇가지 영상처리 기법을 통하여 해결하였다.

2.2 가상 그림자 추출 기법

2.2.1 차영상 구하기

사용자의 그림자 모양 추출을 위해 배경 차분화(background subtraction) 기법을 사용한다. 이 기법의 기본적인 아이디어는 사용자가 포함된 이미지와 사용자가 배제된 초기 배경 이미지의 차이를 구하는 것이다. 배경 차분화 기법은 사물의 동작을 실시간으로 추적하는 응용 분야에 널리 사용되는 기법이지만, 조명이나 외부 요인 등에 의해 야기되는 화면상의 동적인 변화에 매우 민감하다는 문제점이 있다.

이를 해결하기 위해 몇가지 배경 모델과 수정된 기법들이 제안되었다[2]. 이들 중 간단하면서도 많이 사용되는 기법은 최근의 일정 시간 간격 동안의 이미

지들을 이용한 근사 이미지를 구하여 대표 배경 이미지로 사용하는 것이다. Yang과 Levine의 연구에서는 이미지들을 구성하는 각 픽셀들의 평균값보다는 중앙값이 더욱 좋은 결과를 보인다는 것을 보이고, 이를 이용하여 대표 배경 이미지를 구하는 알고리즘을 제안하였다[7]. 본 연구에서는 이에 기반 하여 대표 배경 이미지의 픽셀 값 B_{jk} 를 다음과 같은 공식에 의해 구한다.

$$B_{jk} = \text{median}(I_{jk}^1, I_{jk}^2, \dots, I_{jk}^N) \quad (1)$$

$I_{jk}^1, I_{jk}^2, \dots, I_{jk}^N$ 는 최근 N개의 이미지 시퀀스 픽셀들이고 (j, k)는 픽셀의 위치이다. 차영상을 얻기 위해서는 대표 배경 이미지를 얻기 위한 중앙값도 필요하지만 차영상에 속하는 픽셀로 판별하기 위한 임계치 값이 필요하다. 즉 사용자가 포함된 이미지와 대표 배경 이미지의 차이를 각 픽셀별로 구하여 차이가 임계치를 넘게 되면 해당 픽셀은 차영상에 속하게 되는 것이다. 최소 중앙값 자승법(least median squares method)[6]을 비롯한 몇 가지 기법들이 제안되었지만, 본 연구에서는 제한된 정적인 환경을 가정하고 있으므로 상대적으로 간단한 단일 임계치(uniform threshold) 방식을 사용한다.

대표 배경 이미지와 임계치가 결정되고 나면, 사용자가 포함된 이미지로부터 차영상을 구한다. 차영상을 구성하는 픽셀들을 결정하기 위해 다음과 같이 d_{jk} 를 정의한다.

$$d_{jk} = \begin{cases} 1, & \text{if } |I_{jk}^i - B_{jk}| > T \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

I_{jk}^i 는 사용자가 포함된 이미지의 픽셀이고 B_{jk} 는 대표 배경 이미지의 픽셀이며 T 는 임계치이다. d_{jk} 값이 1인 픽셀들의 색상은 가상 그림자의 전경 색상인 검은색으로 변환되고 나머지 픽셀들은 흰색으로 변환된다.

2.2.2 노이즈 제거

RISA는 그림자 객체를 실시간 애니메이션으로 이용하기 때문에 확실한 윤곽선 실루엣을 요구한다. 그러므로 기존의 배경 차분화 기법을 기반으로 차영상을 구한 후 보다 개선된 가상 그림자를 얻기 위해

몇 가지 노이즈 축소 기법을 적용한다.

(1) 가우시안 노이즈 제거

잘못된 점들로 나타나는 가우시안 노이즈를 제거하기 위해 마스크를 이용하는 희선 기법을 사용한다. 각각의 픽셀은 픽셀을 둘러싸고 있는 마스크에 포함된 전체 픽셀들의 평균값으로 대체된다. 입력 이미지가 흑백이므로 이 과정을 거친 이미지는 회색도(gray level)의 픽셀들을 포함하게 된다. 따라서 간단한 히스토그램 기법에 의해 계산된 임계치를 사용하여 픽셀들의 색상을 다시 흑백으로 변환한다.

(2) 윤곽선 정리

가우시안 노이즈를 제거한 후 더욱 선명한 사용자 윤곽선을 얻기 위해 확장과 침식 기법을 수차례 반복하여 윤곽선에 근접한 픽셀들 중 이상 픽셀들을 제거한다. 이 과정은 픽셀들 사이의 매우 작은 구멍들을 메우는데도 아주 효과적이지만, 반복 횟수에 따라 계산 비용이 매우 많이 드는 문제점이 있다. 그러므로 본 연구의 실험에서는 반복 횟수를 5번 정도로 제한하여 사용하였다.

(3) 외부 검은 영역 제거

다음 단계에서는 그림자의 외부에 남아있는 격리된 검은 영역들을 제거한다. 이 기법은 사용자의 몸에 해당하는 픽셀들은 몸의 중심부에 집중되어 있다는 것, 즉 몸은 연결되어 있다는 특성을 이용한다. 이를 위해, 동일 색상의 인접 픽셀들로 구성된 영역들을 식별하고 식별된 각 영역의 크기를 계산하는 알고리즘을 사용한다[12]. 식별된 영역 중 가장 큰 영역은 그림자 영역이므로 이외의 모든 영역을 배경색인 흰색으로 변환한다.

(4) 스위핑

노이즈 제거의 마지막 단계로, 스위핑 과정을 수행한다. 이 과정의 목적은 가상 그림자 내부의 구멍을 메우는 것이다. 이미지의 각 행마다 인접한 검정 픽셀들 이후에 나타나는 흰색 픽셀들에 대하여 색상을 검게 변환하는 스위핑을 수행한다. 이러한 스위핑은 수행 시간이 매우 작다는 장점이 있으나 잘못된 스위핑은 그림자 외부에 노이즈가 많은 검은 선들을 만들 수 있다. 이를 방지하기 위해 그림자의 무게중심을 기준으로 스위핑의 범위를 제한하는 등의 구현상 제약을 두었다.

3. 머리 및 손동작 탐지

3.1 머리 탐지

본 연구에서는 사용자의 몸에서 머리 부분, 손 부분과 손동작의 탐지가 필요하다. 인간의 몸동작의 추적 및 탐지에 관해서는 다양한 연구들이 수행되어 왔다. 이들을 추적 과정에 사용된 기법의 관점에서 분류하자면 모델 기반 추적(model-based tracking), 영역 기반 추적(region-based tracking), 테두리 기반 추적(active contour based tracking), 특징 기반 추적(feature-based tracking)으로 분류할 수 있다 [2]. 이 중 특징 기반 추적은 객체 전체를 추적하는 대신 두드러진 점이나 선 등의 부특징(sub-feature)들을 사용한다는 점에서 다른 기법들과 구분된다[9]. 본 연구에서는 수정된 특징 기반 추적 기법을 사용한다.

기존의 특징 기반 추적 기법은 움직이는 몸의 각 부분을 분리하여 인식하는 과정과 분리된 몸의 각 부분의 동작들을 추적 탐지하는 기법들이었다[10]. 이의 문제점은 추적 컴퓨팅 시간이 너무 많이 걸려 본 연구의 환경인 일상적인 PC를 사용하는 실시간 환경에는 부적합하다는 것이다. 그러므로 본 연구에서는 일상적인 환경에 적합하도록 새로운 경험적 부특징 기반 추적 알고리즘을 작성하였다. 제안 알고리즘은 컴퓨팅 시간을 빠르게 하기 위하여 머리와 팔에 대한 분리 과정을 수행하지 않고 사용자 몸의 볼록 영역을 탐색하는 볼록 영역 검출 아이디어를 사용하였다.

제안 알고리즘의 기본 아이디어는 프레임을 구성하는 각 픽셀에 대해 해당 픽셀을 가까운 거리에서 둘러싸고 있는 몇 개의 픽셀들만을 분석함으로써 머리와 손의 중심 위치를 찾는 것이다. 동작 인식을 위해서는 손을 들거나 팔을 오른쪽 또는 왼쪽으로 뻗는 것과 같이 미리 정의된 몇 가지 동작들만을 인식하고자 하므로, 머리카나 손과 같이 탐지하고자 하는 영역들은 볼록(convex) 형태라고 가정할 수 있다. 이러한 가정 하에서는 특정 픽셀과 일정한 거리를 가진 주변 픽셀들의 색상이 배경색인 경우, 해당 픽셀을 중심 위치로 간주하는 것이 가능해진다. 그림 2의 (a)에서 머리의 중심 픽셀에서 어느 정도 떨어진 외부 사각형 픽셀들은 배경색 즉 흰색이며 이보다 가까운 내부 사각형 픽셀들은 검은색이다.

머리의 중심 위치 (x, y) 를 찾는 알고리즘은 공식 (3)을 사용한다.

$$\begin{aligned} \text{Color of } (x + \alpha_x, y + \alpha_y) \text{ is white, for } (\alpha_x, \alpha_y) \in \\ O - \{(0, -a), (0, 0)\} \text{ and} \\ \text{Color of } (x + \alpha_x, y + \alpha_y) \text{ is black, for } (\alpha_x, \alpha_y) \in \\ I, \\ \text{where } O = \{ (x, y) \mid x = -a, 0, a \text{ and } y = -a, \\ 0, a \}, \\ I = \{ (x, y) \mid x = -\beta, 0, \beta \text{ and } y = -\beta, \\ 0, \beta \}, \beta = ka > 0, 0 < k < 1 \end{aligned} \quad (3)$$

O 의 점들은 그림 2 (a)의 외부 사각형에 해당하며, I 의 점들은 내부 사각형에 해당한다. 외부 사각형에 해당하는 8개의 점 중 중심점의 하단부에 위치한 하나의 점을 제외한 나머지 7개의 점이 흰색이고, 내부 사각형에 해당하는 모든 점들이 검은색인 경우, 이 중심점을 머리의 중심 위치로 해석하게 된다. 이 중심 위치를 찾는 작업은 a 를 임의의 큰 값(화면 가로 길이의 반)으로 초기화하여 a 를 줄여가며 위 (3)의 조건을 만족하는 최초의 픽셀이 발견될 때까지 반복한다. 발견된 픽셀은 머리의 중심일 수도 있지만 약간 왼쪽으로 치우친 점일 수도 있다. 왜냐하면 (3)의 조건을 만족하는 픽셀들이 몇 개 인접해 있을 가능성이 있고, 머리 중심 검색 알고리즘은 이미지의 위쪽부터 왼쪽에서 오른쪽으로 픽셀들을 검사하기 때문이다. 그러므로 (3)을 만족하는 첫 번째 픽셀이 나타나면 해당 위치를 포함한 일련의 수평 방향으로 인접한 검은 픽셀들 중 머리 부분에 포함된 픽셀들을 구하여 그의 중심을 머리의 중심 위치로 정한다.

설명한 과정에 의해 위로 향하는 가장 큰 볼록 형태인 머리의 중심 위치를 찾을 수 있다. 그림 2의 (a)는 이 과정에 의해 발견된 머리 중심 위치의 사례를 보이고 있다. 몸의 중심부에 가장 가까운 하나의 픽셀을 제외한 외부 사각형의 모든 픽셀들은 배경색인 흰색이며, 내부사각형의 점들은 검은색이다. 최종적인 중심 위치 탐지에 사용되는 a 값은 카메라와 사용자 간의 거리에 따라 달라진다. 머리의 중심 위치를 찾을 때 사용한 중심 위치와 수평하며 머리에 포함된 픽셀들의 개수 l 을 계산하여 손 위치의 탐지를 위한 다음 단계에서 사용한다. l 은 중심 픽셀을 포함하는 머리의 수평 거리이다.

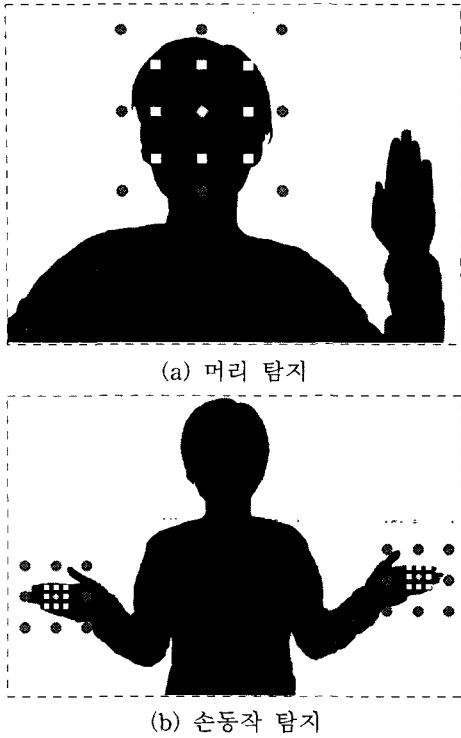


그림 2. 머리와 손동작 탐지 예

3.2 손동작 탐지

본 연구에서 사용하는 손동작은 위로 향하는 것, 오른쪽으로 향하는 것, 왼쪽으로 향하는 것의 세가지이다. 그러므로 손동작 탐지 기법의 목표는 세가지 형태의 손의 중심을 찾아내는 것이다. 이를 위하여 식 (3)과 유사한 형태의 공식 (4)를 사용한다.

$$\begin{aligned}
 &\text{Color of } (x + o_x, y + o_y) \text{ is white, for } (o_x, o_y) \in \\
 &\quad O - S \text{ and} \\
 &\text{Color of } (x + o_x, y + o_y) \text{ is black, for } (o_x, o_y) \in \\
 &\quad I, \\
 &\text{where } S = \{(0, -a), (0, 0)\} \text{ for upward hand} \\
 &\quad S = \{(a, 0), (0, 0)\} \text{ for leftward hand} \\
 &\quad S = \{(-a, 0), (0, 0)\} \text{ for rightward hand} \\
 &\quad (4)
 \end{aligned}$$

식 (4)에서 손의 방향은 아바타를 바라보는 입장에서 손의 방향이다. 즉 그림 2 (b)에서 왼쪽으로 향하는 좌향손은 실제 사용자의 오른손이다. (4)는 식 (3)과 S의 내용만 다른 식이다. 머리의 중심을 둘러

싼 외부 사각형 점들은 몸통에 가까운 부분인 $(0, -a)$ 를 외부 사각형에서 제외한 점들이 흰색이지만 손의 경우는 세가지 방향에 따라 제외되는 점의 위치가 달라진다. 그림 2의 (b)의 왼쪽에 있는 좌향손의 예를 보면 외부 사각형에서 몸통 쪽의 점인 $(a, 0)$ 이 검정이 되어야 한다. 우향손의 경우에는 같은 이유로 $(-a, 0)$ 이 제외된다.

손동작 탐지를 위한 알고리즘은 식 (4)를 이용한 머리 탐지 알고리즘과 같다. 단, 손의 크기는 머리보다 작으므로 a 의 초기값으로 머리 탐지 과정에서 구한 a 를 머리의 수평 지름값인 l 로 스케일링하여 작게 한 값을 사용한다. 그러므로 손동작 탐지 알고리즘은 머리를 탐색하지 않고 넘어가게 된다. a 의 변경은 머리 정한 a 의 하한 값에 도달할 때까지 반복한다. 하한 값을 두는 이유는 머리카락의 끝부분 등과 같은 매우 가는 블록 영역이 손으로 인식되는 것을 방지하기 위한 것이다.

손의 위치가 탐지될 때마다 가까운 위치에 이미 탐지된 손의 위치가 있는지를 검사하고, 이미 탐지된 위치가 있다면 해당 위치는 탐지 결과에서 배제하고 다른 위치의 탐지를 시도한다. 왜냐하면 하나의 손에 대해 식 (4)를 만족하는 픽셀은 몇 개가 인접해있을 수 있기 때문에 이들 중 가장 첫 번째로 발견된 픽셀만 손의 위치로 하는 것이다. 머리 탐색과는 달리 하나의 손의 위치가 탐색된 후 계속해서 탐색하는 이유는 찾아야할 손의 위치가 오른손과 왼손 두 개이기 때문이다.

4. 애니메이션 매핑

카메라로부터 입력된 프레임에 대해 앞에서 기술한 방법에 의하여 손동작이 탐지된다. 손동작 탐지는 양손에 이루어지며 각 손의 방향은 위, 오른쪽, 왼쪽으로 식별된다. 두 손동작의 모양에 따라 다섯 가지 형태의 감정을 표현하는 애니메이션이 사용자의 얼굴 부분에 매핑된다. 현재 RISA는 다섯 가지 형태의 감정 표현을 나타내도록 구현하였지만 두 손동작 입력이 더 많은 종류로 표현 가능하기 때문에 새로운 감정 표현을 나타내도록 쉽게 확장 구현할 수 있다. 표 1은 손동작 입력의 감정 표현에 대한 매핑 관계를 나타낸 것이며 오른손, 왼손은 아바타를 보는 입장이 아니라 사용자 입장에서의 오른손, 왼손을 의미한다.

표 1. 손동작과 감정 표현 간의 관계

손동작		감정 표현
오른손	왼손	
위	위	화남
위	무행동	슬픔
무행동	위	놀람
오른쪽	왼쪽	기쁨
무행동	무행동	무감정

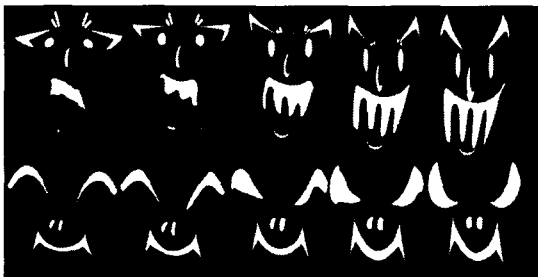


그림 3. 감정 표현 애니메이션 프레임들

감정 표현을 나타내는 애니메이션을 구현하기 위해 이미지 저작도구를 사용하여 각 애니메이션 당 다섯 개의 연속되는 이미지를 작성하였다. 즉 각 애니메이션은 다섯 개의 프레임들로 구성되어있다. 그림 3의 위쪽은 화남을 표현하는 애니메이션 구성 프레임들이고 아래쪽은 기쁨을 표현하는 애니메이션 구성 프레임들이다.

그림 4는 각각 손동작에 따라 무감정, 화남, 기쁨, 놀람을 표현하는 RISA의 예들을 보인 것이며, 표 1의 손동작과의 관계에 맞게 표현됨을 볼 수 있다. 손동작과 해당되는 감정은 의미 없이 연결된 것이 아니라 손동작에 어울리는 감정 표현을 짝이 되도록 하였다. 두손이 모두 무행동이면 무감정이고 두 손을 모두 올리고 있다면 화남 상태, 두 손을 양쪽으로 벌리고 있다면 기쁨을 나타내는 웃음이 어울린다. 그러므로 사용자는 해당 감정에 해당되는 손동작을 좀 더 쉽게 익힐 수 있으며 사용자의 놀이 도구로도 적당하게 사용될 수 있다. 또한 미디어 아트 분야의 전시회에서도 RISA 시스템은 관람자의 손동작에 적합한 표정을 얼굴에 나타냄으로써 관람자의 흥미를 유발시킬 수 있다.

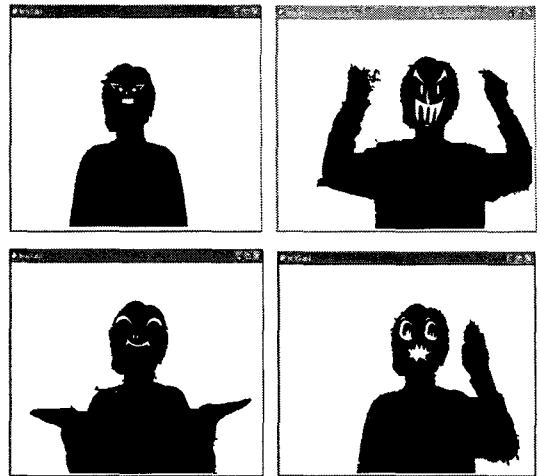


그림 4. RISA 사용 예

5. 실험

구현된 RISA 시스템의 유용성과 탐지 알고리즘의 정확도를 알아보기 위해 몇 가지 다른 조건 하에서 실험을 수행하였다. 실험 대상은 남 1명, 여 1명이며 모두 단발머리 형태이다. 웹캠은 일반적으로 많이 사용되는 저가격의 640X480의 해상도 캠을 사용하였으며 서로 다른 조명 조건을 갖는 가정방과 사무실 방에서 실험을 수행하였다.

식 (3)에서 정의한 a 에 대한 β 의 비율인 k 의 값을 0.65로 하고, 카메라와 사용자 사이의 거리를 0.8~3.0m 범위에서 변화시키고 a 를 114~32 범위에서 변화시켰을 때, 모든 경우에 대해 머리 탐지는 정확하게 수행되었다.

머리의 탐지와는 달리 손동작 탐지는 몇 가지 오류 상황이 발생하였다. 오류의 종류는 크게 두 가지 형태로 볼 수 있는데 첫 번째는 두 개의 위치는 찾았는데 결과가 손이 아닌 다른 신체의 부분(예를 들면 팔)을 포함하는 경우이다. 두 번째는 탐지 알고리즘이 위치 자체를 못 찾은 경우이다. 이 같은 두 가지 오류를 전자는 type1 오류, 후자는 type2 오류라 하겠다. type1 오류는 손이 아닌 신체의 다른 부분이 손을 검색하기 전에 식 (4)를 만족하기 때문에 일어났다. 이는 주로 손과 연결된 부분인 팔 부분에서 일어났다. type2 오류는 손동작을 취할 때 손을 너무 몸통에 밀착했을 때 생기는 것으로 분석되었다. 즉 손이 밀착된 몸통을 하나의 영역으로 인식하기 때문에 오류

가 발생하는 것이다.

표 2는 웹캠을 사용하여 각각의 조건의 방에서 60개의 다른 손동작을 실험한 결과이다. 60개의 다른 손동작은 손목이나 팔꿈치, 손바닥의 각도 등을 바꾸어가면서 서로 다른 동작들을 만들어낸 것이다. 방1은 사무실이고 방2는 가정의 방이다. s 는 손동작 탐지 알고리즘에서 사용될 a 의 초기값을 얻기 위해 머리 중심 부분의 수평 길이인 l 을 스케일링 하는데 사용한 값이다. 즉 a 의 초기값은 s 에 l 을 곱한 값이다.

s 가 작을 때는 작은 블록 윤곽을 갖고 있는 다른 몸의 일부분이 손으로 인식되는 것으로 인해 type1 오류가 더 많이 발생하였다. 반면에 s 가 클수록 손이 머리나 몸통에 인접되어 있는 경우 손을 탐지 못하는 type2 오류가 증가하였다. s 를 여러개로 변화시키며 실험한 결과 s 가 0.3인 경우에 가장 오류가 적었다. 그러므로 표 2에는 s 가 0.3인 경우와 주변값인 0.2, 0.4인 경우의 결과를 나타내었다. 가정방인 방2의 경우는 산란광으로 인한 노이즈가 많아서 가까운 거리에 대해 type1 오류가 많이 발생하였다. 즉 조명의 문제 때문에 가정의 공부방보다는 일반 사무실에서 좀 더 좋은 결과를 얻었다. 하지만 두 조건의 방 모두 거리를 3.0m로 했을 때는 정확도에 큰 차이가 없음을 볼 수 있다. 표 3은 복잡한 배경을 단색을 가진 단순 배경으로 바꾼 경우와 조명을 텡스텐 조명인 고급 조명으로 바꾸어 실험한 결과이다. 복잡 배경과 일반 조명을 사용한 실험에서 좋은 결과를 나타낸 s 가 0.3

인 경우와 거리가 3.0m인 경우만 나타내었다. 두 경우 모두 당연히 일반적인 경우보다 좋은 결과를 보였지만 조명 효과보다는 단순 배경이 오류를 더 줄이는 효과를 보였다.

구현된 RISA 시스템은 미디어 아트 전시장이 카메라와 사용자의 거리가 어느 정도 되는 환경에서는 매우 잘 작동하지만 가정의 방에서 이루어지는 PC 채팅과 같이 카메라와 사용자의 거리가 가까운 경우에는 실행의 정확도가 어느 정도 떨어진다는 문제점을 가지고 있음을 알 수 있었다.

실험에서 나타난 또 다른 문제점은 배경과 같거나 유사한 색의 옷을 입은 사용자의 경우 가상 그림자를 얻는 과정에서 오류를 보일 수 있었다.

6. 결론 및 향후연구

멀티미디어와 웹 환경이 발달함에 따라 인간의 감정 표현을 컴퓨터로 표현하는 이모티콘이나 아바타 같은 기법들이 미디어 아트나 온라인 커뮤니케이션 같은 응용에 많이 사용되고 있다. 많은 아바타들이 소개되었지만 아직 사용자의 움직임에 따라 감정 표현이 애니메이션으로 나타나는 아바타에 대한 연구는 없었다. 본 논문에서는 일상적인 PC 환경에서 사용자의 손동작에 따라 표정이 실시간으로 변하는 실시간 반응형 그림자 아바타인 RISA를 제안하였다. 가상 그림자를 얻기 위해 배경 차별화 기법과 몇

표 2. 손동작 탐지 실험 결과

거리	방	s=0.2			s=0.3			s=0.4		
		type1	type2	정확도	type1	type2	정확도	type1	type2	정확도
0.8m	1	11	10	65.0	5	4	85.0	8	14	63.3
	2	16	7	61.7	13	4	71.7	6	18	60.0
1.5m	1	6	5	81.7	4	0	93.3	7	10	71.7
	2	14	4	70.0	7	5	80.0	4	11	75.0
3.0m	1	4	2	90.0	4	1	91.7	1	5	90.0
	2	6	4	83.3	5	3	86.7	3	4	88.3

표 3. 단순 배경과 고급 조명을 사용한 실험 결과

거리	방	단순 배경(s=0.3)			텡스텐 조명(s=0.3)		
		type1	type2	정확도	type1	type2	정확도
3.0m	1	1	0	98.3	2	2	93.3
	2	3	0	95	4	3	88.3

가지 노이즈 제거 기법들을 사용하였으며 머리와 손 동작의 탐지를 위해 새롭게 고안된 휴리스틱 알고리즘들을 사용하였다. 인간의 여러 감정을 표현하기 위해 감정별로 애니메이션을 구현하였으며 이를 사용자의 손동작에 맞게 실시간으로 사용자 얼굴 부분에 적용시켰다.

정확한 RISA의 실행은 머리와 손동작의 올바른 탐지 결과에 의존한다. 카메라와 사용자의 거리, 조명환경이 다른 방들, 탐지 알고리즘에 사용한 s 값과 같은 파라미터들을 변화해가며 실험한 결과 모든 경우에서 머리의 탐지는 올바르게 수행되었으나 파라미터 값에 따라 손동작 탐지의 정확도는 차이가 있었다. 이들 중 다른 파라미터는 적절한 변경으로 좋은 결과를 얻을 수 있었지만 카메라와 사용자의 거리가 일상 PC 환경과 같이 가까운 경우는 거리를 두는 경우보다 정확도가 떨어지는 결과를 보였다. 하지만 RISA를 간단한 온라인 채팅 환경과 개인적인 PC 놀이 환경에서 사용해본 결과 유용하게 사용할 수 있었다. 또한 RISA는 미디어 아트 전시¹⁾에 사용되어 아트적으로도 유용성이 입증되었다.

향후 연구로는 사용자와 카메라의 거리가 가까울 때 산란광 등에 의해 생기는 부정확성 해결, 사용자의 의도를 표현하는 보다 많은 종류의 손동작 탐지를 위한 탐지 알고리즘의 개선, 시스템의 성능에 영향을 미치는 s 와 같은 파라미터들을 유전자 알고리즘과 같은 기계학습 기법으로 자동 최적화시키는 것 등이다.

참 고 문 헌

- [1] A. Jaimes and J. Liu, "Hotsopt Components for Gesture-Based Interaction," *Proceedings of IFIP Interact*, 2005.
- [2] L. Wang, W. Hu, and T. Tan, "Recent Developments in Human Motion Analysis," *Pattern Recognition*, Vol. 36, No. 3, pp. 585-601, 2003.
- [3] S. Greenberg and C. Fitchett, "Phidgets: Easy Development of Physical Interfaces Through Physical Widgets," *Proceedings of UIST*, pp. 209-218, 2001.
- [4] S.R. Klemmer, J. Li, J. Lin, and J.A. Landay, "Papier-Mâché: Toolkit Support for Tangible Input," *Proceedings of ACM CHI*, 2004.
- [5] M. Ekinici and E. Gedikli, "Real Time Background Model Initialization and Maintenance for Video Surveillance," *Proceedings of Turkish Symposium on AI and Neural Networks*, pp. 164-170, 2003.
- [6] S. Arseneau and J.R. Cooperstock, "Presenter Tracking in a Classroom Environment," *Proceedings of IECON'99*, pp. 145-148, 1999.
- [7] Y.H. Yang and M.D. Levine, "The background primal sketch: an approach for tracking moving objects," *Machine Vision and applications*, Vol. 5, pp. 133-149, 1992.
- [8] M.K. Leung and Y. Yang, "Human Body Motion Segmentation In A Complex Scene," *Pattern Recognition*, Vol. 20, pp. 55-64, 1987.
- [9] A. Jaimes and N. Sebe, "Multimodal Human Computer Interaction-A Survey," *IEEE International Workshop on Human Computer Interaction*, 2005.
- [10] A.F. Bobick and J.W. Davis, "The Recognition of Human Movement Using Temporal Templates," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 23, No. 3, pp. 257-267, 2001.
- [11] Z.B. Simpson, "Installation of Shadow and Flashlight Detection Systems," http://www.mine-control.com/installation_details.html, 2006.
- [12] M. Manohar and H.K. Ramapriyan, "Connected Component Labeling of Binary Image on a Mesh Connected Massively Parallel Processor," *Vision, Computer Graphics and Image Processing*, Vol. 45, pp. 133-149, 1989.

1) 전시 "예술과 과학의 만남", 3인 그룹전(임양미, 정구은, 황세진), 인사동 판훈 갤러리, 2006. 6~7.

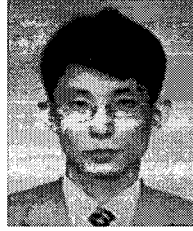


임 양 미

- 1993년 서울산업대학교 매체공학과 (학사)
- 1998년 큐슈예술공과대학교 정보전달 (석사)
- 2004년 8월~현재 중앙대학교 첨단영상대학원 예술공학 (박사과정)

2000년 2월~현재 세영산업(현 (주)엔터티브기획), 고문
 2004년 9월~2006년 8월 성신여자대학교 미디어정보학부 초빙교수

관심분야 : HCI Contents Design, Future Media Art



이 재 원

- 1990년 서울대학교 컴퓨터공학과 (학사)
- 1992년 서울대학교 컴퓨터공학과 (석사)
- 1998년 서울대학교 컴퓨터공학과 (박사)
- 1999년~현재 성신여자대학교 컴퓨터정보학부 교수

관심분야 : 금융공학, 인공지능, 기계학습, 자연언어처리



홍 의 석

- 1992년 서울대학교 계산통계학과 (학사)
- 1994년 서울대학교 계산통계학과 (석사)
- 1999년 서울대학교 전산학과 (박사)
- 1999년~2002년 안양대학교 디지털미디어학부 교수

2002년~현재 성신여자대학교 컴퓨터정보학부 교수
 관심분야 : 소프트웨어공학, 웹기반 컴포넌트 응용