

화상 통신에서의 사생활 보호를 위한 실시간 전경 분리 및 배경 대체

준회원 배 건 태*, 정회원 곽 수 영*, 변 혜 린*

Real-Time Foreground Segmentation and Background Substitution for Protecting Privacy on Visual Communication

Guntae Bae* Associate Member, Sooyeong Kwak*, Hyeran Byun* Regular Members

요 약

본 논문은 모노 카메라로 입력 받은 영상에서 실시간으로 전경과 배경을 분리하여 배경을 자연스럽게 대체 하는 방법을 제안한다. 기존 연구는 대부분 단일 색상의 배경을 이용하여 전경 색에 대한 제약이 있거나, 깊이 정보를 추출을 위한 스테레오 카메라와 같은 장치에 대한 제약이 있거나, 제한적인 전경의 모양 모델을 이용하여 분리할 수 있는 전경의 모양에 대한 제약이 있었다. 이에 본 논문에서는 일반적으로 사용되는 웹캠과 같은 고정된 모노 카메라를 이용하여 실시간으로 전경 분리가 가능한 전경 분리 방법을 제안한다. 또한, 전경 분리의 성능 향상을 위하여 동영상의 시간적인 특징 정보를 이용한 시간적 전경 확률 모델을 제안한다. 또한 분리된 전경과 새로운 배경의 자연스러운 합성을 위한 알파 매트를 이용한 경계선 영역 처리방법과 간단한 후 처리 방법을 제안한다. 제안된 방법은 실제의 화상통신에서 개인의 사적인 정보가 포함된 배경을 자연스럽게 대체시켜 개인의 사생활을 보호할 수 있다.

Key Words : real-time foreground segmentation; natural background substitution; Temporal Foreground Probability Model(TFPM); alpha matte; visual communication.

ABSTRACT

This paper proposes a real-time foreground segmentation and background substitution method for protecting the privacy on visual communication. Previous works on this topic have some problems with the color and shape of foreground and the capture device such as stereo camera. we provide a solution which can segment the foreground in real-time using fixed mono camera. For improving the performance of a foreground extraction, we propose the Temporal Foreground Probability Model (TFPM) by modeling temporal information of a video. Also we provide an boundary processing method for natural and smooth synthesizing that using alpha matte and simple post-processing method.

1. 서 론

정보통신기술의 발달로 인하여 많은 사람들이 기존의 가정용 전화와 개인용 핸드폰을 이용하여 행

하던 의사소통의 많은 부분을 인터넷 폰과 인터넷 메신저와 같은 개인용 메시징 프로그램을 이용하고 있다. 초기에는 텍스트를 중심의 저용량 통신이 대부분이었지만, 웹캠의 등장으로 카메라가 일반화됨

※ 본 연구는 한국과학재단 특정기초연구(R01-2007-000-11683-0)지원으로 수행되었음.

* 연세대학교 컴퓨터과학과(gtbae@yonsei.ac.kr, ksy2177@yonsei.ac.kr, hrbyun@yonsei.ac.kr)

논문번호 : KICS2008-11-511, 접수일자 : 2008년 11월 18일, 최종논문접수일자 : 2009년 4월 6일

에 따라 영상과 음성, 텍스트를 이용한 통신이 증가하고 있다. 화상 통신이 일반화되고 보편화 되면서 이와 비례하게 증가한 것이 개인의 사생활 침해와 보호에 대한 관심이다. 일반적으로 카메라를 이용하여 획득된 영상을 통화나 메신저에 그대로 전송하는 경우 본인의 의사와는 상관없이 보이고 싶지 않는 개인의 사적인 정보까지 노출되는 문제점이 있다. 본 논문에서는 이러한 문제점의 해결을 위하여 통신의 주체가 되는 전경 영역과 개인의 사적인 정보가 포함될 수 있는 배경 영역을 분리하여, 분리된 기존의 배경을 새로운 배경으로 대체하여 자연스럽게 합성함으로써 화상 통신의 장점을 유지하면서 개인의 사생활을 보호할 수 있는 방법을 제안한다.

본 논문은 그림 1에서 보느냐와 같이 고정된 웹 카메라로부터 획득한 영상에서의 전경 영역과 배경 영역의 실시간 분리와 분리된 전경과 새로운 배경의 자연스러운 합성을 목표로 한다. 전경 분리 및 배경대체 기술은 기존에 많은 연구가 존재하지만 본 논문에서는 실시간 처리, 자동 전경 분리, 전경 분리의 정확성 향상, 자연스러운 합성의 4가지 목표에 중점을 두고 있다. 본 논문을 통해 개발되는 기술은 화상 채팅이나 화상 통화, 화상 회의에서 적용되어 화상 통신의 장점을 살리면서 개인의 사생활을 보호할 수 있으며, 그 외에도 화상을 이용한 다양한 통신 분야에 적용가능하다.

II. 관련연구

영상에서 전경 영역을 분리하고 분리된 영상과 다른 영상과 자연스럽게 합성하는 기술은 오래전부터 연구되어온 분야로 다양한 방법들이 존재한다. 각 방법들은 카메라의 수에 따라 단일 카메라 기반과 다중 카메라 기반 방법으로 구분되며, 단일 카메라 기반의 방법론들의 경우 다시 전경 분리에 이용

되는 특징 정보에 따라 색상 정보 기반, 움직임 정보 기반, 모델 기반 방법들로 구분할 수 있다. 다중 카메라 기반의 방법론의 경우 깊이 정보를 이용하는 방법이 주류를 이룬다.

색상 정보 기반의 방법은 초기의 전경 분리 기법으로 영화나 TV에서의 특수효과를 목적으로 개발된 방법¹⁾으로 파란색과 같은 특정 색상의 스크린을 배경으로 이용하여 영상을 촬영하고 촬영된 영상에서 특정 색상에 해당하는 화소들을 제거하여 전경 영역을 획득하고, 제거된 영역을 다른 영상으로 대체하는 방법이다. 이 방법은 간단한 연산만으로 전경 영역을 얻을 수 있어 매우 빠르고 비교적 정확한 전경 영역을 획득할 수 있어 오늘날에도 많이 쓰이는 방법이다. 그러나 배경으로 이용된 색이 전경 영역의 색과 비슷한 경우 정확하게 분리하지 못한다는 점과 단일 색상의 배경이 필요하여 실내 환경에서만 적용가능하다는 문제점이 존재한다. 이중 단일 색상의 배경과 관련된 문제점을 해결하려고 노력한 방법이 배경 차분 기법^{2), 3)}이다. 배경 차분은 입력 영상 중 사람이 등장하지 않는 순수한 배경 영상을 미리 학습하여 배경 모델을 만들어 놓고, 사람이 등장하게 되면 배경 모델과 입력 영상의 차이를 이용하여 사람 영역을 분리 해내는 방법이다. 이 방법의 경우 단일 색상의 배경이 없어도 전경 분리가 가능하며 쉽고 빠르게 전경 영역을 분리할 수 있어 실시간 처리가 가능하다는 장점이 있으나, 카메라가 고정되어 있어야 한다는 것을 전제조건으로 하며, 조명의 변화 민감하며 배경의 색이 조금만 달라져도 학습된 배경 모델과의 차이가 발생하여 전경 영역으로 잘못 분리되는 단점이 있다.

움직임 정보를 기반으로 하는 방법⁴⁾은 배경 차분 기법에서 카메라가 고정되어야 한다는 전제조건을 극복하기 위한 방법으로, 특정 영역이나 화소가 시간적으로 연속된 프레임에서 이동한 방향과 거리에 대한 정보를 이용하여 배경과 전경 영역 간의 차이로 전경 영역을 분리하는 방법이다. 카메라가 움직이는 환경에서도 전경 분리가 가능하다는 장점이 있으나, 정확한 움직임 정보의 추출을 위해서는 많은 계산을 필요로 하여 실시간 처리가 어려우며 약간의 조명 변화에도 민감하다는 단점이 있다.

모델 기반의 방법⁵⁾은 배경 차분과 반대로 전경 영역에 대한 모델을 이용한 방법으로 전경으로 나타날 사람의 몸통, 머리, 팔과 다리에 대한 형태를 학습하여 모델을 만들고, 이를 이용하여 입력 영상에서 각각 모델이 해당하는 부분을 분리해 내는 기술이다. 학습된 모델과 다른 형태의 전경이 나타

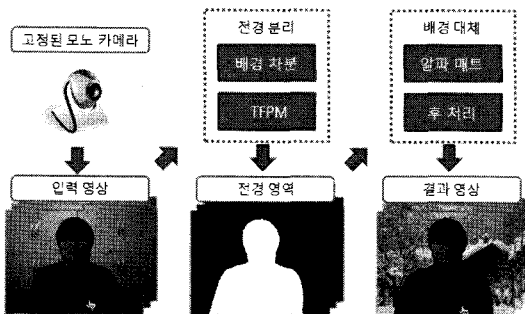


그림 1. 제안하는 시스템의 흐름도

나거나 사람이 손을 드는 등의 다양한 자세를 취했을 때에는 전경 분리가 불가능하다는 단점이 있다.

다수의 카메라를 이용하는 방법^{7, 8)}은 사람의 눈과 같이 카메라를 일정한 간격을 두고 설치하여 각 카메라에서 촬영된 영상간의 시차를 이용하여 영상 내에 존재하는 각 물체들의 깊이 정보를 구하고 이를 이용하여 전경과 배경 영역을 분리하는 방법이다. 배경이 복잡하거나 변하는 등의 다양한 배경 환경에서도 이용가능하다는 장점이 있으나, 카메라의 가격이 고가이며 깊이 정보에 의존적이라는 단점이 있다.

이밖에 자연스러운 합성을 위한 연구로는 전경 영역에 대한 투명도를 나타내는 알파값을 추정하는 매칭 기술^{9, 10)}이 대부분이다. 영상에서 확실한 배경과 전경 영역에 대한 정보를 이용하여 미결정 영역에 대한 알파값을 추정하고 이를 영상 합성에 이용하면 매우 자연스러운 합성이 가능하다. 그러나 한 장의 영상에서 알파값을 추정하는데 수초에서 수분의 시간이 필요하여 실시간 처리가 불가능하다는 단점이 있다.

본 논문에서는 대부분의 화상 통신이 실내 환경에서 이루어지며 카메라가 고정되어 있다는 점을 고려하여 고정된 모노카메라를 이용한 화상 통신에서 실시간으로 전경을 분리하고 배경을 대체하여 자연스럽게 합성하는 방법을 제안하고 한다.

III. 전경 분리

고정된 모노카메라를 통하여 획득된 영상에서 전경과 배경 영역을 분리하기 위한 방법으로 배경 차분(background subtraction)방법을 이용한다. 움직이는 물체가 나타나지 않은 순수한 배경 영상의 획득이 가능한 경우 이를 학습하여 배경 모델(background modeling)의 생성이 가능하다. 배경 모델이 생성되면 이후부터는 간단한 비교 연산만으로 쉽게 전경 영역을 획득 할 수 있어 매우 빠르다는 장점이 있는 반면 배경 모델을 학습하기 위해서는 많은 시간과 메모리 공간이 필요하다는 단점이 있다. 또한 조명이 변하는 경우 학습된 배경과의 차이로 인하여 정확하게 전경 영역 분리 하지 못한다는 단점과 전경 영역의 일부분이 학습된 배경과 색이 유사할 경우 분리해내지 못하는 문제점이 있다. 이에 본 논문에서는 실시간 배경 학습 방법과 색의 유사성으로 인한 문제점 해결을 위하여 시간적 전경 확률 모델을 이용한 성능 향상 방법을 제안한다.

3.1. 배경 차분

우선 본 논문에서는 시간과 메모리를 많이 필요로 하는 배경 학습을 순차적인 방법을 이용하여 실시간 처리가 가능하도록 하였다. 기존의 방법들은 배경 모델의 생성을 위하여 여러 프레임의 순수한 배경 영상을 모아두었다가 한 번에 학습하는 이른바 배치 학습방법을 주로 이용하였다. 이로 인하여 여러 프레임 영상의 저장을 위한 메모리 공간과 많은 데이터를 처리할 시간이 필요하다는 단점이 있었다. 이러한 배경 모델링 방법 중 대표적인 예가 바로 Horprasert 알고리즘¹¹⁾으로 밝기와 색의 왜곡된 정도를 이용하여 그림자의 영향에 덜 민감하다는 특징이 있다. 기본적으로 Horprasert의 배경 모델링은 영상 내의 임의의 화소(pixel) p 를 다음과 같은 4가지 요소 $\langle \mu_p, \sigma_p, a_p, b_p \rangle$ 로 구성한다. 여기서 μ_p 는 화소 p 에서의 컬러 기댓값, σ_p 는 RGB 컬러 값의 표준편차, a_p 밝기 왜곡의 변화(variation of brightness distortion), b_p 는 색도 왜곡의 변화(variation of chromaticity distortion)를 나타낸다. Horprasert 알고리즘은 일정 시간 동안의 고정된 배경 영상을 저장하여 일괄적으로 배경을 학습하기 때문에 학습시간과 메모리의 소비가 많아 실시간 처리가 불가능하다는 문제점이 있다.

본 논문에서는 이러한 문제점을 개선하여 실시간 처리가 가능한 순차 Horprasert 알고리즘¹²⁾을 이용하였다. 순차적 배경모델링 방법은 매 프레임마다 배경을 학습할 수 있도록 수정한 방법으로 학습시간과 메모리 소비를 줄일 수 있다는 장점이 있다. 또한 영상 내에 사람이 검출되는 시점까지 학습하여 그 시점까지의 배경모델을 얻을 수 있어 언제 사람이 등장할지 모르는 실제 환경에서 적용이 가능하다는 장점이 있다.

입력 영상에서의 전경 분리는 영상내의 임의의 화소에 대한 2진 분류 문제로 수식(1)과 같이 임의의 화소 p 에 대해 밝기 왜곡에 대한 변화와 색도 왜곡에 대한 변화를 이용한 객체 유사도를 기준으로 평가한다.

$$l(p) = -\log(f^b(\alpha_p|p)) - \eta \log(f^c(\gamma_p|p)) \quad (1)$$

평가는 수식(2)와 같이 만약 객체 유사도가 임계값 θ_1 보다 큰 경우 해당 화소를 전경 영역으로 그렇지 않은 경우는 배경으로 분류한다.

$$O_{PS}(p) = \begin{cases} 1 & \text{if } l(p) > \theta_1 \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

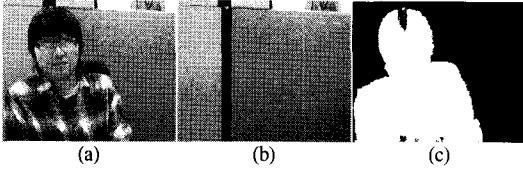


그림 2. 배경 모델 및 차분 결과 (a)입력 영상 (b)학습된 배경 모델 (c)차분 결과

그림 2는 입력 영상에 대하여 순차적으로 학습한 배경 모델의 평균 영상과 이를 이용한 차분 결과를 나타낸 것이다.

3.2. 시간적 전경 확률 모델

배경 차분을 이용한 전경 분리는 하나의 화소에 대하여 색상 정보와 공간적 위치 정보만을 이용하였지만 동영상의 중요한 특징은 바로 시간적 특징 정보이다. 동영상에서 연속적인 두 프레임은 거의 차이가 나지 않는다는 특징이 있다. 이는 이전 프레임에서 전경으로 분리되었던 화소들이 다음 프레임에서도 전경으로 분리될 확률이 높다는 것을 의미한다. 이에 본 논문에서는 이러한 시간적 특징을 이용하여 전경 분리의 성능을 향상시킬 수 있는 시간적 전경 확률 모델(TFPM: Temporal Foreground Probability Model)을 제안한다. TFPM은 이전 프레임까지의 화소별 전경 분리 확률 모델로 수식(3)과 같이 이전 프레임과 다음 프레임의 가중치 합으로 나타낼 수 있다.

$$TFPM^t(p) = (1 - \alpha)TFPM^{t-1}(p) + \alpha R^t(p) \quad (3)$$

where $R^t(p) = [0, 255]$, $TFPM^0(p) = 0$

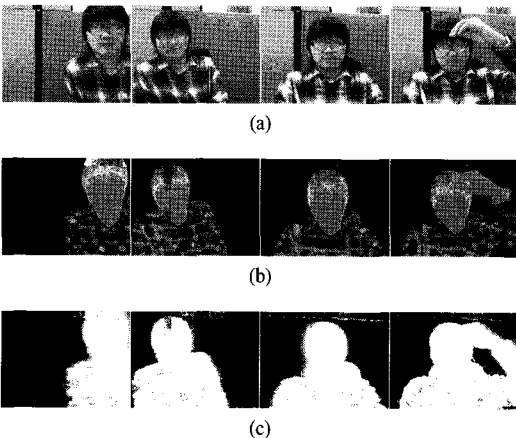


그림 3. 시간적 전경 확률 모델 (a)입력된 5개 시퀀스 (b) 객체 유사도 (b)시간적 전경 확률 모델

수식(3)에서 $TFPM^{t-1}(p)$ 는 화소 p의 0에서부터 t-1 프레임까지의 전경 확률이고 $R^t(p)$ 는 화소 p의 전경 분리의 결과로 전경일 경우 255를 배경일 경우 0을 갖는다. 수식(3)이용한 모델 생성에서 가장 중요한 부분은 전경 영역의 특징에 따른 가중치의 결정이다. 만약 전경의 움직임이 빠른 경우에는 가중치(α)를 크게 하여 바로 이전 프레임에서의 결과를 비중을 높이고, 반대의 경우에는 가중치를 낮게 하여 오랫동안 누적된 시간적 전경 확률 모델의 비중을 높인다. 화상 통신에서 대부분의 경우 사람이 급격하게 움직이는 경우가 적다는 점을 고려하여 가중치를 낮게 설정하였다.

아래 그림 3은 입력 영상에 대하여 배경차분을 위한 객체 유사도와 시간적 전경 확률모델의 생성 결과를 나타낸 것으로 흰색(255)에 가까울수록 해당 화소가 전경일 확률이 높은 것이고 반대로 검정색(0)에 가까울수록 확률이 낮다는 것을 의미한다. 배경 차분에 의한 전경 분리 시 만약 사람의 머리 부분이나 몸에 흠이 발생한 경우 그 영역의 크기가 작더라도 합성 시 어색함을 느끼게 하는 주요원인이 된다. 그러나 시간적 전경 확률 모델을 이용하게 되면 화상대화 시 배경과 색이 비슷하여 발생하는 부분적인 흠과 같은 문제점을 해결 할 수 있다는 장점이 있다.

IV. 배경 대체 및 후처리

본 논문에서는 전경 분리 단계에서 시간적 전경 확률 모델을 이용하여 정확하게 분리된 전경 영역과 새로운 배경을 자연스럽게 합성하여 화상통신에서의 어색함을 최소화하면서 실시간 처리가 가능한 방법을 제안한다.

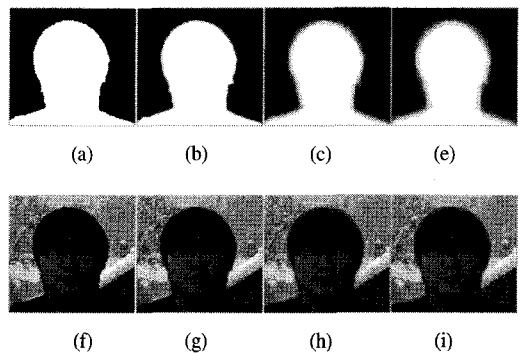


그림 4. 알파매트를 이용한 영상 합성 결과 (a)(b)(c)(d)는 평균 필터의 윈도우 크기별 (0, 3, 8, 10) 알파 매트 (f)(g)(h)(i)는 알파매트를 이용한 합성 결과

합성된 영상에서 어색함을 느끼는 주된 원인은 영상의 경계선 영역에서 단절감으로 인한 부자연스러움이다. 일반적인 영상의 경우 물체의 경계선 부분은 물체와 배경이 겹쳐 보인다. 이러한 이유로 경계선은 물체의 색과 배경의 색이 혼합되어 보이는 특징이 있다. 이러한 특징을 고려하지 않은 단순 영상 합성의 경우 단절감을 느끼게 된다. 이에 본 논문에서는 자연스러운 합성을 위하여 알파 블렌딩(Alpha Blending)^[13] 기법을 이용하며 이는 아래 수식(4)로 나타낼 수 있다.

$$C = aF + (1 - a)B \quad (4)$$

위의 수식에서 C는 합성된 영상, F는 전경 영상, B는 배경 영상을 나타내며, a는 알파 매트릭스 전경 영상의 투명도를 나타내며 0~1사이의 값을 가진다. 본 논문에서는 알파 매트릭스 중 경계선 영역의 알파값을 수평/수직 평균 필터를 이용하여 수정하여 자연스러운 합성이 가능하도록 하였다. 알파값 수정에 이용된 평균 필터의 윈도우의 크기에 따른 알파 매트릭스 합성 결과를 아래 그림 4에 나타내었다. 그림 4 (a)는 필터 윈도우의 크기가 0인 경우는 단순 합성에서의 알파 매트릭스에 해당하며 알파값은 0과 1로 구성된다. 이를 이용하여 합성한 결과는 그림 4 (b)와 같이 경계선 영역에서 단절감으로 인하여 어색함을 느끼게 된다. 이와 반대로 그림 4 (e)와 같이 필터의 크기가 너무 큰 경우 그림 4 (i)와 같이 너무 넓은 경계선 영역이 투명하게 되어 어색함을 줄 수 있다. 본 논문에서는 실험적으로 윈도우의 크기를 3으로 설정하여 이용하였다.

영상 합성에서 부자연스러움의 또 다른 원인은 바로 영상간의 색상 차에 의한 어색함이다. 영상의 색상 보정과 영상 합성과 관련하여 많은 다양한 알고리즘들이 있으나 대부분은 많은 처리 시간을 필요로 하여 실시간 처리 부적합하다. 본 논문에서는 다른 방법들에 비해 간단하고 빨라 실시간 처리가 가능한 대각선 모델^[14]을 이용하여 색상을 보정한다.

대각선 모델을 이용한 색상 보정은 새롭게 대체될 배경의 색상을 분리된 전경 영역과 자연스럽게 어울릴 수 있도록 보정하는 방법으로 수식 (5)으로 나타낼 수 있다.

$$B \times M = F \quad (5)$$

수식 (5)에서 B는 대체될 새로운 배경을 F은 원래의 배경을 나타낸 것이고, M은 분리된 원래의

배경과 새로운 배경을 이용하여 추정된 대각선 모델로 수식 (6)으로 나타낼 수 있다.

$$M = \begin{bmatrix} \alpha & & \\ & \beta & \\ & & \gamma \end{bmatrix} \quad (6)$$

이때 대각선 모델을 구성하고 있는 α, β, γ 각각은 원래 배경과 새로운 배경의 RGB 채널별 평균 명암도(intensity)의 비율을 나타낸다.

V. 실험 및 평가

본 논문에서 제안한 방법의 실험을 위하여 Windows XP에서 Visual C++ 6.0을 이용하여 구현하였으며, 펜티엄-IV의 CPU 3.0 GHz와 2GB RAM의 하드웨어에서 실험하였다. 실험 데이터는 다양한 실내 환경에서 Logitech 사의 QuickCam IM 모노카메라를 이용하여 촬영한 동영상 데이터로 일반적인 영상 통신에 이용되는 320 * 240 해상도이다. 실험 데이터는 화상 통신의 환경을 고려하여 상반신 위주로 촬영하였고, 사람의 움직임 정도와 배경과 사람의 색 유사도를 다양하게 하여 실험하였다. 실험 데이터의 대표 영상을 아래 그림 5에 나타내었다.

5.1. 전경 분리의 정확성 평가

먼저 전경 분리의 정확도를 평가하기 위하여 참 영상(Ground Truth)과 결과영상을 화소마다 비교하였다. 참영상은 그래픽 툴(Adobe Photoshop 6.0)을

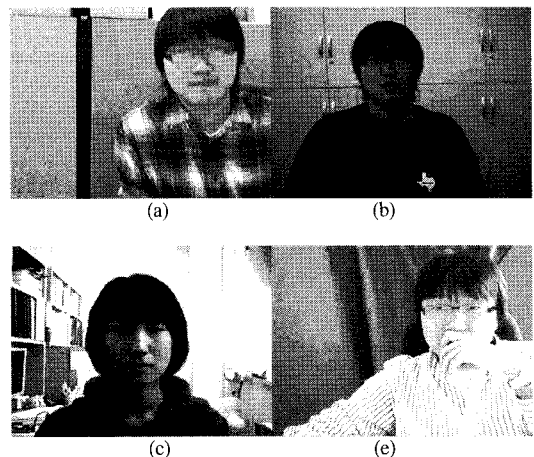


그림 5. 실험 데이터의 대표 영상 (a)Dataset1 (b)Dataset2 (c)Dataset3 (d)Dataset4

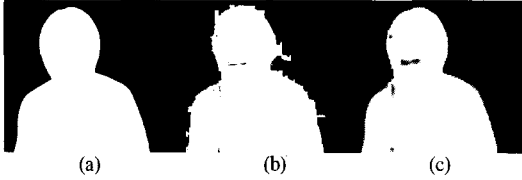


그림 6. 참 영상 및 결과 영상 비교 방법 (a)원영상 (b)참영상 (c)전경 분리 결과 영상 (d)과량색: 상위추출 오류 화소 빨강색: 하위추출 오류 화소

이용하여 사람이 수작업으로 전경을 분리하여 영상을 제작하였다. 모든 프레임에 대하여 참영상을 생성하기에는 제작하는 시간적 어려움이 있어, 영상의 매 10번째 프레임단위로 참영상을 제작하였다.

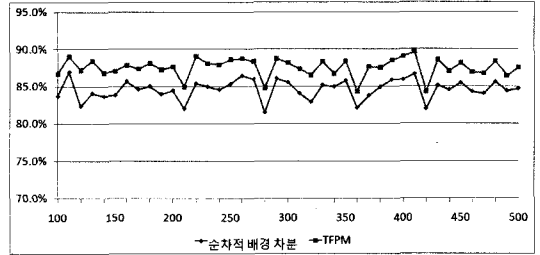
그림 6 (b)는 그림 6 (a)의 입력 영상에 대해 그래픽 툴을 이용하여 제작한 참영상을 보여준 것이고 그림 6 (c)는 전경 분리 결과 영상이다. 영상 분리 결과를 평가하기 위해서 이 두 개의 영상을 화소 단위로 비교하였다. 영상 분리에는 추출되어야 하는데 추출되지 않은 하위추출 오류 화소(under-extraction error pixel)와 추출되지 않아야 하는데 추출된 상위추출 오류 화소(over-extraction error pixel)크게 두 가지로 나누어 볼 수 있다. 두 개의 오류를 측정하여 에러율(error rate) 계산할 수 있다. 그림 6 (d)에 나타난 과량색 영역은 상위추출 오류 화소를 나타내고 빨강색 영역을 하위추출 오류 화소를 나타낸 것이다. 이를 수식 (7)으로 나타내었다.

$$\begin{aligned} E_U &= (G - (G \cap R)) / G \\ E_O &= (R - (G \cap R)) / R \end{aligned} \quad (7)$$

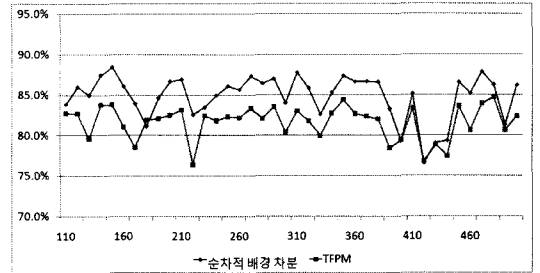
수식 E_U 는 하위추출 오류율을 나타낸 것이고 E_O 는 상위추출 오류율을 나타낸 것이다. 수식(7)에서 G 는 참영상 화소를 나타내고 R 은 영상 분리 결과 화소를 나타낸다. 수식 (7)의 두 개의 오류률의 값은 낮을수록 성능이 좋은 것이고 수식 (8)은 두 개의 오류률을 참고하여 영상 분리의 정확도를 나타낸 것이므로 값이 높을수록 정확한 것이다.

$$A = 1 - (E_U + E_O) \quad (8)$$

위의 수식(8)을 이용하여 실험 데이터 Dataset2와 Dataset4에 대하여 프레임별로 영상 분리의 정확도를 측정하여 그림 7에 나타내었다. 실험데이터 Dataset2 그림 5 (b)의 경우 조명의 영향의 거의 받지 않아 전체적으로 안정적인 분류 성능을 보였으나, 실험데이터 Dataset4 그림 5 (d)의 경우에는 몇몇 프레임



(a)



(b)

그림 7. 영상 분리 정확성 평가 실험데이터 (a)Dataset2 (b)Dataset4

에서 조명의 영향으로 등장인물의 머리카락 색상이 배경과 유사하게 변하여 다소 불안정한 분류 성능을 보였으나 전체적으로 제안하는 전경 분리의 정확성이 향상되었음을 알 수 있다.

전체 실험데이터에 대하여 시간적 전경 확률 모델에 의한 전경 분리의 정확성 향상을 실험하기 위하여 각 데이터마다 순차적 전경 분리 결과와 시간적 전경 확률 모델을 적용한 분리 결과를 측정하여 표 1에 나타내었다.

모든 실험데이터에 대해서 화소단위 영상 분리 결과보다 시간적 전경 확률 모델을 이용한 결과 영상이 더 좋은 결과를 나타냄을 수치적으로 보여주었다. Dataset1과 Dataset2는 전경과 배경의 색이 유사한 부분이 있어 전경 분리의 성능을 저하시키는 요인이 있었고 Dataset4는 조명의 영향으로 전경 분리가 정확하지 않았다. 그러나 세 데이터 모두

표 1. 전경 분리의 정확성 평가(단위:%)

| 분리 방법 | Dataset1 | Dataset2 | Dataset3 | Dataset4 |
|-----------|----------|----------|----------|----------|
| 순차적 전경 분리 | 89.01 | 84.62 | 89.79 | 81.62 |
| TFPM | 91.65 | 87.56 | 90.80 | 84.69 |
| 개선 율 | 2.64 | 2.94 | 1.01 | 3.07 |

표 2. 단계별 평균 처리시간(단위: ms)

| 단계 | Dataset1 | Dataset2 | Dataset3 | Dataset4 |
|----------------|----------|----------|----------|----------|
| 배경 차분 | 17.38 | 17.38 | 17.19 | 16.90 |
| TFPM 갱신 | 1.34 | 1.53 | 1.87 | 1.54 |
| 알파 매트 생성 | 10.83 | 10.99 | 10.88 | 11.22 |
| 영상 합성 및 후처리 | 6.66 | 6.38 | 6.42 | 6.38 |
| 처리 시간 | 36.21 | 36.28 | 36.36 | 36.04 |

제안한 방법을 적용 후 성능이 향상되었다는 것을 알 수 있다.

그러나 Dataset3의 경우 순차적 배경 차분 기법이 비하여 아주 작은 개선율을 보였다. 이 수치는 아주 작다고 판단 할 수 있으나 새로운 배경으로 합성한 최종 결과 영상을 눈으로 확인하였을 때 더욱 자연스럽다는 느낌을 주었다. 그림 8에는 본 논문에서 제안한 방법으로 영상을 분리하여 배경을 대체한 예를 나타낸 것이다.

5.2. 전경 분리 및 배경 대체의 실시간성 평가

실제 화상 통신 분야에 적용하기 위해서는 제안하는 방법이 실시간 처리가 가능해야 한다. 이러한 실시간 처리 가능성을 평가하기 위하여 크게 배경 차분, 시간적 전경 확률 모델 갱신, 알파 매트 생성, 영상 합성 및 후 처리의 4단계로 나누어 평균 처리 속도를 측정하였다. 모든 실험데이터에 대하여 각 단계 별로 평균 처리 속도를 측정하였다. 표 2는 단계별 평균 처리 속도를 표로 나타내어 비교한 것이다.

각 단계별 처리 시간을 측정 해 본 결과 영상을 입력받아 최종적으로 영상을 합성하는 데까지 한 프레임에 처리하는데 평균적으로 36.22 ms 시간이 소요되어 약 27.6 FPS(Frame Per Seconds)의 처리 성능을 나타내었다.

VI. 결 론

본 논문에서는 화상 통신에서 영상의 배경에 포함되어 있는 개인의 사적인 정보의 노출 문제를 통신의 주체가 되는 전경 분리해내어 새로운 배경으로 대체시켜 해결하였다. 이때 전경 분리의 정확성 향상을 위하여 동영상의 시간적 특징을 이용한 시간적 전경 확률 모델을 제안하였다. 이를 이용하여

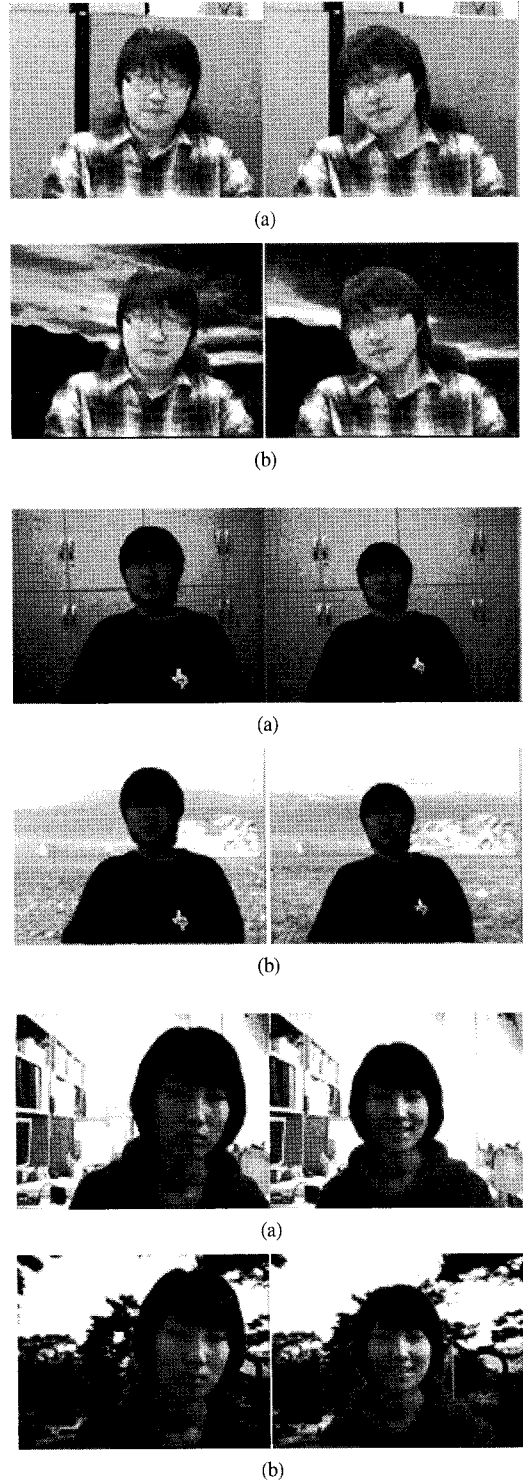


그림 8. 영상 분리 및 배경 대체의 결과 (a)입력 영상 (b)배경 대체 결과
Fig 8. Foreground segmentation and background substitution results (a)input sequence (b)sub

전경을 보다 강인하게 분리할 수 있었으며, 실시간 처리가 가능하였다. 또한 배경 대체로 인한 어색함을 최소화 할 수 있는 합성 및 후처리 방법을 제안하였으며, 제안하는 방법을 이용하여 자연스러운 합성 영상을 얻을 수 있었다. 마지막으로 제안하는 방법으로 입력 영상에서 최종 배경 대체 영상 합성하는데 1초당 27.6 프레임의 실시간에 가까운 처리 성능을 보여 화상을 이용한 다양한 실시간 통신에 적용이 가능함을 나타내었다. 그러나 전경 분리를 위해서 미리 사람이 존재하지 않는 순수한 배경 영상이 존재해야한다는 점과 배경 모델을 위하여 배경 영상을 미리 학습해야하는 다는 점이 극복해야 할 문제점으로 남아있으며, 이를 극복하기 위한 연구와 하드웨어가 제한적인 모바일 환경으로의 확장을 위한 연구가 향후 연구 과제로 남아있다.

참 고 문 헌

[1] A. R. Smith and J. F. Blinn, "Blue Screen Matting", in SIGGRAPH '96: Proceeding of the 23rd annual conference on Computer graphics, pp. 259-268, 1996.

[2] C. Stauffer, W. Grimson, "Adaptive background mixture models for real-time tracking", in Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition, pp. 246-252, 1999

[3] P. Kumar, S. Ranganath, W. Huang, "Queue based fast background modelling and fast hysteresis thresholding for better foreground segmentation", The 2003 Joint Conference of the Fourth ICICS and PCM, Vol. 2, pp. 15-18, 2003.

[4] J. Badenas, J. M. Sanchiz, F. Pla, "Motion-based segmentation and region tracking in image sequences", Pattern Recognition, Vol. 34, No. 3, pp. 661-670, 2001.

[5] H. Luo, Eleftheriadis, A., "Model-based segmentation and tracking of head-and-shoulder video objects for real time multimedia services", IEEE Transactions on Multimedia, Vol. 5, No. 3, pp. 379-389, 2003.

[6] L. Zhao, L.S. Davis, "Closely coupled object detection and segmentation", in Proceedings of the 10th IEEE International Conference on Computer Vision(ICCV'05), Vol. 1, pp. 454-461, 2005.

[7] V. Kolmogorov, A. Criminisi, A. Blake, G. Cross and C. Rother, "Bi-layer segmentation of binocular stereo video", in Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), Vol. 2, No. 2, pp. 1186-1193, 2005.

[8] V. Kolmogorov, A. Criminisi, A. Blake, G. Cross and C. Rother, "Probabilistic fusion of stereo with color and contrast for bi-layer segmentation", IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), Vol. 76, No. 2, pp. 1480-1492, 2006.

[9] Y. Chuang, B. Curless, D. Salesin, and R. Szeliski, "A bayesian approach to digital matting", in Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition(CVPR 2001), Vol. 2, pp. 264-271, 2001.

[10] J. Wang and M. Cohen, "Optimized color sampling for robust matting", in Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2007), pp. 1-8, 2007

[11] T. Horprasert, D. Harwood and L.S. Davis "A Statistical Approach for Real-time Robust Background Subtraction and Shadow Detection", in Proceedings of the 7th IEEE International Conference on Computer Vision, Frame Rate Workshop(ICCV'99), pp. 1-19, 1999.

[12] J.H. Ahn and H. Byun, "Human Silhouette Extraction Method using Region Based Background Subtraction", Mirage 2007: Computer Vision/Computer Graphics Collaboration Techniques and Applications, Lecture Notes in Computer Science, Vol. 4418, pp.412-420, 2007.

[13] Blinn, J.F, "Compositing. 1. Theory", IEEE Computer Graphics and Applications, Vol. 14, pp. 83-87, 1994.

[14] G. Y. Tian, D. Gledhill, D. Taylor, D.

Clarke, "Colour Correction for Panoramic imaging", in Proceedings of the Sixth International Conference on Information Visualization(IV'02), pp. 483-488, 2002.

배 건 태 (Guntae Bae)

준회원



2005년 2월 한남대학교 컴퓨터
멀티미디어학과 졸업
2008년 2월 연세대학교 컴퓨터
과학과 석사
2008년 3월~현재 연세대학교 컴
퓨터과학과 박사과정

<관심분야> 영상처리, 컴퓨터 비전 및 패턴 인식

곽 수 영 (Sooyeong Kwak)

정회원



2003년 2월 계명대학교 컴퓨터
공학과 졸업
2005년 2월 연세대학교 컴퓨터
과학과 석사
2005년 3월~현재 연세대학교 컴
퓨터과학과 박사과정

<관심분야> 영상처리, 컴퓨터 비전 및 패턴 인식

변 혜 란 (Hyeran Byun)

정회원



1980년 2월 연세대학교 수학과
졸업
1987년 2월 연세대학교 수학과
석사
1987년 6월 Univ. of Illinois at
Chicago Computer Science,
M.S.

1993년 12월 Purdue Univ. Computer Science, Ph. D.

2004년 3월~현재 연세대학교 컴퓨터과학과 정교수

<관심분야> 패턴 인식, 컴퓨터 비전 시스템