

얼굴 검출을 이용한 샷 유형 감지 시스템

백영태*, 박승보**

Shot Type Detecting System using Face Detection

Yeong Tae Baek*, Seung-Bo Park**

요약

본 논문은 얼굴 검출을 이용한 샷의 유형을 판단하는 방법론을 제시한다. 클로즈업 샷이나 미디엄 샷, 롱 샷과 같은 샷의 유형은 영화의 서사 구조를 파악하는 주요한 단서이다. 클로즈업을 통해 감독은 등장인물의 감정 상태를 묘사하고 롱 샷을 통해 인물이 처한 상황이나 배경을 묘사하게 된다. 인물의 심리나 감정의 변화, 인물이 처한 상황을 묘사하는 샷의 여러 유형은 인물과 카메라의 거리에 의해 결정된다. 따라서 화면에 등장하는 인물의 얼굴 크기를 알아내어 샷의 유형을 판단할 수 있다. 이를 위해 본 논문에서는 얼굴 검출을 통해 샷의 유형을 감지하는 방법론을 제시하고 시스템으로 구현하여 성능을 평가한다. 평가실험에서 클로즈업 샷과 미디엄 샷의 감지 성능은 95%와 90%로 비교적 높게 나타났지만 얼굴의 윤곽이 불분명한 롱 샷의 경우 53.3%로 측정되었다.

▶ Keywords : 샷 종류, 얼굴 검출, 사용자 인터페이스, 하르, 샷 경계 추출

Abstract

In this paper, we propose the method that decides the shot types using face detection technique. The shot types, such as close-up shot, medium shot, and long shot, can be applied as useful information for understanding narrative structure of movies. The narrative structure of movie is builded by characters. Also their mental and emotional changes become inextricably bound up with them of narrative. The shot types are decided by distance between character and camera. If put together above them, shot types can be found by using detection technique of face size of characters and understand narrative of movie. To do this, we propose the methodology to detect shot type by face detecting and implement the system to do it. Additionally, we evaluate the

• 제1저자 : 백영태 • 교신저자 : 박승보

• 투고일 : 2012. 09. 03, 심사일 : 2012. 09. 20, 게재확정일 : 2012. 09. 24.

* 김포대학교 멀티미디어과(Dept. of Multimedia, Kimpo College)

** 경희대학교 경영대학(School of Management, Kyung Hee University)

※ 이 논문은 2012학년도 김포대학교의 연구비 지원에 의하여 연구되었음

performance of the system. The implementation system has been evaluated as 95% for close-up shot detection and 90% for medium shot detection, while 53.3% is just detected for long shots.

▶ Keywords : Shot type, Face detection, User interface, Haar, Shot boundary detection

I. 서 론

영화에서 샷(shot)의 유형은 장면의 상황을 묘사하거나 등장인물의 감정 상태를 묘사하기 위한 수단으로 사용된다. 클로즈 업(Close-up)의 경우 등장인물의 얼굴을 확대하여 감정 상태나 변화를 관객에게 전달하기 위해 사용된다. 롱 샷(Long shot)은 등장인물이 위치한 주변 환경을 같이 보여주어 공간적·시대적 배경에 대한 설명이나 등장인물이 처한 상황을 설명하기 위해 사용된다[1,2]. 따라서 영화에서 샷의 유형을 알아내면 감독이 샷이나 장면을 통해 관객에게 전달하고자 하는 의도를 파악하는데 도움이 될 수 있고, 이를 활용하여 스토리 관점의 영화 검색 시스템을 구축할 수도 있다. 클로즈 업 샷을 찾아내면 관객에게 인물의 심리나 감정 상태를 묘사하려고 한다는 것을 알 수 있으며, 롱 샷은 관객에게 등장인물이 처한 상황이나 스토리의 배경을 설명하려 한다는 것을 파악할 수 있다. 클로즈 업 샷이나 롱 샷은 등장인물과 카메라 사이의 거리에 의해 결정된다. 다시 말해 카메라가 인물을 얼마나 크게 확대하여 촬영하느냐에 의해 결정되는 것이다. 확대에 대한 정도는 화면의 크기에 대한 등장인물의 얼굴의 크기의 비율로 결정될 수 있다. 얼굴의 크기를 검출하여 현재 샷이 클로즈 업 샷인지 롱 샷인지를 판단할 수 있다. 따라서 본 논문은 얼굴 검출 기술을 이용하여 현재의 샷의 유형이 무엇인지를 판단하는 방법을 제안하고 이를 시스템으로 구현하는 것을 목표로 한다.



그림 1. 영화 노팅힐에서 샷의 유형
Fig. 1. Shot types in a Movie "Notting Hill"

샷의 유형은 그림 1에서 보여지는 것처럼 크게 클로즈 업 샷, 미디엄 샷, 롱 샷의 3가지 종류들로 구별할 수 있다[3].

클로즈 업 샷은 얼굴의 반응을 강조하여 크게 보여주는 극적인 샷으로 그림 1의 a처럼 일반적으로 얼굴 전체와 턱밑의 어깨 선 부분까지 포착하는데 이때 얼굴이 전체 화면의 대부분을 차지하게 된다[3]. 미디엄 샷은 인물의 허리나 그 약간 밑 부분을 포착하여 인물의 팔꿈치와 거의 같은 높이로 할 수 있다. 인물이 클로즈 업 샷에 비해 더 작게 표현되므로 머리 위의 빈 공간인 헤드룸이 클로즈 업 샷에 비해 일정영역을 차지하도록 촬영된다. 미디엄 샷을 통해 인물의 시선 방향이나 움직임 또는 대화 상대의 표현을 하는데 사용된다. 미디엄 샷은 경우에 따라 2인의 대화 장면에서 사용되기도 한다. 롱 샷은 인물의 전신을 포함하면서 발 밑의 영역까지 촬영한다. 또한 미디엄 샷에 비해서 훨씬 넓은 헤드룸을 갖게 된다. 롱 샷은 화면 안에서 인물의 움직임이 전부 표현될 수 있기 때문에 화면이 인물을 따라갈 필요가 없이 고정되어 있을 수 있으며, 인물 주변의 배경이나 다른 인물을 함께 묘사할 수 있다. 이렇게 샷의 유형은 카메라와 인물 사이의 촬영 거리에 의해 결정된다. 즉 얼굴의 크기를 검출하여 얼굴이 화면에서 차지하는 비율을 계산하여 샷의 유형이 무엇인지를 판단할 수 있다.

이를 위해 본 논문에서는 다음과 같이 얼굴 검출 기술과 샷 경계 추출 기술을 활용한다.

- Haar와 피부색을 이용한 얼굴 검출 방법: 현재 프레임에서 얼굴이 있는 후보 영역을 추출하기 위해 Haar를 이용하였으며, 후보 영역이 얼굴인지 아닌지를 판단하기 위해 눈과 입이 포함 되었는지를 판단하는 Haar를 추가적으로 진행하였다. 또한 피부색의 포함 정도를 판단하여 얼굴의 후보 영역을 추가적으로 판단하도록 하였다[4,5,6].
- 얼굴의 크기와 샷의 유형간의 결정 규칙: 샷에서 나타난 얼굴들의 크기와 화면 크기 사이의 비율을 이용해서 샷의 유형을 판단할 수 있는 규칙을 제안한다.
- 히스토그램을 이용한 샷 경계 추출 방법: 샷의 유형을 판단하기 위해 얼굴을 검출하기 이전에 샷의 경계를 파악하는 작업이 선행되어야 한다. 샷의 경계를 판단하기 위해 히스토그램의 변화량을 추출하였으며, 변화량이 특정값 이상인 경우에 샷의 경계로 인식되도록

하였다[7].

추출된 얼굴 크기와 화면에 대한 비율을 계산하기 위해 화면의 크기와 얼굴의 크기를 알아내어 계산하였다. 얼굴의 크기와 화면 크기 간의 비율에 기반한 샷의 유형을 판단 규칙들을 제안하였으며 규칙들을 이용하여 샷의 유형을 판단하였다.

본 논문에서는 이와 관련된 전반적인 기술들과 시스템 구현 결과를 설명한다. 이를 위해 2장에서는 배경 지식과 관련된 연구들에 대해 설명한다. 3장에서는 본 논문에서 사용된 기술들과 기술들을 통합한 샷 유형 감지 시스템에 대해 설명하며, 4장에서는 구현된 감지 시스템의 설명과 성능 평가에 대해 논의한다. 그리고 마지막으로 연구의 결과에 대해 5장에서 정리한다.

II. 배경 지식 및 관련 연구

1. 샷의 유형

렌즈나 카메라의 이동이 없는 상태에서 매우 단순한 피사체의 움직임만을 촬영하는 단순 샷은 한 사람 이상의 인물이나 피사체를 포착한다. 이 단순 샷의 유형은 크게 분류하여 클로즈 업 샷, 미디엄 샷, 롱 샷의 3가지로 구별하거나 좀 더 세분화하여 표 1과 같이 11개의 샷의 유형으로 분류할 수도 있다[3].

표 1. 샷의 유형
Table 1. Shot Types

3가지 유형	세분화한 분류
클로즈 업 샷	익스트림 클로즈 업 샷
	빅 클로즈 업 샷
	클로즈 업 샷
	미디엄 클로즈 업 샷
	오버 쇼울더 샷
미디엄 샷	미디엄 샷
	미디엄 롱 샷
	미디엄 2 샷
롱 샷	롱 샷
	베리 롱 샷
	익스트림 롱 샷

감독이 전달하고자 하는 메시지나 등장인물의 감정 변화에 따라 세분화된 다양한 형태의 샷의 유형이 사용될 수 있다. 하지만 본 논문에서는 유형간의 경계가 확실한 3가지 샷의 유형을 사용하였다.

클로즈 업 샷은 얼굴 전체와 턱 밑의 어깨 선 부분까지 포

착하는 것으로 머리 공간인 헤드룸은 매우 작거나 거의 없게 표현된다. 이렇게 표현된 클로즈 업 샷에서 얼굴은 화면의 50% 이상의 비율을 차지하게 된다. 영화의 영상은 TV의 화면과 달리 1.85:1이거나 2.35:1의 비율을 나타낸다. 가로와 세로에 비해 월등히 넓으면서 가변적이기 때문에 얼굴과 화면의 비율은 가로의 비율보다는 세로에서 차지하는 비율만을 고려하여야 한다. 다시 말해 클로즈 업 샷에서 얼굴의 세로 길이는 화면의 세로 길이에 대해 50% 이상의 비율을 차지하게 된다.

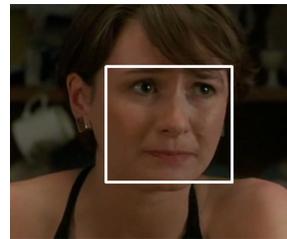


그림 2. 얼굴 영역
Fig. 2. Face Area

얼굴의 영역은 다음 장에서 설명할 얼굴 검출에 의해 추출되는 영역으로 그림 2와 같이 눈썹 바로 위부터 입술 바로 아래까지를 포함한다.

미디엄 샷은 허리 약간 아래까지 잡는다고 하여 웨이스트 샷(Waist Shot)이라고 불리기도 한다[3]. 그림 1의 b처럼 헤드룸이 일정량 존재하며 허리 약간 아래까지 인물을 포착한다. 얼굴이 화면에 대해 차지하는 비율은 20% 이상이 된다. 미디엄 샷은 클로즈 업과 달리 한 화면에 2명 이상이 포착되기도 하므로 2개 이상의 얼굴에 발견된 경우 그 값들의 평균 값을 얼굴의 비율로 활용할 필요가 있다.

롱 샷은 그림 1의 c처럼 인물의 전신은 물론 넓은 헤드룸과 발 밑의 공간을 함께 묘사할 수 있다. 또한 인물의 좌우 공간에 다양한 배경을 함께 표현할 수 있다. 인물과 배경에 대한 묘사가 가능하므로 한 장면의 앞 부분에 주로 설정되곤 한다[2]. 전신과 헤드룸과 발 밑 공간이 함께 포착되므로 얼굴은 화면에 비해 8% 안팎의 비율을 유지하게 된다.

2. 얼굴 검출

얼굴 검출에 많이 활용되는 기술은 Haar-like 특징을 이용한 방법과 피부색을 이용한 방법이다. 또는 두 개의 기술을 혼합하여 빠르게 화면에서 얼굴을 검출하기도 한다.

동영상에서 얼굴 검출을 하는 것은 매우 빠른 연산 속도가 요구된다. Haar-like 특징을 이용한 방법은 적분 이미지를

이용하여 빠르게 연산이 가능한 방법론으로 복잡한 연산을 줄여서 실시간 객체 검출에 효과적이다[8]. 얼굴 영역 내에서 간단하고 객체 간의 구별 능력이 있는 특징들을 사용하여 해결한다. Haar-like 특징들은 그림 3처럼 Edge, Line, Center-surround 등으로 정의된다[5].

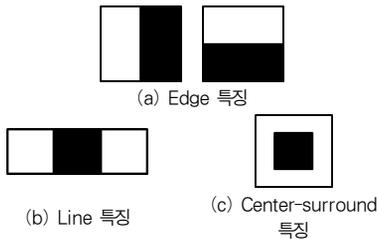


그림 3. Haar-like 특징
Fig. 3. Haar-like Features

Haar-like 특징만을 이용할 경우 영화와 같이 다양한 개체들이 촬영되고 조명의 변화가 심한 경우 얼굴이 아닌 개체들이 오인되어 검출되는 경우가 많다. 또한 해상도가 높은 경우 실시간으로 얼굴을 검출 하지 못하는 경우가 있다. 이를 개선하기 위해 피부색 영역만을 사전에 검출하여 탐색해야 할 얼굴 영역을 줄여주게 되면 좀 더 빠르고 정확한 얼굴 검출 결과를 얻을 수 있다[5,6]. 피부색을 이용한 얼굴 후보 영역 검출은 색공간에서 이루어진다. 영화의 경우 얼굴 위에 그림자나 조명의 변화가 빈번하게 나타날 수 있으므로 그림자에 강인한 HSV 색공간으로 변환하여 피부색 영역 검출을 진행한다. 피부색은 식 1을 만족하는 픽셀들로 결정할 수 있다.

$$(S(p_i) \geq 0.2) \text{ and } (0 \leq H(p_i) \leq 30) \quad (1)$$

3. 샷 경계 검출

샷은 카메라의 촬영이 끊이지 않고 이어지는 연속된 이미지(프레임)들의 집합이다. 샷에서 다음 샷으로 넘어갈 때 이미지의 연속성은 보장되지 않으므로 이미지 간의 차이는 매우 크게 된다. 이러한 속성을 이용하여 샷 간의 경계를 추출하는 대표적인 방법이 히스토그램을 이용한 샷 경계 검출 방법이다. 히스토그램은 이미지의 명암값 프로파일을 보여주는 것으로 그림 4의 중간 부분처럼 픽셀들의 밝고 어두움을 256개의 계조로 분포시킨 그래프이다. 이미지 전체의 각 픽셀들을 읽어서 픽셀이 해당하는 256개의 계조의 명암값들 중에 1개를 증가시키는 작업을 반복하게 되면 이미지에 대한 히스토그램이 완성되게 된다. 샷 내의 프레임들은 연속적이기 때문에 이

미지 간의 변화가 미미하다. 하지만 샷과 샷의 경계에서는 완전히 다른 이미지로 변화한다. 따라서 현재 프레임에서 이전 프레임 간의 차영상에 대한 히스토그램을 추출할 경우 밝은 영역에 높은 분포를 갖게 된다.



그림 4. 이미지의 명암 히스토그램 그래프
Fig. 4. Brightness Histogram Graph of Image

그림 4처럼 연속된 프레임인 샷 내의 대부분의 프레임들의 차영상에서는 밝은 영역이 적게 나타나지만 샷의 경계에서는 이미지 전반적으로 밝게 나타난다. 따라서 샷의 경계를 추출하기 위해서는 특정 레벨 이상의 값을 취하게 되면 샷의 경계에서 유달리 큰 값을 얻게 된다.

차영상에 대해 시간적으로 변화하는 히스토그램의 변화량을 그래프로 표현하면 값이 특이하게 커지는 부분이 샷의 경계가 된다. 즉 그림 5처럼 차영상의 히스토그램에서 특정 레벨 이상의 밝기를 갖는 값만을 합하여 연속된 그래프로 표현할 경우 그림 5의 b에서 보여지는 것처럼 샷의 경계 부분에서 매우 높은 값을 보여주게 된다. 히스토그램 변화량 그래프에서 특정값 이상을 나타내는 프레임이 샷의 경계 프레임이 된다[7]. 이 방법은 80% 이상의 비교적 높은 검출 성공률을 보여주는 방법이지만 이 방법의 경우 카메라 가까이에서 인물이 이동하거나, 카메라가 이동하면서 촬영할 경우 배경이 움직이기 때문에 차영상이 매우 밝게 나타나게 되어 샷의 경계로 오인되는 문제점이 있다. 이의 개선을 위해 차영상과 함께 특징점의 추적을 고려하면 샷의 경계 추출 성공률을 높여줄 수 있다. 하지만 특징점의 추적은 많은 연산 시간을 필요로 하여 샷의 경계 추출과 함께 얼굴 추적을 진행하여야 하는 본 논문의 경우 적용하는데 문제가 있다. 따라서 차영상의 히스토그램

변화량만으로 샷의 경계를 검출하도록 한다.



그림 5. 차영상과 히스토그램 변화량 그래프
Fig. 5. Differential Image and Histogram Variation Graph

III. 얼굴 검출 기술을 이용한 샷 유형 감지 시스템

얼굴 검출 기술을 이용한 샷 유형 검출을 위해서는 그림 6의 시스템 구조도에 보여지는 것처럼 프레임에서 얼굴을 정확하게 검출하는 모듈과 프레임들이 속해 있는 샷의 경계를 추출하는 모듈이 필요하다.

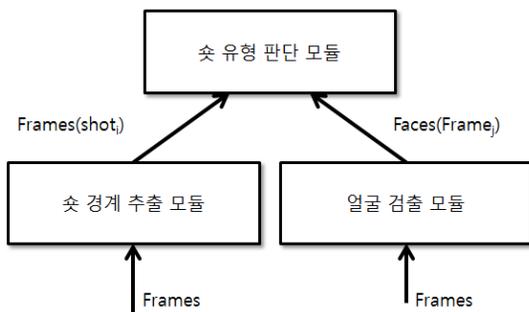


그림 6. 시스템 구조도
Fig. 6. System Architecture

프레임을 받아서 얼굴의 영역을 추출하고, 연속된 2개의 프레임을 받아서 샷의 경계를 추출하는 작업이 선행적으로 이루어져야 하며, 이 두 개의 모듈들을 통해 얻어지는 얼굴 영역과 샷의 경계를 이용하여 얼굴영역이 샷 내의 프레임들에서 차지하는 비율을 계산하여 샷의 유형을 판단한다. 샷의 유형을 판단하기 위한 각 모듈들에 대해 구체적으로 살펴보자.

1. 얼굴 검출 모듈

얼굴 영역을 추출하기 위해 본 연구에서는 피부색으로 얼굴 후보 영역을 추출한 후 Haar-like 특징을 이용하여 얼굴 영역을 추출하였다.

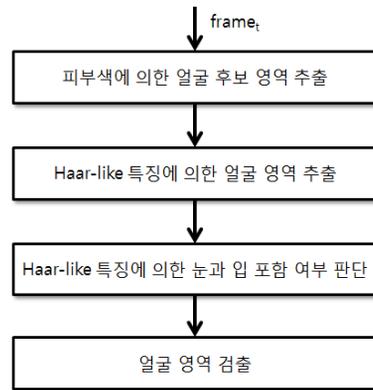


그림 7. 얼굴 검출 흐름도
Fig. 7. Face Detection Flow Chart

영화에서는 다양한 색과 모양의 개체들이 촬영되기 때문에 Haar-like 특징을 이용한 방법만을 사용하여 프레임에서 얼굴 영역을 추출할 경우 얼굴이 아닌 개체들이 추출되기도 한다. 그래서 추출된 얼굴 영역들 중에서 추가적으로 내부에 눈과 입을 포함하는 얼굴 영역들만을 얼굴로 최종 검출하였다. 눈과 입 역시 그에 맞는 Haar-like 특징을 이용하여 검출 하였다. 이러한 과정을 정리하면 그림 7과 같은 흐름도로 정리된다.

먼저 입력된 프레임에서 식 1에 의해 피부색에 해당하는 부분들만 먼저 추출하여 얼굴이 아닌 영역은 검게 칠한다. 이어서 미리 학습된 Haar-like 특징을 이용하여 얼굴 영역을 추출한다. 이때 OpenCV[9]와 Swiki의 얼굴 학습 데이터들 [10] 활용한다. 얼굴 영역이 추출되고 나면 잘못 검출된 얼굴들을 제거하기 위해 눈과 입의 포함여부를 판단한다. 잘못 검출된 얼굴에서 눈과 입을 추가적으로 검출함으로써 잘못된 값을 줄여주게 된다.

추가적으로 동영상에서 잘못 검출된 얼굴 영역을 제거하기 위해 샷 내에서 1 프레임에서만 검출된 얼굴 영역은 배제시켰다.

2. 솟 경계 추출 모듈

솟의 경계를 추출하기 위해 그림 8의 흐름도와 같이 먼저 연이은 프레임들(frame_t, frame_{t-1})의 차영상을 계산하여 추출하였다. 그리고 차영상에 나타나는 변화의 크기를 측정하였다.

솟의 내부에서 변화를 발생시키는 것은 피사체나 배경의 움직임이다. 피사체의 움직임으로 인해 피사체의 내부 영역에서 발생하는 차영상은 밝기 변화가 매우 낮게 된다. 하지만 솟의 경계와 같이 이미지 자체가 바뀌는 경우에는 차영상 전체적으로 매우 밝기 변화가 크게 나타난다. 따라서 솟 경계에서 나타나는 밝기 변화를 효과적으로 적용하기 위해 히스토그램에서 특정 레벨 이상의 값만을 변화량으로 반영하였다. 본문에서는 그 값을 100으로 하였다.

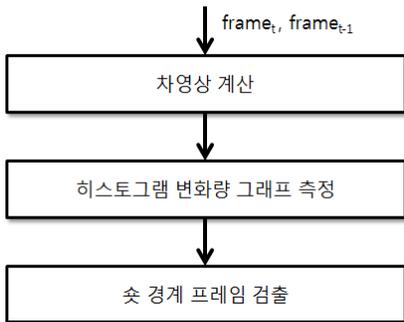


그림 8. 얼굴 검출 흐름도
Fig. 8. Face Detection Flow Chart

이렇게 얻어진 히스토그램 변화량 그래프에서 특정 레벨 이상의 변화량을 나타내는 프레임을 새로운 솟의 시작 프레임으로 설정하였다[7].

3. 솟 유형 판단 모듈

솟의 유형은 검출된 얼굴이 화면에서 차지하는 비율을 통해 계산되어진다. 얼굴 영역은 그림 2처럼 눈썹 바로 위에서 입술 아래까지의 영역이다. 영화에서 화면의 가로 크기는 유동적이므로 세로 길이만을 얼굴 영역과 비교하기 위한 지표로 사용하였다. 얼굴 영역 역시 세로 길이만을 사용하였다. 솟의 유형은 클로즈 업 솟, 미디엄 솟, 롱 솟의 3가지 분류를 사용하였다. 한 솟 내에서 얼굴 영역이 여러 개가 추출된 경우 잘못 검출된 얼굴 영역의 영향을 줄여주기 위해 식 2와 같이 각 얼굴 영역이 나타난 프레임수를 함께 고려하여 평균값을 계산하였다.

$$FSize(shot_k) = \frac{\sum_{i=1}^n FrmCnt(face_i) \times FSize(face_i)}{\sum_{i=1}^n FrmCnt(face_i)} \quad (2)$$

$FSize(s)$: s 에 해당하는 얼굴의 세로길이(Face Size)

$FrmCnt(face)$: $face$ 가 나타나는 프레임들의 개수 (Frame Count)

n : 솟 내에서 나타나는 얼굴의 개수

대화의 장면에 속한 솟이나 롱 솟에서 2개 이상의 얼굴 영역이 추출되거나 잘못된 얼굴 영역이 추출된 경우에 평균값을 계산하게 된다. 식 2는 솟 내에서 많이 검출된 얼굴의 영향력을 강화하기 위한 것으로 많이 검출된 얼굴을 솟의 얼굴 영역으로 사용하겠다는 것을 의미한다. 이렇게 얻어진 솟의 얼굴 영역을 이용해서 솟의 유형을 판단하게 된다.

표 2. 솟의 유형 판단 규칙
Table 2. Rules to Decide Shot Types

솟의 유형	규칙
클로즈 업 솟	$FSize(Shot) \geq ScreenSize \times 0.5$
미디엄 솟	$ScreenSize \times 0.3 > FSize(Shot)$ and $FSize(Shot) \geq ScreenSize \times 0.2$
롱 솟	$ScreenSize \times 0.09 \geq FSize(Shot)$ and $FSize(Shot) \geq ScreenSize \times 0.07$

얼굴의 크기가 화면에서 차지하는 비율에 의해 솟의 유형이 표 2의 규칙에 의해 결정된다.

클로즈 업 솟은 얼굴의 세로크기가 화면의 세로 길이에 비해 50% 이상인 경우로 하였다. 얼굴 영역이 눈썹부터 입술까지 이므로 그 이외의 얼굴 영역까지 고려하면 실제 얼굴이 차지하는 비율은 80% 이상이 된다. 즉 화면 대부분을 얼굴이 차지하는 클로즈 업 솟의 특징을 대변하는 것이다.

미디엄 솟은 얼굴이 화면에서 20%~30% 사이의 크기를 차지하는 경우이다. 이것은 실제 얼굴영역이 화면에서 30%~45% 정도의 영역을 차지하는 것을 의미한다. 화면에서 머리 위의 헤드룸 부분과 몸통 부분에 대한 고려를 하여 얼굴이 차지하는 영역을 결정한 것이다.

롱 솟은 얼굴 영역이 화면 크기에 비해 7%~9% 정도의 영역을 차지하는 경우로 실제 얼굴영역이 10%~13% 정도의 영

역을 차지하게 된다. 헤드룸이 화면의 10%, 몸 전체의 길이가 화면의 70%, 발 밑의 공간이 화면의 10% 정도를 차지한다고 했을 때 얼굴은 화면의 10% 남짓 정도를 차지하게 된다.

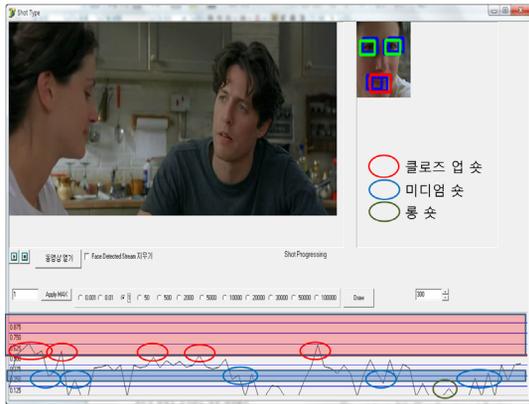


그림 9. 구현 결과 화면
Fig. 9. Screen Shot of Implementation Result

IV. 시스템 구현 및 평가

제안된 방법론에 의해 샷의 유형을 결정하는 프로그램을 개발하였다. 얼굴 검출과 샷의 경계를 추출한 후 표 2의 샷의 유형 판단 규칙을 사용하여 그림 9와 같이 샷 중에서 클로즈 업 샷, 미디엄 샷, 롱 샷에 해당하는 샷을 추출하였다.

시스템은 델파이 7을 이용하여 운영체제 윈도우 7 하에서 개발하였다. 얼굴 검출은 OpenCV 2.0의 Haar 라이브러리를 이용하여 구현하였다. 얼굴 전면에 대해서만 추출할 수 있도록 하였으며 OpenCV Swiki의 Frontal Face용 Haar cascade 파일을 사용하였다[8]. 얼굴 영역을 확장하기 위해 눈과 입이 있는지를 판단하기 위해 OpenCV Swiki의 Right Eyes, Left Eyes, Mouth 용 Haar cascade 파일을 추가적으로 활용하였다.

이러한 과정을 거쳐 구현된 프로그램의 결과 화면을 그림 9에 보여주었다. 프로그램은 영화를 프레임 단위로 읽는 부분과 샷의 경계를 추출하는 부분, 프레임에서 얼굴 영역을 추출하는 부분, 그리고 맨 아래에 얼굴이 화면에서 차지하는 비율을 계산하여 샷의 순서대로 그래프로 표현하여 샷의 유형을 판단하는 부분으로 구성하였다.

샷 유형 감지의 성능을 판단하기 위해 영화 노팅힐의 각 샷들에 대해 평가를 진행했다. 노팅힐에서 인물이 등장한 샷들을 100개를 추출하여 각 샷의 유형을 얼마나 잘 검출하는

지를 판단하였다. 클로즈 업 샷은 40개, 미디엄 샷은 30개, 롱 샷은 30개를 추출하여 성능을 측정하였다.

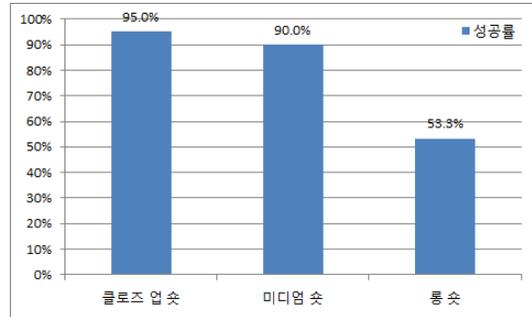


그림 10. 샷 유형 감지 성능
Fig. 10. Shot Type Detection Performance

클로즈 업 샷이나 미디엄 샷의 경우 얼굴의 형태가 확실히 드러나서 얼굴 검출에 의해 얼굴의 영역이 거의 확실히 얻어진다. 따라서 클로즈 업 샷의 감지 성능은 95.0%이고 미디엄 샷은 90.0%로 비교적 높게 측정되었다. 하지만 그림 10의 롱 샷의 경우 얼굴의 윤곽이 확실하지 않은 경우가 많아 53.3%로 성공률이 낮게 측정되었다. 롱 샷의 경우 감지 성공률을 높여 주기 위해 얼굴 검출 이외에 보행자나 몸 인식과 같은 추가적인 방법이 필요할 것으로 판단된다.

V. 결론

샷의 유형은 영화의 서사구조와 밀접한 관련이 있다. 감독은 클로즈 업 샷을 통해서 인물의 감정을 묘사하고 롱 샷을 통해 인물이 처한 시간적·공간적 배경을 묘사한다. 이러한 샷의 유형을 알아내는 작업은 영화의 서사구조를 파악하는데 매우 유용한 방법이 될 수 있다. 이를 위해 본 논문에서는 얼굴 검출 기술을 이용하여 클로즈 업 샷과 미디엄 샷, 롱 샷을 판별하기 위한 규칙을 제시하여 샷의 유형을 판단하는 방법을 제안하고 시스템으로 구현하였다.

본 시스템에서 얼굴 검출이 잘 되는 클로즈 업 샷이나 미디엄 샷은 인식 성공률이 각각 95%와 90%로 비교적 잘 감지하였지만 얼굴이 너무 작게 보여지는 롱 샷의 경우 인식률이 53.3%로 감지하지 못하는 경우가 많이 발생하였다. 이의 개선을 위해 롱 샷에서 얼굴 검출 보다는 보행자 인식이나 몸 인식과 같은 기술을 이용하여 롱 샷 감지 성능을 향상시킬 필요가 있다. 향후 이와 관련된 기술을 적용하여 샷 유형 감지 성능을 향상 시키도록 하겠다.

참고문헌

- [1] Roy Thompson, "Grammar of the Edit," Butterworth-Heinemann, pp. 24-43, 1993.
- [2] Jonggab Kim, "The Rhetoric of Close-up: the Face of Greta Garbo," The Culture and Image, vol. 4, no. 2, pp. 5-28, December 2003.
- [3] Luca Canini, Sergio Benini, and Reccardo Leonardi, "Affective Analysis on Patterns of Shot Types on Movies," Proceeding of 7th International Symposium on Image and Signal Processing and Analysis (ISPA 2011), pp. 253-258, september 2011.
- [4] Heeman Lee, "Implementing Augmented Reality By Using Face Detection, Recognition And Motion Tracking," Journal of the Korea Society of Computer and Information, vol. 17, no. 1, pp.97-104, January 2012.
- [5] Donghyeon Kim, Jaehyun Im, Daehee kim, Taekyung Kim, and Joonki Paik, "Face Detection and Tracking using Skin Color Information and Haar-Like Features in Real-Time Video," HCI 2009, pp. 146-149, February 2009.
- [6] Joong-Gyo Jeong, Sang-Sung Park, and Dong-Sik Jang, "Real-Time face detection using the Skin color and Haar-like feature," Journal of the Korea society of computer and information, vol. 10, no. 4, pp. 113-121, September 2005.
- [7] Seung-Bo Park, Heung-Nam Kim, Hyunsik Kim, and Geun-Sik Jo, "Exploiting Script-Subtitles Alignment to Scene Boundary Detection in Movie," Proceeding of IEEE International Symposium on Multimedia (ISM2010), pp. 49-56, December 2010.
- [8] Eun-Kyoung Kim, HyunJu Kim, Hyun Mi Jo, and Yong-Hwan Lee, "Real-time Face Detection and Tracking using the AdaBoost Algorithm," Proceeding of Korean Institute of Information Technology Summer Domestic Conference, pp. 379-383, May 2011.
- [9] OpenCV, <http://opencv.org/>
- [10] OpenCV Swiki,

<http://alereimondo.no-ip.org/OpenCV>

저 자 소 개



백 영 태

1989 : 인하대학교
전자계산학과 이학사
1993 : 인하대학교
전자계산공학과 공학석사
2002 : 인하대학교
전자계산공학과 공학박사
1993-1998 : 대상정보기술(주)
정보통신연구소 선임연구원
1998-현 재 : 김포대학교
멀티미디어과 부교수
관심분야 : 멀티미디어 정보검색,
웹교육시스템, 모바일시스템
Email : hannae@kimpo.ac.kr



박 승 보

1995 : 인하대학교
전기공학과 공학사
1997 : 인하대학교
전기공학과 공학석사
2011 : 인하대학교
정보공학과 공학박사
현 재 : 경희대학교
경영대학 학술연구교수
관심분야 : 멀티미디어 정보검색,
스토리 모델링, 감정검출
Email : molaal@naver.com