

향상된 코덱으로 압축된 프레임에서 고속 얼굴 검출 기법 연구

*윤소정 **유성근 ***엄유미

서울과학기술대학교

*jamie88y@gmail.com **orcogre@gmail.com **youmi1003@nate.com

Study of Fast Face Detection in Video frames compressed by advanced CODEC

*Yoon, So-Jeong **Yoo, Sung-Geun ***Eom, Yumie

Seoul National University of Science and Technology

요약

최근 얼굴 인식 기술과 하드웨어의 빠른 발전으로 인해 실시간 얼굴 검출이 가능한 다양한 어플리케이션이 제시되고 있다. 특히 네트워크의 발달과 영상 장비의 저 비용화로 IP 기반의 네트워크 감시 카메라와 얼굴 검출 기술을 이용한 스마트 감시 카메라의 요구와 저장된 감시카메라의 영상에서 얼굴 검출을 할 수 있는 스마트 감시 시스템의 요구가 증대되고 있다. 그러나 대부분의 감시 시스템은 네트워크 대역폭과 저장 용량을 감소시키기 위하여 영상을 압축하고 있다. 압축된 영상을 전부 디코딩 하고 모든 프레임에서 얼굴 검출을 하는 것은 시스템 성능 요구사항을 증대시키므로 압축된 영상을 이용한 빠른 얼굴 검출 기법이 요구되고 있다. 본 논문은 기존의 Haar like features와 adaboost 학습기 등의 고속화된 얼굴 검출 알고리즘과 모션 정보를 이용한 프레임 저감기법을 이용하여 압축된 프레임에서 고속으로 얼굴검출을 하는 방법을 제시하고 방송 응용분야에 대해 논의 하고자 한다.

Recently, various applications using real-time face detection have been developed as face recognition technology and hardware grows. While network service is developing and video instruments costs lower, it is needed that smart surveillance camera and service using network camera based on IP and face detection technology. However, videos should be compressed for reducing network bandwidth and storage capacity in surveillance system. As it requires high-level improvement of system performance when all the compressed frames are processed in a face detection program, fast face detection method is needed. In this paper, not only a fast way of algorithm using Haar like features and adaboost learning and motion information but also an application on broadcast system is suggested.

1. 서론

생체 인식 기술 중에서 최근 얼굴 인식 기술은 방법 및 신분 확인의 목적으로 많이 사용되고 있다. 이는 2013년 미국 FBI가 보스턴 마라톤 폭과 사건과 관련하여 범인을 검거하는데 주변 CCTV(Closed-Circuit Television)의 영상과 얼굴 인식(face recognition) 프로그램을 이용하여 검거에 도움을 받았다고 발표한 사실에서 알 수 있다. [1] 이처럼 얼굴 인식 기술이 다방면에서 쓰이는데 비해 빠르게 얼굴을 인식하여 처리하는 기술은 모든 영상 프레임에서 얼굴을 검출해 내고 인식해야 하기 때문에 계산량이 많아지고 처리시간이 오래 걸린다는 문제 때문에 고속 얼굴 검출이 어렵다.[2]

이에 따라 고속 얼굴 검출을 위해 하르 유사 특징(Haar-like features)과 adaboost 학습기, cascade 분류기 등의 요소를 포함한 Viola and Jones [3] 알고리즘 등이 제안되어 왔으나 수행을 위한 개선 방법으로 프레임 저감 기법 등의 알고리즘이 필요하다.

따라서 본 논문은 앞서 얘기한 고속 얼굴 검출의 성능 향상과 계

산 효율을 높이기 위해 향상된 코덱로 압축된 영상의 프레임 단위에서 수행할 수 있는 알고리즘을 제시하고 이를 적용하여 실험할 것이다. 논문의 구성은 다음과 같다. 먼저, 기존 논문들을 통해 압축 도메인에서 기존의 고속 얼굴 검출 기법은 연구한다. 둘째, H.264/AVC와 HEVC에서의 고속 얼굴 검출 기법을 연구한다. 셋째, 도출된 알고리즘을 설명하고 마지막으로 실험 결과 및 결론을 맺는다.

2. 고속 얼굴 검출 기법 분석

2.1 압축 코덱을 고려한 고속 얼굴 검출 기법 분석

기존 고속 얼굴을 탐색하고 검출하는 방법 중 뛰어난 성능을 보이는 방법은 Viola and Jones [3] 가 제시한 것으로 빠르고 정확도 높은 탐색률을 보여준다. 이 방법의 주요 기여 요소들로는 직접영상(integral image), 하르 유사 특징(Haar-like features), adaboost 학습기, cascade 분류기가 있다. Viola and Jones의 알고리즘은 얼굴 검출 속도와 정확도가 기존의 방법보다 높다는 특징이 있는데 이는 다음과

같은 이유가 있다. 하르 유사 특징(Haar-like feature)의 합연산의 소모적인 계산을 피하고자 한 픽셀을 그 위측과 좌측의 모든 픽셀들의 면적합에 의해 만들어 지는 집적영상을 어떤 크기와 위치에서도 일정한 시간으로 계산이 가능하게 하였다. 또, adaboost 알고리즘과 cascaded 분류기를 이용하여 우수한 Haar-like feature들을 선별한 약한 분류기(weak classifier)를 구성한 뒤, 에러에 가중치를 적용하여 점차 강한 분류기(strong classifier)를 구성한다. 이와 같은 분류기의 직렬 cascaded 연결 구조는 수행 속도가 가속화되는 동시에 점점 강력한 분류기를 통과함으로써 정확도 또한 향상시킨다. 이러한 고속 얼굴 검출 기술의 발전에 큰 역할을 했다.

Viola and Jones 알고리즘은 이전의 알고리즘보다 고속 얼굴 검출의 처리 속도를 크게 단축시켰지만 모든 프레임에 대한 얼굴 탐색 및 검출을 시행해야하므로 계산량이 많은 것이 문제다. 이를 해결하기 위해 Viola and Jones 알고리즘과 응용·수정한 다양한 알고리즘 및 방법이 연구되어 왔다. Lee et al.[4]의 알고리즘은 프레임 단위에서 계산을 줄이기 위해 프레임간의 시간적 유사성(temporal correlation)을 이용하여 얼굴을 추적함으로써 연산해야할 프레임의 수를 감소시킬 수 있는 방법으로 제시되었다.[5]

2.2 H.264/AVC에서의 고속 얼굴 검출 기법 분석

H.264/AVC에서의 고속 얼굴 검출 기법은 움직임 추정을 통한 움직임 벡터를 통해 이뤄진다고 할 수 있다. 매크로블럭(macroblocks; MBs)으로 나뉜 프레임들은 시간적으로 순차적인 전(previous) 프레임과 현재(present) 프레임내의 탐색 영역(searching window)에서 후보블록을 탐색함으로써 MBs의 위치를 추정(estimation)하고 움직임 벡터(motion vector)를 만든다.[7] H.264/AVC 이전 코덱에서는 시간적 유사성(temporal correlation)과 공간적 유사성(spatial correlation)을 이용해 영상을 압축했다면 H.264/AVC에서는 움직임 정보를 프레임 공간 내에서 예측하는 인트라 예측 모드(Intraprediction)와 참조 프레임과 현재 프레임의 차이에서 예측하는 인터 예측 모드(Inteprediction)로 압축을 시행한다. 이에 얼굴 탐색 알고리즘을 인터 예측 모드를 이용해 탐색한다면 관심 영역의 크기를 줄일 수 있고 이에 따라 연산량을 감소시킬 수 있다.[4]

3. 제안하는 알고리즘의 개요

본 알고리즘의 목표는 H.264/AVC 코덱 환경에서 사용하고 있는 움직임 추정 과정을 통해 형성된 움직임 벡터를 활용하여 얼굴 검출 및 인식을 수행 시의 연산량을 줄이는 데에 있다. Viola and Jones[3]의 알고리즘에서의 요소인 하르 유사 특징(Haar-like features)을 이용하였다. 먼저, 이 알고리즘은 압축된 영상을 움직임 벡터 계산기(motion vector caculator)라 명명한 결정자에게 넣는다. 이 계산기는 파싱(parsing)과 비슷한 기능을 하면서 벡터(vector)값을 스칼라(scala)값으로 바꾸어 일정한 한계치(T)를 넘으면 프레임 내 움직임이 있는 것으로 판단하는 기능을 한다. 움직임이 있다고 결정이 내려지면 얼굴 검출을 하게 되고 얼굴 인식기로 보내지게 된다. 이 때 움직임과 얼굴이 검출되지 않았을 경우, 그 프레임은 버려지게 되며 연산하지 않아도 된다. 이로써 처리해야 하는 프레임 수를 줄여줌으로써 고속 얼굴 검출을 구현할 수 있게 된다. 그림 1은 움직임 벡터를 이용한 얼굴 검

출 및 인식 방식에 대한 순서도이다.

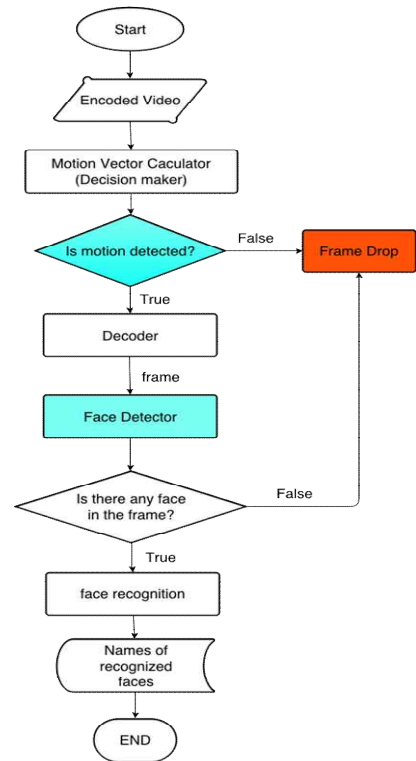


그림 1. 움직임 벡터를 이용한 얼굴 검출 및 인식 방식에 대한 순서도

4. 실험 환경

실험을 위해 arecontvision사에서 제공하는 H.264/AVC로 인코딩된 Trade Show Booth 영상을 이용하였다. 이 영상은 1920X1056 해상도에 초당 17 프레임으로 인코딩 된 영상이고 총 프레임 수는 4494이다. 분석에 사용된 PC 는 Intel I7 4770 에 32GB RAM 을 장착한 PC 이고 영상을 분석한 소프트웨어는 Intel사에서 제공하는 OpenCV 2.4.8.0 라이브러리와 Visual Studio 2010버전을 이용하여 작성하였다.

본 실험은 인물이 많이 출현하는 영상을 이용하여 제안한 알고리즘을 적용한 방법과 적용하지 않은 방법으로 나눠서 실험한다. 평가 요소로는 1)프레임 처리 속도와 2)얼굴 검출 정확도이다. 두 실험 상황에서 평가 요소들을 측정하고 비교하여 제안한 알고리즘의 성능을 시험한다.

5. 실험 방법 및 결과

실험에서 전반적으로 이용할 프로그램은 하르 유사 특징(haar like features) 중 frontalface_alt_tree와 eye_tree_eyeglasses를 이용하여 얼굴을 검출 방식을 사용한다. 이에 더하여 움직임 벡터를 검출하여 처리할 프레임을 정함으로써 처리에 효율성과 성능치 개선을 목표로하였다. 먼저 움직임 벡터를 이용하지 않고 모든 프레임에 대해 얼굴을 검출하는 실험을 진행한다. 그 뒤 동일한 영상을 알고리즘을 적용한 검출 프로그램을 통해 얼굴 검출을 시행한다. 이 두 실험에 대한 대조군으로 프로그램에서 추출한 영상 내 모든 프레임(총 1000 프레임)에서 육안

(naked eyes)으로 검출할 수 있는 얼굴의 수가 될 수 있다. 이 프로그램에 경우 얼굴 검출된 프레임은 이미지로 저장하도록 하여 비교 가능하게 한다. 또한 움직임 검출이라 판정할 픽셀 개수(면적; 크기)를 나타내는 T가 중요한 개념인데 영상의 크기(size)에 종속된다. 이를 표현하면 다음과 같다.

$$\text{width} * \text{height} * k = T \text{ (k는 화면 비율 계수)}$$

위 식에서 본 것과 같이 화면의 크기에 따라 움직임을 판별하는 윈도우(window; mask) 사이즈가 변해야 함을 알 수 있다. 또한 알고리즘에 따라 움직임 벡터가 검출된 부근의 블록을 얼굴 검출기로 보내어 움직임이 없는 부분에서는 검출이 행해지지 않도록 하여 계산량을 낮추려 한다.

표 1은 위의 방법으로 시행한 실험의 결과를 나타낸 것이다. (단, 하나의 얼굴도 검출하지 못한 프레임은 0% 정확도로 하고 얼굴과 얼굴이 아닌 것을 같이 검출 한 경우인 오류검출의 정확도를 50%로 두었다.)

	T	프레임 처리 시간 (초)	검출된 프레임 수	불검출 프레임 수	오류검출 프레임 수	얼굴 검출 정확도(%)
Full frame detection	2,000	2,305.41	782	112	255	69.4
Algorithm applied detection	1,000	1,140.30	347	57	103	68.7
	2,000	352.08	72	14	20	66.7
	2,500	170.02	16	4	3	56.2

표 1. 알고리즘 적용한 실험 비교 결과

알고리즘을 적용하지 않은 경우에는 정지해 있는 물체에도 얼굴이라 검출하였으며 빛이나 화상의 밝기에도 민감하게 반응하여 잘못된 검출을 행했다. 하지만 알고리즘을 적용한 검출기에 경우, T의 크기를 크게 함에 따라 정확도가 달라졌지만 모든 프레임을 검출하는 알고리즘과 비교하였을 때 정확도에는 큰 차이가 없었다. 특히, T값이 2,000일 경우, 두 검출을 비교해 보면 알고리즘이 적용된 검출기가 처리 시간이 약 14배 빠르고 정확도는 2.7 퍼센트 포인트(pp) 낮게 측정되었다. 이에 프레임 처리 시간은 줄어들면서 비교적 대등한 정확도로 검출된 것을 알 수 있다. 하지만 이는 검출된 프레임에 한한 정확도이고 대조군과 비교하였을 경우, 모든 프레임을 적용하는 검출기가 새로운 얼굴에 대한 검출도가 더 높은 것으로 사료된다. 그림 2는 같은 프레임에서 두 검출기를 비교한 것이다. 새로운 얼굴에 대해서 검출 정확도가 다른 것을 확인할 수 있다.



그림 2. 두 검출기에서 얼굴 검출 프레임 비교

6. 결론

실시간 얼굴 검출 기술은 Viola and Jones [3]의 알고리즘을 통해 비교적 빠르고 정확한 검출이 가능하게 되었다. 최근 스마트폰의 보급에 따른 얼굴 인식 어플리케이션(application)과 ব্যবধ으로 쓰이는 등 기술적 진보를 이뤘다. 하지만 많은 데이터량을 고속으로 처리할 수 있는 얼굴 검출 기술이 대두되었다. 이에 본 논문은 H.264/AVC 코덱에서의 움직임 정보를 이용하여 불필요한 프레임을 줄여 계산량을 줄이는 알고리즘이 제안하였다. 실험을 통해 프레임 저장하는 편이 시간 대비 정확도가 높으며 이에 성능비도 대등한 것으로 나타났다. 하지만 제안한 알고리즘은 처리 속도가 빠르나 새로운 얼굴에 대해 검출도가 낮아 신뢰도면에서 높다고 말할 수 없다. 이는 앞으로의 얼굴 검출 및 인식 기술에서 해결해 나가야 할 것이다.

고속 얼굴 검출 기술의 발전은 최신 영상 데이터량이 많아지는 현상에 적격일 것으로 보인다. 방송 콘텐츠 편집 및 저장 시에 얼굴을 검출하고 인식하여 태그(tag)하여 얼굴 단위로 편집 및 저장이 용이하게 할 수 있다. 이는 도래하는 UHD시대의 워크플로우(workflow)라 할 수 있을 것이다.

감사의 글

본 연구는 (주)네오시큐의 "실시간 영상분석과 얼굴인식을 결합한 지능형 CCTV 영상분석 시스템 향상 연구"사업의 연구 결과로 수행되었습니다.

참고문헌

[1] <http://www.indaily.co.kr/client/news/newsView.asp?nIdx=16101&Ncate=F1002&Mcate=M1004>(2014. 5. 14)

[2] 노현진·김태완·정연모, "HD 영상의 실시간 얼굴 검출을 위한 LBP 연산의 하드웨어 설계," 전자공학회 논문지, 제 48 권 SD 편 제 10 호 pp67-71, 2011

[3] P. Viola and M. Jones, "Robust real-time object detection," International Journal of Computer Vision 57(2), 137-154, 2004

[4] T. Lee, J. Kim and H. Lee, "Fast Face Detection with Motion Prediction Using H.264 Motion Vectors," International Conference on Electronics, Information and Communication(ICEIC), 2012. (accepted)

[5] 김선웅, "H.264/AVC 정보를 이용한 얼굴 탐색과 추적", 전기컴퓨터공학부 석사논문, 서울대학교, 2012

[6] 카도노신야, 키쿠치 요시히로, 스즈키 테루히코, 『H.264 AVC 비디오 압축 표준』, 홍릉과학출판사, 2005

[7] Chiariglione, Leonardo (Ed.), *The MPEG Representation of Digital Media*, 2012