

# CCTV에서 얼굴 탐색 시간 단축 기법

박선우\*, 홍지훈\*, 원일용\*

\*서울호서전문학교 사이버해킹보안과

e-mail:supersw\_@naver.com, houn8499@naver.com, clcccc@shose.ac.kr

## A time reduction method For the Face Recognition CCTV

Seon-Woo Pak\*, Ji-Houn Hong\*, Ill-Yong Won\*

\*Cyber Hacking Security, Seoul Hoseo Technical College

### 요 약

동영상에서 학습을 통한 특정인의 탐지는 한 개의 프레임 당 많은 시간이 소모된다. 따라서 특정인의 위치를 파악하면서도 탐색할 프레임의 수를 줄이는 것이 필요하다. 우리는 관찰되는 장소의 특징을 이용하여 일정 범위를 생략하는 탐색 기법을 제안하고 성능을 측정하였다. 실험결과 제안한 방법에서 어느 정도 유의미한 결과를 얻을 수 있었다.

### 1. 서론

최근 영상 녹화 기술의 발달과 비용의 감소로 인해 다양한 영상 녹화장치들이 일반화되는 추세이다. 특히 CCTV는 개인을 감시한다는 부정적인 면도 있지만, 범죄사건 발생 시 관련된 증거를 수집하는데 큰 도움을 주고 있다. 특히 CCTV의 녹화 영상에서 추적하고자 하는 범인의 시간대별 등장을 확인하는 것은 중요한 문제이다. 대부분의 현장에서는 이러한 작업을 수작업으로 진행하고 있는데, 검색해야 하는 영상의 길이가 길어질수록 어려움이 있다. 이에 수집된 영상에서 추적하고자 하는 범인을 자동으로 인식하고 추적하는 기술에 대한 연구가 활발하게 이루어지고 있다[1].

이러한 목표를 달성하기 위해 일반적으로 사용되는 방법은 다음과 같다. 탐색하고자 하는 목표 얼굴 이미지가 확보되면, 탐색하고자 하는 동영상에서 각각의 프레임을 작업 대상으로 해당 얼굴을 포함하는 영상과 얼굴 위치는 찾는다. 작업 영상에서 얼굴 영역으로 예측되는 부분에서 비교할 얼굴을 추출하고 이것을 이용하여 목표 영상과의 유사도를 측정하여 기준 이상의 유사도를 가지면서 가장 유사도가 높은 얼굴을 목표 영상으로 판단한다[2].

이러한 접근법은 몇 가지 단점을 가지고 있는데, 그중 가장 큰 것은 처리 속도이다. 예를 들어 탐색하고자 하는 영상이 1시간이고 초당 20프레임이고 해당 프레임에서 얼굴을 추출하고 유사도를 계산하는 시간을 1초라고 가정한다면 대략 20 시간이라는 막대한 처리 시간이 필요하다는 것이다. 또 탐색해야 하는 동영상의 개수가 여러 개라면 더욱 많은 처리 시간이 필요할 것이다. 본 논문은 이러한 동영상에서 얼굴 인식 시 탐색 시간을 단축하기 위한 방법에 대한 것이다.

우리가 제안하는 방법의 핵심적 아이디어는 탐색 가능성이 높은 프레임을 선정하여 해당 프레임만을 집중적으로 탐색하는 접근법이다.

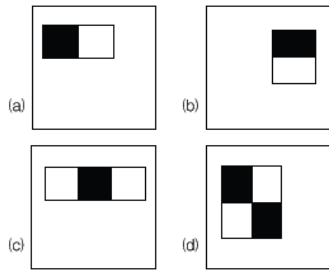
본 논문의 구성은 다음과 같다. 2장에서는 관련된 기존 기반 기술에 대하여 언급하였다. 3장에서는 제안하는 시스템을 설명하였다. 4장에서는 제안한 시스템의 성능 측정을 위해 실험과 결과를 언급하였다. 마지막 5장에서는 결론 및 향후 과제에 대하여 논의하였다.

### 2. 얼굴 영역 탐색 및 인식

일반적으로 얼굴 영역을 탐지하는 방법으로 많이 사용하는 것은 Haar-like Adaboost 알고리즘이다. 이 알고리즘은 얼굴 영역의 특징을 찾기 위해 Haar-like feature를 정의하고, 이것을 대상 위에 겹쳐 놓은 뒤 밝은 영역에 속한 픽셀 값들의 평균에서 어두운 영역에 속한 픽셀 값들의 평균 차이를 구하여 기준 값 이상이면 유사 하르 특징이 있다고 판단하는 방법이다[3].

Haar-like 마스크는 영상에서 지역적 특징을 검출하기 때문에 얼굴 영상의 특정 부분이 어느 정도 가려져도 다른 특징들이 이를 보완하게 된다.

(그림 1)은 일반 Haar-like 특징을 보여주는 것이다. (그림 1)의 (a)는 수직의 특징 성분, (b)는 수평의 특징 성분, (c)는 수직 특징 성분, (d)는 대각선의 특징 성분을 추출하는데 사용한다[3].



(그림 1) Haar-like 특징

아다부스트의 기본 아이디어는 약한 분류기를 선형적 결합에 의해 강한 분류기로 생성한다는 점이다. 약한 분류기를 학습시킨 후 올바르게 분류된 데이터에 대해서는 높은 에러 가중치가 부과되고 올바르게 분류된 데이터에 대해서는 낮은 에러 가중치를 부과한다. 또 실시간 검출에 있어서 유리하며, 복잡하고 다양한 배경에서 얼굴을 추출해 낼 수 있다는 것이 장점이다.

얼굴 인식을 하는 방법에는 다양한 방법이 있는데, 우리가 사용한 방법은 PCA이다. 이 방법은 전체 영상 공간에서 얼굴을 가장 잘 표현할 수 있는 벡터를 찾는다. 이후 얼굴영상에서 일치하는 공분산 행렬의 고유벡터를 찾고, 이것을 이용하여 유사도를 측정하여 가장 유사도가 높은 것을 해당 얼굴로 인식하는 방법이다[4].

### 3. 얼굴 탐색 시간 단축 시스템

우리는 제안하는 방법의 핵심적 아이디어는 탐색하는 프레임의 수를 줄이는 것이다. 목표 객체가 영상에 들어와서 사라지는 시간에 포함되어 있는 프레임 수를  $m$  이라고 하면 우리가 생략할 수 있는 프레임 수  $n$ 은 아래와 같이 표현 할 수 있다. 식에서  $x$ 는 마지막 프레임에서 정면 얼굴이 관측될 때까지의 프레임수를 의미 한다.

$$n \leq m - x$$

대상 얼굴이 발견된  $i$ 번째 프레임  $f(i)$ 를 중심으로 이후  $m$ 프레임만큼 얼굴이 나온다고 가정 할 수 있기 때문에 이 영역은 다시 탐색할 필요가 없다.

여기서 중요한 것은 생략하는 프레임 범위  $m$  내에 목표 영상이 있다면, 이 영상은 탐지하기 어려울 것이다. 따라서 적절한  $m$ 의 값과  $x$ 의 값을 설정하는 것이 시스템의 성능에 큰 영향을 미치게 된다는 점이다.

아래 표는 이러한 아이디어를 반영한 영상 탐색 방법의 의사코드 이다.

<표 1> 동영상에서 빠른 얼굴 탐색

<p><math>m</math>: 탐색할 영상의 총 프레임 수</p> <p><math>n</math>: 영상이 나타났다가 없어지는 평균 시간 동안의 프레임 수</p> <p><math>x</math>: <math>n</math>에서 정면 얼굴을 만나기까지의 최소 프레임 수</p> <p><math>f(i)</math>: 현재 탐색중인 프레임</p> <ol style="list-style-type: none"> <li>1. 탐색할 목표 얼굴을 학습함.</li> <li>2. 목표 얼굴 영상 및 탐색 동영상을 시스템에 입력</li> <li>3. 모든 대상 동영상 프레임에 대하여             <ol style="list-style-type: none"> <li>2.1 만약 <math>f(i)=m-x</math> 이면                 <ol style="list-style-type: none"> <li>2.1.1 <math>f(i)</math>에서 얼굴영역 인식</li> <li>2.1.2 인식된 모든 얼굴 후보 영역에 대하여 유사도 측정</li> <li>2.1.3 유사도가 기준 값 이상이면서 가장 유사도가 높은 얼굴을 목표 얼굴로 판단.</li> </ol> </li> <li>2.2 아니면 3으로 가기</li> </ol> </li> </ol>
--

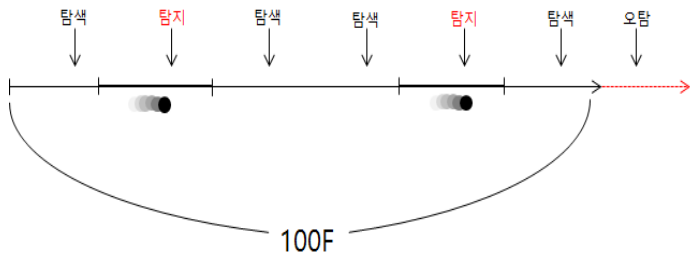
### 4. 실험 및 결과

제안한 시스템의 유용성을 평가하기 위해 C/C++을 이용하여 시스템을 구현하였고 영상처리는 OpenCV 라이브러리를 사용하였다.

320 x 240 해상도와 초당 30 프레임의 특성을 갖는 동영상을 서로 다른 장소에서 적절한 시간범위 내에서 촬영하여 10개 사용 하였다.

관찰을 통해 한사람이 등장에서 지나가는 평균 시간  $n$ 을 구하고, 최소 정면 확보 시간  $x$ 를 구하였다. 이 값을 기반으로 전체 동영상에서 탐색 하여 목표 얼굴이 탐지된 프레임을 측정하였다. 얼굴 영역 감지 및 얼굴인식은 2장에서 언급한 방법을 사용 하였다.

측정 결과는 아래 표와 같다. 표에서 No는 동영상 번호를 의미하고,  $n$ -step는 실제로 탐색한 프레임수를 의미한다. 실제로 목표 영상이 존재하는 범위에서 탐지를 했을 때, 탐지한 회수를 TP, 탐지하지 못한 회수를 FP로 표기 하였다. 또 실제 목표 영상이 아닌 곳에서 탐지 했을 때, 얼굴이 아님을 탐지하는 것을 TN, 얼굴이라고 오판한 회수를 FN으로 표시하였다.  $n$ -detect는 실제 얼굴 영역인데 전혀 탐지 시도가 되지 않은 영역의 수를 의미 한다. <그림 2>는 이러한 개념을 보여준다.



(그림 2) 측정 결과의 예

<표 2> 실험결과 표

No	m	n <sup>-</sup> step	n	x	TP	TN	FP	FN	n <sup>-</sup> detect
1	9000	113	90	10	32	70	10	1	2
2	7200	90	90	10	27	55	8	0	0
3	9000	113	90	10	30	66	12	5	2
4	12600	158	90	10	51	80	20	7	1
5	9000	113	90	10	33	70	10	0	0
6	18000	225	90	10	70	130	20	5	0
7	9000	113	90	10	30	70	11	2	0
8	10800	135	90	10	40	70	20	5	1
9	9000	113	90	10	31	66	13	3	2
10	9000	113	90	10	35	63	10	5	0

## 5. 결론 및 향후 과제

CCTV에서 얼굴 인식을 통한 특정인을 지속적으로 탐색하여 추적하는 시스템에 대한 연구는 필요하다. 현재의 기술적 방법을 동원하여 모든 프레임을 작업 대상으로 검색하면 1시간분의 동영상 처리도 엄청나게 많은 시간이 소모된다. 이러한 문제를 해결하기 위해 일정 범위의 프레임을 작업 대상에서 생략하는 방법으로 시간을 단축하는 검색 시스템을 제안하였다. 이 아이디어의 핵심은 한 장소에 인물이 나타나서 사라지는 동안의 평균적 시간을 이용하여, 첫 검색이 이루어진 프레임으로부터 그 평균적 시간까지의 프레임은 다시 검색할 필요가 없다는 가정에 기반하고 있다.

제안된 방법의 유용성을 측정하기 위해 실험을 실시하였으며 결과는 어느 정도 의미가 있었다. 다만 영상 검색을 어떤 목적으로 사용하는가에 따라 다른 의미를 부여할 수도 있다.

향후 과제는 탐색 시간을 좀 더 줄이기 위해 병렬 시스템을 도입하는 방법에 대한 연구가 필요하다. 또 여러 장소에서 동시에 검색된 자료를 이용하여 범위 내에서의 특정인의 시간과 장소를 고려한 이동 경로를 자동으로 만드는 연구가 필요하다.

## 참고문헌

[1] Woo-Sung Kang, Jin-Hee Na, Ho-Seok Ahn, Jin

-Yong Choi “Face Recognition System with SVDD-based Incremental Learning Scheme” 2006.9 The journal of Korea Robotics Society

[2] Dong-Yeop Lyu, Yong-Hwan Lim, Sun-Hee Yun, Jung-Min Seo, Chang-Hun Lee, Geun-Su Lee, Sang-Mun Lee “A User Authentication System Using Face Analysis and Similarity Comparison” 2005.11 Journal of Korea Multimedia Society

[3] Young-Hee Hong “Separated Haar-like Feature based Face Detection“ 2011.2 Thesis of Master in Sung-sil University

[4] Jung-Ui Kim “A Study on Real-time Face Recognition using PCA” 2006.08 Thesis of Master in Dong-shin University