

정규논문 (Regular Paper)

방송공학회논문지 제20권 제3호, 2015년 5월 (JBE Vol. 20, No. 3, May 2015)

<http://dx.doi.org/10.5909/JBE.2015.20.3.430>

ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

감시 영상에서 움직임 정보 분석을 통한 폭력행위 검출

강 주 형^{a)}, 곽 수 영^{b)†}

Violent Behavior Detection using Motion Analysis in Surveillance Video

Joohyung Kang^{a)} and Sooyeong Kwak^{b)†}

요 약

최근 범죄 예방을 위해 폭력행위 검출에 대한 영상 분석 기술에 대한 요구가 증가되고 있다. 영상을 이용한 행동 인식 기술을 많은 연구되고 있지만, 폭력행위에 대한 검출 기술은 상대적으로 텔레비전 또는 영화의 폭력장면 검출에만 초점이 맞추어져 있다. 영화에서 촬영된 폭력 장면에는 주로 피를 흘리는 모습들이 자주 등장하기 때문에 움직임 정보와 색상 정보를 함께 사용하는 방법을 많이 사용하였다. 하지만 실제 CCTV에서 촬영된 폭력행위의 경우 피가 묻은 장면은 자주 발생하지 않기 때문에 색상 정보를 이용한 폭력행위 검출에는 한계점이 존재한다. 본 논문에서는 영상에서의 움직임 벡터를 이용하여 감시영상에서의 폭력 행동을 검출하는 알고리즘을 제안하고자 한다. 제안하는 방법은 공개 데이터인 USI 데이터와 실제 폭력 행위가 발생한 YouTube 데이터를 사용하여 검출결과를 나타내었다.

Abstract

The demand of violence detection techniques using a video analysis to help prevent crimes is increasing recently. Many researchers have studied vision based behavior recognition but, violent behavior analysis techniques usually focus on violent scenes in television and movie content. Many methods previously published usually used both a color(e.g., skin and blood) and motion information for detecting violent scenes because violences usually involve blood scenes in movies. However, color information (e.g., blood scenes) may not be useful cues for violence detection in surveillance videos, because they are rarely taken in real world situations. In this paper, we propose a method of violent behavior detection in surveillance videos using motion vectors such as flow vector magnitudes and changes in direction except the color information. In order to evaluate the proposed algorithm, we test both USI dataset and various real world surveillance videos from YouTube.

Keyword : violence detection, behavior recognition, surveillance video

a) 전자부품연구원 멀티미디어IP연구센터(KETI)

b) 한밭대학교 전자-제어공학과(Hanbat National University)

† Corresponding Author : 곽수영(Sooyeong Kwak)

E-mail: sykwak@hanbat.ac.kr

Tel: +82-42-821-1167

ORCID:<http://orcid.org/0000-0002-4064-5108>

Manuscript received March 9, 2015; revised May 11, 2015; accepted May 11, 2015.

1. 서론

최근에는 강도, 폭력 등의 강력 범죄 및 학교폭력 등을 사전에 예방하기 위해 CCTV를 활용한 범죄예방 시스템에 대한 요구가 지속적으로 증가하고 있다. 이러한 범죄예방 시스템들은 각종 범죄로부터 실시간, 상시적으로 상황을

인식하여 폭력행위가 발생하였을 때 즉시 경보 및 대처가 가능하도록 하기 위하여 CCTV에서 폭력행위를 검출하는 영상 분석기술을 필요로 한다^{[1][2][3]}. 하지만 지능형 영상감시분야에서 행동 분석에 대한 기술은 꾸준히 연구되고 있으나 폭력행위에 대한 검출기술은 상대적으로 연구가 활발히 이루어지지 못하고 있다. 초기의 폭력행위 검출에 대한 연구는 영화에서 폭력 장면을 검출하기 위해 시작되었다. 영화에서 촬영된 폭력 장면에는 주로 피를 흘리는 모습들이 자주 등장하기 때문에 움직임 정보와 색상 정보를 함께 사용하는 방법을 많이 사용하였다^[4]. 하지만 실제 CCTV에서 촬영된 폭력행위의 경우 피가 묻은 장면은 자주 발생하지 않기 때문에 색상 정보를 이용한 폭력행위 검출에는 한계점이 존재한다. 이러한 한계점을 극복하기 위해 감시시스템에서 폭력행위 검출을 위한 알고리즘들은 일반적으로 걸어가거나 뛰어가거나 하는 사람들과 다른 움직임 정보를 나타내는 특징을 이용하기 위해 움직임 정보를 이용한 서술자를 생성하는 방법을 많이 사용하였다.

Chen 등이 개발한 MoSIFT(Motion Scale Invariant Feature Transform) 서술자는 움직임이 강한 부분에서 SIFT

(Scale Invariant Feature Transform)알고리즘으로 주요 특징점을 추출하여 비주얼 워드(Visual Word)를 생성하고, 추출된 워드(word)에 대해 HOG(Histogram of Gradient)와 HOF(Histogram of Optical Flow)를 결합한 디스크립터를 생성하여 SVM(Support Vector Machine)으로 폭력과 비폭력 행위를 검출하도록 하였다^[5]. 하지만 이러한 BoW(Bag of Words)기법의 알고리즘의 경우 실시간 폭력행위를 검출하기 어렵다는 단점이 있다. Hassner 등은 실시간 감시 시스템에 적합한 알고리즘 개발을 위해 움직임 정보만을 이용한 ViF(Violent Flows) 서술자를 제안하였다. ViF 서술자는 움직임 벡터의 크기 값이 큰 부분만을 검출하여 이진화 한 뒤, 검출된 영역의 평균을 SVM으로 학습하여 폭력행위를 검출하도록 하였다^[6]. 하지만 ViF 서술자의 경우 멀리서 촬영한 군중들의 폭력행위를 검출하는 것에는 적합하나 소수의 사람이 등장하는 폭력행위 장면 검출에는 적합하지 않다.

본 논문에서는 감시카메라 환경에서 소수의 사람이 폭력행위를 할 때 나타나는 움직임의 방향변화를 이용하여 폭력 행동을 검출하는 알고리즘을 제안한다. 즉, 사람이 걷거

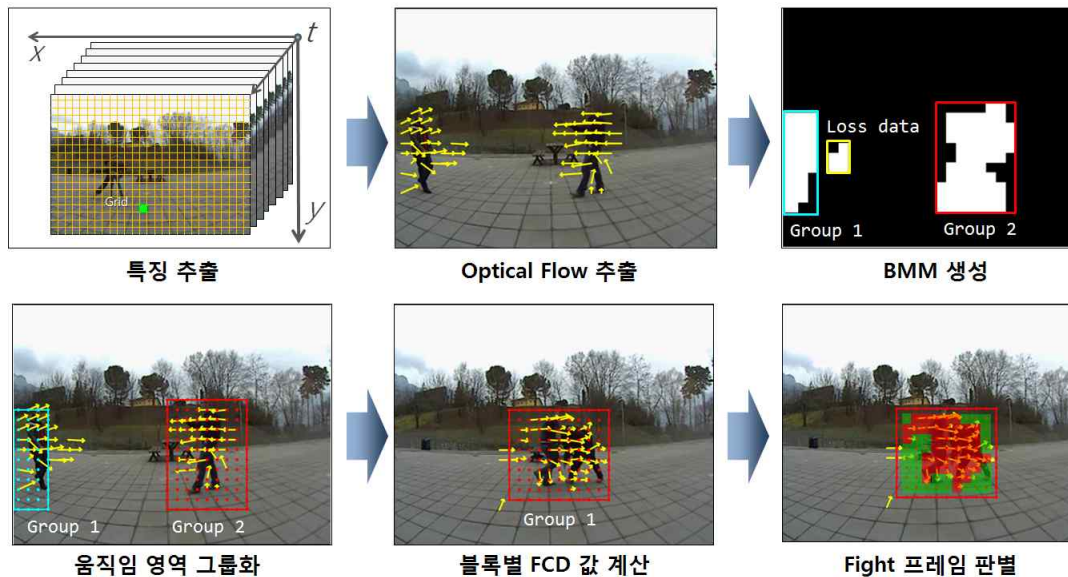


그림 1. 제안한 알고리즘의 개요
 Fig. 1. Overview of the proposed algorithm

나 뛰는 행동을 할 경우에는 움직임의 방향변화가 급격하게 발생하지 않지만 폭력행위의 경우 연속적인 프레임에서 움직임의 벡터 방향 변화가 산발적으로 나타나는 것에 착안하여 새로운 서술자를 제안하였다.

본 논문의 주요 알고리즘은 그림 1과 같다. 먼저, 영상을 $m \times n$ 의 그리드(grid)형태로 나누어 각 블록($B(m \times n)$)별로 optical flow를 이용하여 움직임이 발생하는 영역을 검출한다. 움직임이 발생하는 영역과 발생하지 않은 영역을 구분하기 위해 이진화 영상을 생성하였으며, 이렇게 생성된 영상을 Binary Motion Map(BMM)이라고 정의한다. BMM 영상에서 움직이고 있는 객체는 하나의 그룹으로 묶어주기 위하여 그룹핑 단계를 거친후, 각 그룹별 움직임의 방향 변화량 Force of Change-in-Direction(FCD)값을 계산하여 최종적으로 각 프레임단위로 폭력행위를 검출하게 된다.

II. 움직임 영역 추출 및 그룹화

본 논문에서 검출하고자 하는 폭력행위의 경우 객체의 개별 추적이 어려운 환경이므로 영상의 전역에서 움직임 정보를 추출하고 이를 이용한 새로운 서술자를 개발하여 폭력행위를 검출한다. 제안하는 방법은 움직임 정보 추출을 위해 영상 전역에서 optical flow 알고리즘을 사용하여 광류 흐름을 추출하였다. Optical flow 알고리즘은 연속하는 두 프레임 간 차이를 통해 움직임을 추정하는 방법 중

하나로써 명도를 서서히 변화시켜 영상에서 발생하는 움직임의 방향과 속도를 벡터로 나타내는 방식이다. Optical flow 알고리즘은 flow를 추출하는 밀집도에 따라 sparse optical flow와 dense optical flow로 구분할 수 있다. Sparse optical flow는 영상에서 주요 특징점을 추출한 뒤, 추출된 특징점에 대해서 움직임을 추출하는 방법으로 대표적인 알고리즘으로는 Lucas Kanade 알고리즘이 이에 해당된다. Dense optical flow는 영상의 각각의 픽셀에 대해 움직임 정보를 추출하는 방법으로 대표적인 알고리즘으로는 Gunner Farneback 알고리즘이 이에 해당된다^[8]. 일반적으로 영상에서 움직임 정보가 큰 경우 sparse optical flow가 적합하고, 움직임 정보가 작은 비디오를 분석할 경우 dense optical flow가 적합하다. 또한, Dense optical flow는 sparse optical flow보다 속도가 느린 단점이 있다. 본 논문에서는 영상을 2020 크기의 블록 단위로 분할하고 각 블록에서 sparse optical flow의 대표적인 알고리즘인 Lucas-Kanade optical flow 알고리즘을 이용하여 움직임 크기와 방향을 추출하였다.

본 논문에서는 움직임이 발생하는 영역과 발생하지 않은 영역을 구분하기 위해 20×20 단위의 블록마다 추출된 움직임 정보를 이용하여 이진화 영상을 생성하였으며, 이렇게 생성된 영상을 BMM(Binary Motion Map)이라고 정의하였다. 각 블록에서 추출된 움직임 정보는 크기 벡터가 일정 임계값 이상인 부분을 움직임이 발생하는 영역으로 판단하였다. BMM 영상에서 움직이고 있는 객체는 하나의 그룹으로 나타나고, 두 객체가 서로 떨어져서 움직이는 경우 두

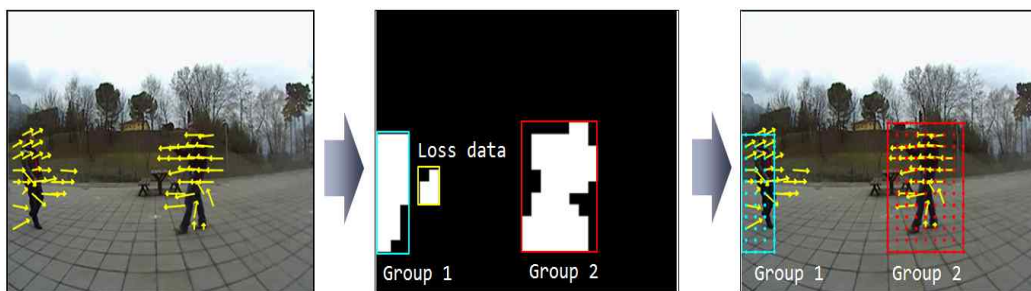


그림 2. K-means 알고리즘을 이용한 움직임 그룹핑
Fig. 2. Motion grouping using a K-means algorithm

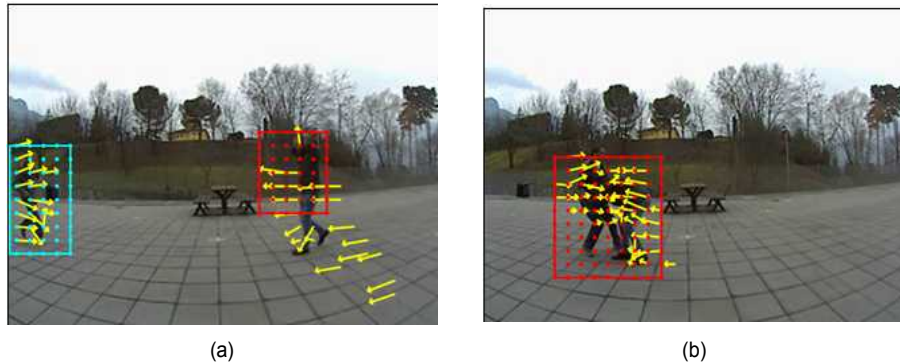


그림 3. 움직임 방향의 예 (a) 비폭력행위, (b) 폭력행위
 Fig. 3. The example of motion direction (a) Nonviolence (b) Violence

개의 객체 그룹으로 나타난다. 하지만 두 객체가 폭력행위를 하는 경우 두 객체는 하나의 객체로 나타나게 된다. 이렇게 움직임 영역을 하나의 그룹으로 그룹화하기 위해 그림 2와 같이 K-means 알고리즘을 통해 군집화 과정을 거치게 된다. 이렇게 군집화 된 하나의 그룹마다 움직임 벡터를 분석해보면 그림 3-(a)와 같이 사람이 걷거나 뛰는 경우, 하나의 그룹화 된 영역에서 추출된 움직임 방향은 동일하게 나타난다. 하지만 그림 3-(b)와 같이 두 사람이 폭력행위를 하는 장면에서는 두 사람이 붙어있기 때문에 하나의 그룹으로 결정되며, 이 그룹화 된 영역 안에서의 움직임 벡터 방향이 일정하지 않고 충돌이 일어나는 특징이 있다. 본 논문에서는 위와 같이 두 명 이상의 사람이 하나의 그룹으로 그룹화 된 영역 내에서 발생하는 움직임 벡터의 방향 변화에 초점을 두고 이를 이용하여 폭력행위를 검출하는 알고리즘을 제안한다.

III. 움직임 벡터의 방향변화를 통한 폭력행위 검출

본 논문에서는 사람이 걷거나 뛰는 행동을 할 경우에는 움직임의 방향변화가 급격하게 발생하지 않지만 폭력행위의 경우 연속적인 프레임에서 움직임의 벡터 방향변화가 산발적으로 나타나는 것에 착안하여 움직임의 방향 변화

강도를 나타낼 수 있는 FCD(Force of Change-in-Direction)를 제안하였다. FCD를 계산하기 위해 앞장에서 소개한 움직임 정보 추출 후 움직임 벡터의 각도를 4개의 방향으로 양자화 하였고, BMM 영상에서 하나의 그룹으로 판단된 영역별로 4방향으로 양자화 된 움직임 벡터의 히스토그램을 생성한 뒤 히스토그램을 정규화 하여 그룹별 각 방향에 대한 비율을 계산한다. 즉, BMM에서 그룹 1로 결정된 영역에 대해 4방향 히스토그램이 생성되고, 그룹 2로 결정된 영역에 대해 4방향 히스토그램이 생성된다.

이전 프레임과 현재 프레임의 방향 변화를 관찰하여 이들에 대해 가중치를 결정하여 각 블록($B(m \times n)$)별 방향 변화 강도 FCD를 계산한다. 이때 사용되는 가중치는 이전 블록과 현재 블록의 방향 변화가 없는 경우는 가중치는 부여하지 않았으며($\omega_0=0$), 수직 방향($90^\circ, 270^\circ$)으로 방향 변화가 발생하는 경우(ω_1), 반대 방향(180°)으로 방향 변화가 발생하는 경우(ω_2) 높은 가중치를 부여하여 각 블록별 방향 변화 강도 $FCD_{B(m \times n)}$ 을 계산하였다. 각 블록별 FCD를 계산하는 과정 $FCD_{B(m \times n)}$ 은 수식 1과 같으며 계산의 예는 그림 4와 같다. \rightarrow 방향으로 판별된 블록의 FCD 계산 방법은 그림 4의 첫 번째 줄과 같으며, \leftarrow 방향으로 판별된 블록의 경우 세 번째 줄과 같이 계산될 수 있다. 수식 1에서 d 는 양자화 된 방향의 개수($d=4$), ω_k 는 방향 변화에 따른 가중치, dr_i 는 히스토그램의 방향별 비율이다.

$$FCD_{B(m \times n)} = \sum_{i=1}^d w_k \times dr_i \quad (1)$$

그림 5는 영상 내 그룹이 2개 존재할 때와 영상 내 그룹

이 1개 존재할 때 FCD를 계산한 결과를 나타낸 것이다. 그림 5에 포함되어 있는 영상은 두 명이 천천히 걸어서 다가오다가 싸우는 행동의 시나리오로 구성되어 있다. 그림 5-(a)의 경우 그룹 1과 그룹 2는 각각에 대해 방향 히스토그램

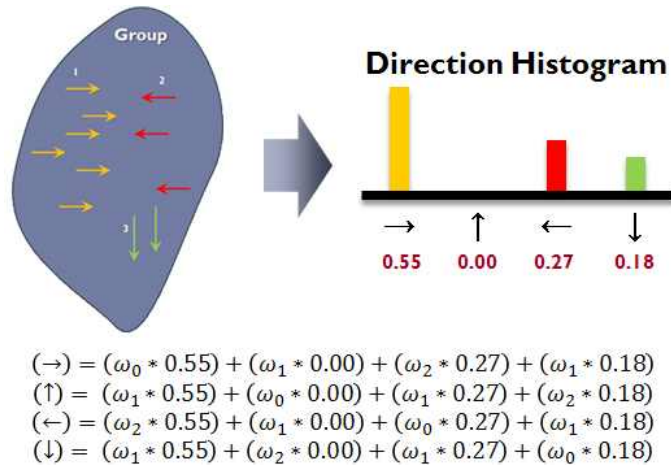


그림 4. 블록(20x20)단위 FCD 계산의 예
 Fig. 4. Example of FCD calculation for each block(20X20)

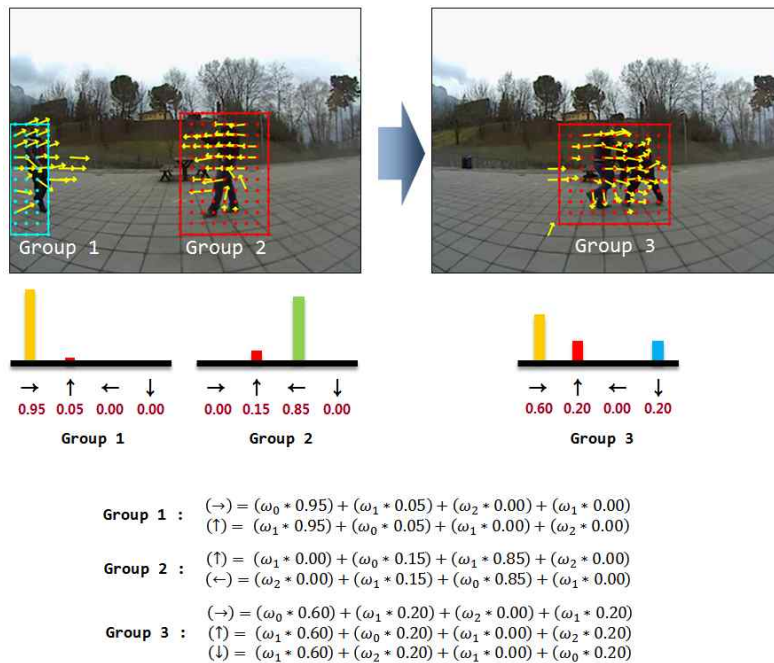


그림 5. 폭력행위와 비폭력행위의 FCD 연산의 예
 Fig. 5. Example of FCD calculation on violence and nonviolence

램이 하나의 특정 빈(bin)이 높게 나타나는 현상이 발생된다. 즉 방향이 하나의 방향으로 일관성 있게 나타나는 특징이 있다. 이 때 각 블록별 FCD 값을 계산하게 되면 그룹 1의 경우 움직임이 2개의 방향으로만 나타났기 때문에 2방향에 대한 FCD 값이 계산되며, 방향1(→)과 방향2(↑)의 경우 $\omega_0 = 0$ 이기 때문에 FCD값이 낮을 수밖에 없다. 반면에 그림 5-(b)와 같이 싸우는 행동을 하는 경우 두 명이 붙어서 싸우기 때문에 이진화 결과도 두 명이 하나의 그룹으로 결정될 수밖에 없고, 그림 3의 방향 히스토그램의 경우도 다양한 방향들이 섞여서 나온다는 결과를 볼 수 있다. 이 때 각 블록별 FCD 값을 계산하게 되면 싸우지 않는 행위보다는 높은 FCD값을 갖는 블록이 많을 수밖에 없다. 아래 식은 각 그룹별 FCD 계산 결과이다.

본 논문에서 제안한 FCD를 이용하여 폭력행위를 검출하기 위해 각 블록에서 계산된 FCD를 최근 10 프레임동안 누적하고, 누적한 결과 값에 대한 임계값을 적용하여 각 블록을 폭력블록과 비폭력블록으로 구분하였다. 이렇게 구분된 폭력블록이 K-means 알고리즘을 통해 할당된 하나의 그룹 영역에서 30%이상 차지하는 경우 현재 프레임을 폭력행위가 발생한 프레임으로 최종 판단한다. 블록별 누적된 FCD 값에 대한 예시는 그림 6과 같다.



그림 6. FCD 누적 결과
 Fig. 6. The result of cumulative FCD

IV. 실험결과

본 논문에서는 제안하는 알고리즘의 성능 검증을 위해 공개 데이터인 USI(UNITN Social Interaction)데이터^[7]와 실제 폭력행위가 있는 YouTube 데이터를 수집하여 테스트 하였다. USI 데이터는 총 64개의 비디오로 구성되어 있으며 Fight, Shake, Hug, Talk의 4가지 행동을 포함하고 있다. 하지만 이 데이터의 경우 연출한 장면이기 때문에 실제 폭력행위가 있는 감시영상을 YouTube를 통해 데이터를 수집하여 추가적인 실험 결과를 제시하였다. 그림 7과 그림 8은 제안한 FCD를 이용하여 폭력행위를 검출한 결과로써 그림 7은 연출된 폭력행위의 USI 데이터 결과이고 그림 8은 실제 폭력행위가 발생한 YouTube 데이터의 실험결과이다.

본 논문에서는 폭력행위와 비폭력행위 비디오에 대한 성능을 평가하기 위해 폭력행위가 발생하는 프레임에 대해 폭력행위에 대한 ground truth로 지정하였다. 또한, 지정된 ground truth 구간에서 80% 이상의 프레임이 폭력행위로 올바르게 판별되는 경우 폭력행위가 판명된 비디오로 간주 하였다. 즉 1개의 비디오에서 10번의 폭력행위 프레임이 존재하는 경우 8개 이상의 프레임을 폭력행위로 검출하면 올바르게 판단한 것으로 가정하였다. 본 논문에서 제안한 FCD를 이용한 폭력행위 검출 성능은 표 1과 같다. USI 데이터 실험 결과 폭력행위에 해당하는 16개 비디오에 대해 약 81.25%의 인식률을 보였고, 비폭력행위에 해당하는 비디오 48개에 대해 약 93.75%의 인식률을 보였으며, 실제 폭력행위가 발생한 YouTube 데이터에 대해 100%의 인식률을 보였다. USI 데이터에서 오분류가 발생하는 경우는 사람과 사람 사이에서 발생한 상호작용이 매우 짧은 시간 발생하거나, 상호작용이 발생하는 객체가 카메라로부터 멀리 떨어져 있는 경우에 발생하였다. 이러한 오분류가 발생하는 이유는 USI 데이터의 촬영 환경이 고정된 카메라 시점을 사용하고 있으나, 촬영된 위치가 일반적인 CCTV의 위치가 아닌 정면에서 촬영되어 원근감에 의한 움직임 정보 추출 성능이 떨어지기 때문이다. 이와 같은 이유로 발생한 오분류 현상은 USI-Shake, USI-Hug에서 발생되었으며, USI-Shake에서 발생한 오분류 현상은 카메라로부터 가까운 거리에서 두 사람이 악수를 하며 손을 흔들 때 발생하였

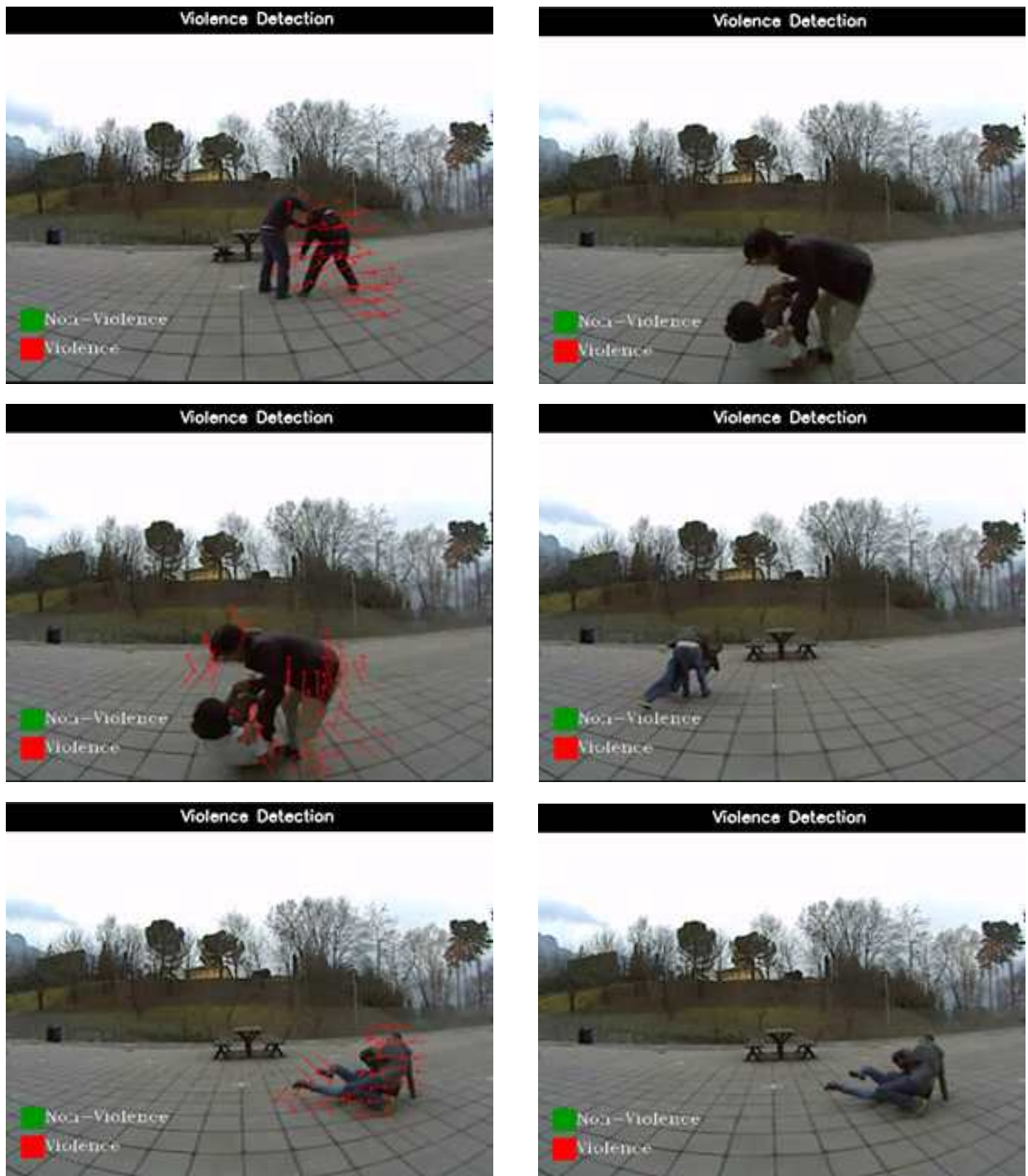


그림 7. 실험 결과 - USI_Fight
Fig. 7. The experiment results - USI_Fight

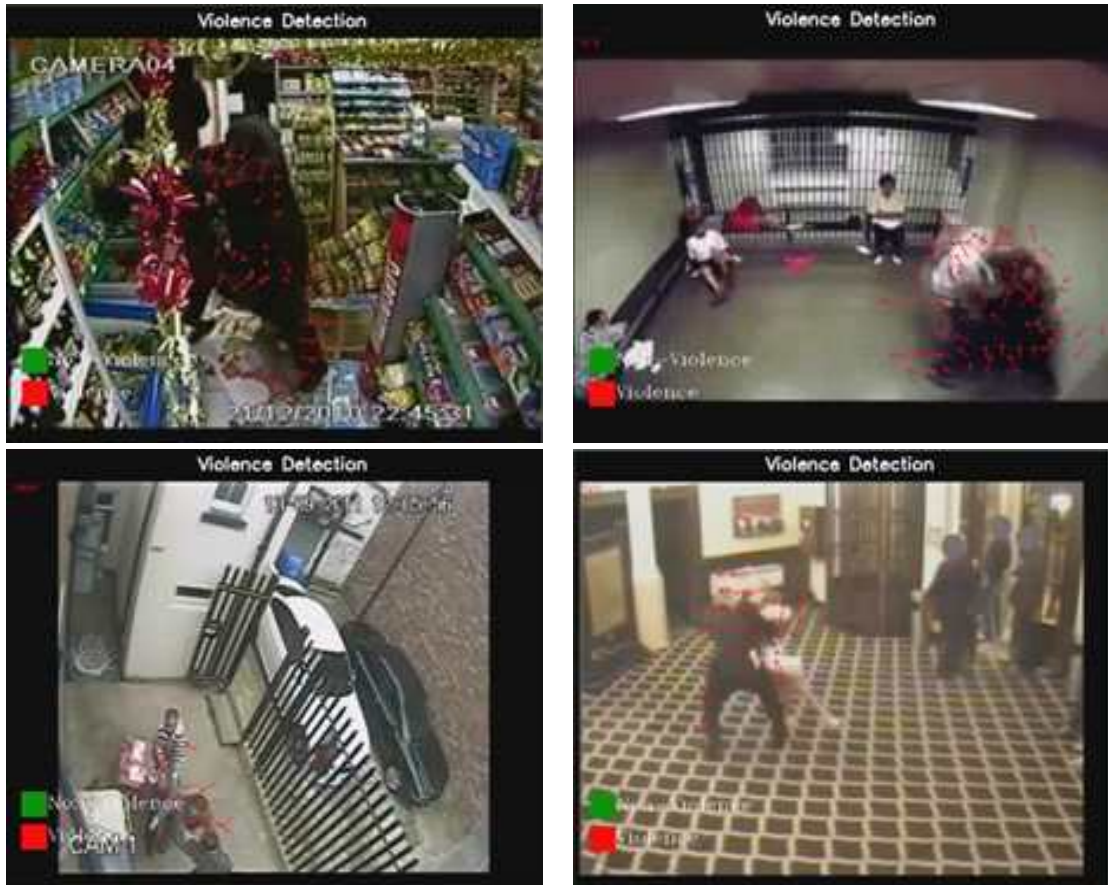


그림 8. 실험 결과 - YouTube 데이터
 Fig. 8. The experiment results - YouTube data

표 1. 폭력행동 검출 결과
 Table 1. The result of violence detection

DATA	SCENE	#of FIGHT	#of NON_FIGHT	Fight Detection	Non-FIGHT Detection
USI	Fight	16	0	13	3
	Shake	0	16	1	15
	Talk	0	16	0	16
	Hug	0	16	2	14
YouTube	Fight	6	0	6	0

다. 그리고 USI-Hug에서 발생한 오분류 현상은 그림 9-(a)와 같이 두 사람이 서로 안으려고 하는 순간의 프레임에서 추출된 FCD가 하나의 그룹에서 추출되었고, 이 때 FCD 값이 폭력행위 때 발생하는 FCD와 유사한 결과 값을 갖게 되어 Fight로 오분류 된 것으로 분석되었다. 이 외에 발생한

오분류 현상은 그림 9-(b)와 같이 객체가 카메라로부터 가까운 지점에서 이동하면서 뒤에서 이동하는 객체와 같은 그룹으로 설정되어 오분류 된 것으로 분석되었다. 폭력행위를 폭력행위로 판단하지 못한 비디오의 경우 폭력행위가 발생하는 객체가 카메라로부터 멀리 떨어져 있기 때문에

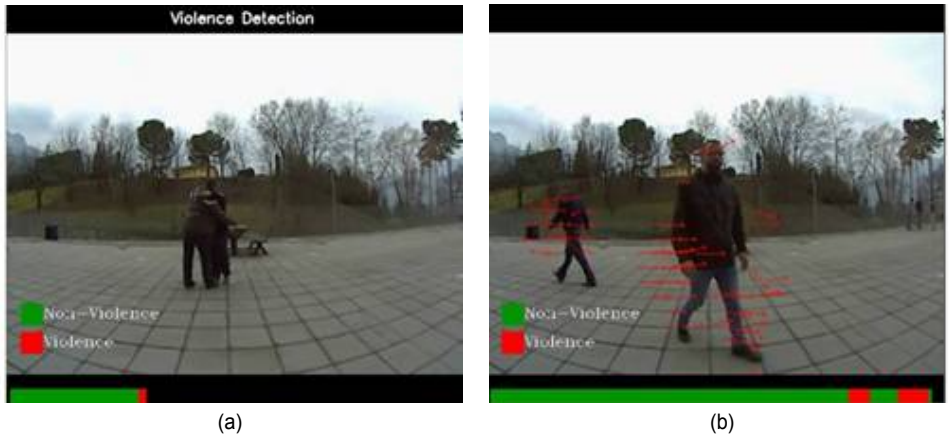


그림 9. 오검출의 예
 Fig. 9. The examples of false alarm

추출할 수 있는 움직임 정보가 적은 이유로 검출하지 못한 것으로 분석되었다. 위와 같이 발생한 오분류를 감소시키기 위해서는 움직임 정보뿐만 아니라 추가적인 특징 정보가 요구되며, 다양한 실제 폭력행위 데이터 분석을 통해 분류의 오류를 줄이고자 한다.

V. 결론 및 향후 연구방향

본 논문에서는 FCD 특징을 이용하여 CCTV 환경에서 발생할 수 있는 폭력행위 검출 방법을 제안하였다. 제안하는 방법의 성능을 검증하기 위해 공개 데이터인 USI 데이터를 사용하였으며, 실세계에서 일어날 수 있는 실제 폭력행위에 적용할 수 있도록 실제 폭력행위가 발생한 YouTube 데이터를 사용하여 추가 실험을 진행하였다. 실험 결과 640×480 영상에서 평균 20fps 속도로 검출이 가능하기 때문에 실시간 검출이 가능하다. 본 논문에서 제안한 폭력행위 검출 방법은 실세계에서 발생한 폭력행위를 검출할 수 있는 가능성을 제시하였으며, 다양한 환경에서 발생할 수 있는 폭력행위에도 적용할 수 있는 연구들이 요구된다. 또한, 환경의 변화나 객체의 크기 변화에도 변함없는 성능을 보여줄

수 있도록 개선할 예정이다.

참고 문헌 (References)

- [1] http://www.nia.or.kr/bbs/board_view.asp?BoardID=201111281502566361&id=12294&Order=010100&search_target=&keyword=&Flag=010000
- [2] Ho sik Park, "Real-time Recognition and Tracking System of Multiple Moving Object", J-KICS, pp. 421-427, 2011.
- [3] Guntae Bae, "Detection of Abnormal Behavior by Scene Analysis in Surveillance Video", J-KICS, pp. 744-752, 2011.
- [4] W. Zajdel, J.D. Krijnders, T. Andringa, and D.M. Gavrilu, "CASSANDRA: audio-video sensor fusion for aggression detection", IEEE Int Conf. on Advanced Video and Signal Based Surveillance, pp. 200-205, 2007.
- [5] M. Chen, L. Mummert, P. Pillai, A. Hauptmann, and R. Sukthankar, "Exploiting multi-level parallelism for low-latency activity recognition in streaming video", ACM SIGMM Conf. Multimedia systems, pp. 1-12, 2010.
- [6] T. Hassner, Y. Itcher, and O. Kliper-Gross, "Violence flows: real-time detection of violent crowd behavior", IEEE Int Conf. on Computer Vision and Pattern Recognition, pp. 1-6, 2012.
- [7] USI Dataset, <http://mmlab.science.unitn.it/USID/>
- [8] Gunnar Farneback, "Two-frame motion estimation based on polynomial expansion", Lecture Notes in Computer Science, pp. 363-370, 2003.

저 자 소 개



강 주 형

- 2012년 2월 : 한밭대학교 제어계측공학과 학사
- 2014년 2월 : 한밭대학교 제어계측공학과 석사
- 2014년 3월 ~ 현재 : 전자부품연구원 멀티미디어IP연구센터 연구원
- 주관심분야 : 영상처리, 지능형 감시시스템



곽 수 영

- 2010년 2월 : 연세대학교 컴퓨터과학과 공학박사
- 2010년 3월 ~ 2011년 1월 : 삼성전자 영상디스플레이사업부 책임연구원
- 2011년 2월 ~ 현재 : 한밭대학교 전자-제어공학과 부교수
- ORCID : <http://orcid.org/0000-0002-4064-5108>
- 주관심분야 : 영상처리, 컴퓨터비전, 지능형시스템