일반논문 (Regular Paper)
방송공학회논문지 제22권 제2호, 2017년 3월 (JBE Vol. 22, No. 2, March 2017)
https://doi.org/10.5909/JBE.2017.22.2.234
ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

움직임 변화 특성기반의 실시간 폭력영상 검출

김 광 수[a], 김 웅 태[a], 곽 수 영[a]‡

Real-time Violence Video Detection based on Movement Change Characteristics

Kwangsoo Kim[a], Ungtae Kim[a] and Sooyeong Kwak[a]‡

요 약

본 논문에서는 비디오 영상내 사물의 움직임의 방향과 크기의 변화를 이용한 새로운 서술자를 정의하고 이를 기반으로 하여 실시간으로 폭력 영상을 검출하는 방법을 제안한다. 새로 정의된 서술자는 폭력 행위의 움직임의 크기 및 방향 변화량이 일반적인 움직임에 비해 매우 크다는 관찰에 착안한 것이다. 일정한 프레임 동안의 서술자 값으로 이루어진 서술자 특징 벡터를 얻었고, 이것은 SVM(Support Vector Machine)으로 학습된 분류기를 통하여 폭력행위와 비폭력행위를 구별하는 데에 사용되었다. 제안하는 방법의 성능을 검증하기 위해 ViF(Violent Flow) 알고리즘과 세 종류의 데이터셋을 이용하여 비교 실험을 수행하였고, 모든 경우에서 더 우수한 성능을 보임을 확인하였다.

Abstract

A real-time violence detection algorithm based on a new descriptor using the magnitude and direction changes of movement in images is proposed. The descriptor was developed from the observation that the changes of violent actions are much larger than those of normal movements. Descriptor feature vectors consisting of descriptor values during several frames are obtained and these are inputs to SVM(Support Vector Machine) classifier for discriminating violence actions from and non-violence actions. Comparison experiments between the ViF(Violent Flow) and the proposed algorithm were conducted with three different types of datasets. The experimental results show that the proposed algorithm outperforms the ViF in every case.

Keyword : violence detection, movement descriptor, SVM classifier

a) 한밭대학교 전자·제어공학과(Dept. of Electronics and Control Engineering, 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 February 1, 2017; Revised March 13; Accepted March 13.

Ⅰ. 서 론

폭력이 사회문제로 대두되기 시작한 1960년대부터 폭력적인 영화나 TV 프로그램의 시청이 인간의 폭력적 행동을 부추긴다는 주장이 수십 년 전부터 제기되어왔다. 특히 Gerbner는 영상매체인 텔레비전은 자극적인 시각메시지를

전달함으로써 청소년들의 인지, 태도, 행동에 가장 큰 영향을 미치는 미디어라고 주장한다[1]. 또한, 한국 언론재단의 2013년 언론 수용자 의식조사에 따르면, 성인은 하루 평균 334.4분이라는 긴 시간 동안 미디어에 노출되어 있으며, 폭력적인 미디어의 숫자는 나날이 증가하고 있다[2]. 이러한 배경으로 인하여 폭력 영상을 자동으로 검출하여 자극적인 영상을 필터링 할 수 있는 기술이 요구되고 있다[3].

본 논문에서는 저수준의 특징 정보를 이용하여 실시간으로 폭력 영상을 검출하는 방법에 대해 제안한다. 일반적으로 저수준 특징기반의 폭력행위 검출 방법에 대한 연구는 폭력행동에서만 나타나는 특징정보들을 기반으로 한 서술자를 만들고, 이 서술자를 이용하여 폭력행동이 존재하는 비디오와 폭력행위가 존재하지 않는 비디오를 학습시켜 분류하는 방향으로 수행되고 있다. 저수준 특징으로는 Bermejo[4]의 MoSIFT(Motion Scale-Invariant Feature Transform)를 이용한 방법과 Hassner[5]의 ViF(Violent Flow)를 사용한 방법, 그리고 Wang[6]의 3차원 궤적 서술자를 이용한 방법을 대표적인 예로 들 수 있다. Bermejo는 전미아이스하키리그(NHL)를 촬영한 비디오를 대상으로 몸싸움이 일어나는 영상에 대한 MoSIFT서술자를 생성하였다. MoSIFT는 Motion과 SIFT의 합성어로 HOG(Histogram of Oriented Gradients) 서술자와 HOF(Histogram of Oriented optical Flow) 서술자를 결합한 것이다. 그러므로 Bermejo의 방법은 폭력행위에서 나타나는 외형 정보와 움직임 정보를 함께 서술자로 표현하여 폭력행위를 검출한다. MoSIFT는 외형정보와 움직임정보를 함께 표현하는 장점을 지니지만, 서술자를 생성하는데 연산시간이 길어 실시간으로 이용하기에는 어려운 단점을 지니고 있다. Hassner는 군중의 폭력행위 검출에 초점을 맞추어 ViF 폭력 서술자를 개발하였다. ViF 서술자는 이전 프레임의 광류와 현재 프레임의 광류의 크기와 방향을 비교하여 변화가 일정 크기 이상인 지역을 이진영상으로 만들고 일정 프레임을 누적한 히스토그램을 이용하여 생성한 서술자이다. ViF 서술자는 군중의 폭력행위 검출에는 적합하지만 소수 사람의 개별 폭력 행위 검출에는 적합하지 않다. Wang은 비디오 영상에서 움직임이 있는 영역의 궤적을 추적하는 3차원 궤적 서술자를 개발하였다. Wang은 입력되는 영상을 일정 프레임동안 누적하여 Cuboids를 만들고, Slow Feature를 이용하여 하나의 히스토그램으로 표현하는 방법을 이용하였다. Wang의 방법은 영상 전체의 움직임을 표현하기는 좋지만, 실시간 검출이 어렵다는 단점이 있다. 본 논문에서는 기존 연구가 가지고 있는 문제점 중 하나인 실시간 이슈를 해결하면서 검출의 정확도를 높일 수 있는 방안에 대해 제안하고자 한다.

Ⅱ. 움직임의 방향 및 크기 변화를 이용한 서술자

폭력행위 영상과 일반적인 사람의 움직임만 포함된 영상을 비교해 보면, 사람이 걷거나 뛰는 행동을 할 때에는 움직임의 방향이 한쪽으로 일정하고 크기 또한 일정하지만, 폭력행위의 움직임의 방향은 예측하기 어렵고 그 크기의 변화 또한 매우 크다는 것을 알 수 있다. 즉 폭력행위의 움직임은 대체적으로 급변하는 움직임이 많고, 비폭력행위에서는 급변하는 움직임이 비교적 적다. 이러한 관찰 결과에 착안하여 본 논문에서는 움직임 서술자를 제안한다.

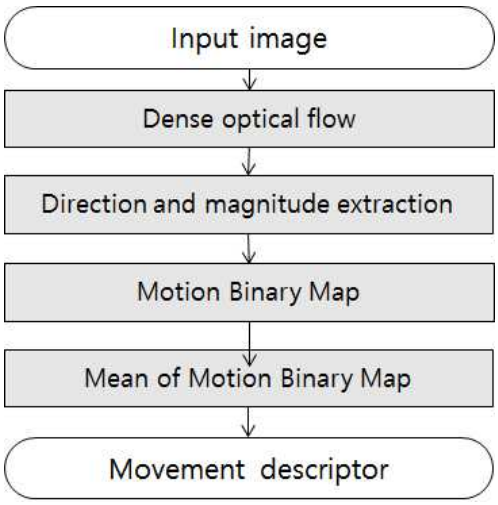

그림 1. 움직임 서술자 생성의 흐름도
Fig 1. Flowchart for movement descriptor generation

본 논문에서 제안하는 움직임 서술자는 그림 1과 같이 4개의 단계를 거쳐 생성된다. 먼저, 영상이 입력되면 15x15(픽셀)의 서브영역 단위별로 광류(dense optical flow)를 사용하여 사람의 움직임 정보를 추출한다. 서브영역으로

나누면 연산시간을 줄일 수 있고, 서술자의 차원이 너무 커지는 것을 방지할 수 있다. 각 서브영역의 움직임 정보를 추출하는 방법은 서브 영역내의 모든 움직임의 평균을 이용하는 방법과 가운데 픽셀의 움직임 정보를 대표로 사용하는 방법이 있다. 연산 속도를 빠르게 하기 위해 본 논문에서는 서브영역의 가운데 픽셀의 움직임 정보를 대표 움직임 정보로 사용하였다.

프레임 시각 t일 때, 서브영역 (x,y)에서의 대표 움직임 정보는 $\vec{v}(x,y,t)$로 표현된다. 벡터 $\vec{v}(x,y,t)$의 크기는 식(1)을 이용하여 정규화한다.

$$M(x,y,t)=\frac{|\vec{v}(x,y,t)|}{\max_{x,y}|\vec{v}(x,y,t)|} \tag{1}$$

이때 $\max_{x,y}|\vec{v}(x,y,t)|$ 값은 프레임 내에서 가장 큰 크기를 의미하여 이를 기준으로 정규화 한다. 매 프레임 마다 이전 프레임과의 크기 및 방향의 변화량을 식(2)를 이용하여 하나의 값으로 표현한다.

$$\begin{aligned} &D_M(x,y,t)\\ &=w|M(x,y,t)-M(x,y,t-1)|+\\ &(1-w)\cos^{-1}\left(\frac{\vec{v}(x,y,t)\cdot\vec{v}(x,y,t-1)}{|\vec{v}(x,y,t)||\vec{v}(x,y,t-1)|}\right)/180^\circ \end{aligned} \tag{2}$$

여기서 w는 $0<w<1$을 만족하는 상수이다. 일반적으로 $w<0.5$로 설정하였는데 이것은 움직임의 크기 변화보다는 방향 변화가 폭력행위와 비폭력행위 사이에 더 큰 차이를 보이기 때문이다.

세 번째 단계에서는 계산된 D_M을 이용하여 영상에서 움직임의 변화가 큰 영역을 추출한다. 즉, 다음 식 (3)를 이용하여 D_M이 사전에 설정된 경계값보다 큰 영역만을 추출한 움직임 이진맵(Motion Binary Map)을 생성한다.

$$MBM(x,y,t)=\begin{cases}1 & \text{if } D_M>th\\ 0 & otherwise\end{cases} \tag{3}$$

그리고 각 프레임 별 이진맵을 다음 식(4)를 이용하여 T 프레임 단위의 평균 MBM을 생성한다.

$$\overline{MBM}(x,y,t)=\frac{1}{T}\sum_{t-T}^{t}MBM(x,y,t) \tag{4}$$

최종적으로 이렇게 구한 평균 MBM을 30 프레임동안 누적하여 히스토그램을 만들고 각 서브영역별 누적값을 원소로 갖는 움직임 서술자를 생성한다. 이렇게 움직임 서술자를 얻어내는 방법을 그림 2에 나타내었다. 320x240 크기의 입력 영상을 15x15 크기의 서브영역으로 나누었기 때문에 서브영역은 총 21x16개다. 움직임 서술자는 좌측 상단 서브영역의 누적값부터 차례로 원소로 갖는 336차원의 벡터이다. 이렇게 추출된 움직임 서술자와 SVM(Support Vector Machine)을 이용하여 폭력영상과 비폭력 영상을 분류하도록 학습시킨다.

III. 실험결과

본 논문에서 제안하는 방법을 평가하기 위하여 공개 데이터인 Hockey[7]데이터와 Movie[8]데이터 그리고 폭력행위

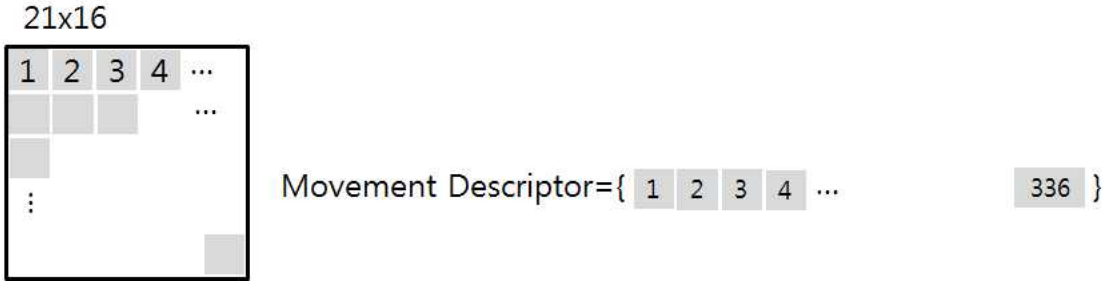

그림 2. 움직임 서술자 특징 벡터
Fig 2. Movement descriptor feature vector

를 담고 있는 Youtube 데이터를 실제로 수집하여 사용하였다. 사용된 데이터의 구성을 표 1에 나타내었다. Hockey 데이터는 NHL 하키 리그 영상 중 폭력행위 데이터 500개와 비폭력행위 데이터 500개로 구성되어있다. Movie데이터는 영화장면의 폭력행위 100개와 비폭력행위 100개로 구성되어있다. 두 데이터의 폭력행위는 특수한 상황에서의 폭력행위 장면이기 때문에 별도로 Youtube에서 폭력데이터 100개와 비폭력데이터100개로 이루어진 200개의 데이터를 추가로 구성하였다. 실시간 검출에 이용하기 위해 Youtube 데이터는 Hockey, Movie 데이터에 비해 더 긴 평균재생시간을 갖는 영상으로만 구성하였다. 그림 3은 각 실험 데이터들의 영상을 하나씩 예로 보여준 것이다.

표 1. 실험에 사용된 데이터
Table 1. Experiment dataset

	number of violent videos	number of non-violent videos	total videos	average duration time(sec)
Hockey	500	500	1000	0.5
Movie	100	100	200	2.3
Youtube	100	100	200	10.4

제안된 기법의 폭력 검출 성능은 F-measure를 이용하여 평가하였다. F-measure는 precision과 recall을 하나의 값으로 표현할 수 있는 방법이며 식(5)과 같이 정의된다. 각 데이터별 실험 결과를 표 2에 요약하였다. Hockey 데이터에 대한 성능이 평균적으로 조금 더 낮은데, 이것은 Hockey 데이터의 경우 운동하는 장면을 촬영하다 보니 카메라가 빠르게 움직이고 카메라 앵글의 변화도 많아 상대적으로 이미지에 노이즈가 많이 포함되었기 때문으로 보인다.

$$F-measure = \frac{2\times \mathrm{Precision}\times Recall}{(\mathrm{Precision}+Recall)} \quad (5)$$

표 2. 데이터별 실험 결과
Table 2. Experimental results for each dataset

	Precision(%)	Recall(%)	F-measure(%)
Hockey	80.6	79.0	79.8
Movie	81.2	81.1	81.1
Youtube	79.0	85.1	81.9

제안된 기법과 기존 방법의 성능 비교를 위해 실시간 검출이 되면서 성능이 우수하다고 알려져 있는 ViF 서술자 방법[5]과 제안된 방법의 성능을 비교하였다. 실험 데이터로는 Youtube 데이터와 Hockey데이터를 이용하였고, 실험 결과를 그림 4에 나타내었다. 그림 4에서 보는 것과 같이 제안한 방법의 폭력영상 검출 precision이 약 79%이고, ViF 서술자 방법의 precision은 약 72%로써 제안하는 방법이 더 우수한 성능을 보여주었다. 제안한 방법의 recall값도 ViF 방법보다 약 5% 정도 높은 성능을 보여주었다. ViF

그림 3. 실험 데이터의 예 (a) Hockey 데이터(b) Movie 데이터 (c) Youtube 데이터
Fig 3. Examples of experiment dataset (a)Hockey data (b)Movie data (c) Youtube data

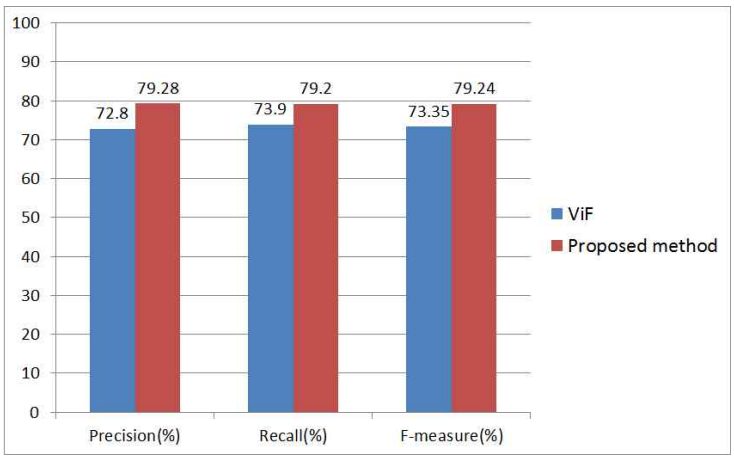

그림 4. 관련 연구와의 성능 비교
Fig 4. Performance comparison between the proposed method and the ViF method

그림 5. 오검출의 예
Fig 5. The examples of false alarm

서술자의 경우 움직임의 크기 변화만을 서술자로 표현하였는데 비해 본 논문에서 제안하는 방법은 폭력행위의 경우 움직임의 크기보다는 방향 변화가 일관성이 없다는 특징을 고려하여 움직임의 방향 정보에 더 큰 가중치를 둔 서술자를 개발하였다. 그러므로 두 방법의 성능 비교 평가 결과는 폭력행위 검출에 움직임 벡터의 방향 변화가 주요한 특징이 될 수 있음을 시사 한다고 할 수 있다. 연산 시간의 측면에서는 제안하는 알고리즘이 1초당 약 15프레임의 처리가 가능하다.

또한, 폭력행위가 아닌데 폭력이라고 판단되는 오검출의 경우, 그림 5와 같이 달려와서 안는 영상처럼 폭력행위 때 발생하는 움직임과 유사하여 잘못 분류되는 경우가 발생하는 것으로 분석되었다. 위와 같은 오분류 현상을 감소시키기 위해서는 향후 유사한 동작에 대한 세부 분석이 필요하다.

IV. 결 론

본 논문에서는 저수준의 특징정보를 이용하여 폭력행위를 실시간으로 검출하는 방법을 제안하였다. 폭력행위가 사람이 걷거나 뛰는 행동과는 움직임의 방향 변화와 크기 변화 정도가 다른 점을 반영할 수 있는 서술자를 제안하였고, 실험을 통해 제안하는 알고리즘이 기존 알고리즘에 비해 더 우수한 폭력 행위 검출 성능을 보여줌을 확인하였다. 제안하는 방법은 지능형 CCTV를 이용하여 폭력 행위를 자동으로 검출하는 시스템 개발에 유용하게 응용될 수 있을 것이다.

참 고 문 헌 (References)

[1] G. Gerbner, and L. Gross, "Living with television: The violence profile," *Journal of Communication*, Vol. 26, No. 2, pp. 172~194, 1976.

[2] Korea Press Foundation, “Survey of media audience,” 2013
[3] J. Kang, and S. Kwak “Violent Behavior Detection using Motion Analysis in Surveillance Video”, *Journal of broadcast engineering*, Vol. 20 No. 3, pp. 430-439, 2015
[4] E.B. Nievas, O.D. Suarez, G.B. Garcia, and R. Sukthankar, “Violence detection in video using computer vision techniques,” *Proceeding of International Conference on Computer Analysis of Images and Patterns*, pp 332-339, 2011
[5] T. Hassner, Y. Itcher, and O. Kliper-Gross, “Violent Flows: Real-time detection of violent crowd behavior,” *Proceeding of IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 1-6, 2012
[6] K. Wang, Z. Zhang, and L. Wang, “Violence Video Detection by Discriminative Slow Feature Analysis,” *Communications in Computer and Information Science*, Vol. 321, pp. 137-144, 2012
[7] I. Laptev and T. Lindeberg, “Space-time Interest Points,” *International Journal of Computer Vision*, Vol 64, No. 2, pp. 107-123. 2005
[8] I. Laptev, M. Marszałek, C. Schmid, and B. Rozenfeld, ,“Learning Human Actions from Movies,” *Proceeding of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1-8, 2008

저 자 소 개

김 광 수

- 2004년 2월 : 서울대학교 전기컴퓨터공학부 공학박사
- 2004년 1월 ~ 2007년 3월 : 삼성전자 통신연구소 책임연구원
- 2007년 4월 ~ 2008년 2월 : 현대자동차 차량정보기획팀 과장
- 2008년 3월 ~ 현재 : 한밭대학교 전자제어공학과 부교수
- ORCID : http://orcid.org/0000-0002-3011-2666
- 주관심분야 : 모바일로봇, 통계신호처리, 제어공학

김 웅 태

- 2013년 2월 : 한밭대학교 제어계측공학과 학사
- 2016년 2월 : 한밭대학교 제어계측공학과 석사
- 2016년 3월 ~ 현재 : 이래오토모티브 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