

H.264 압축과 SVDD를 이용한 영상 감시 시스템에서의 비정상 집단행동 탐지*

오 승 근,† 이 종 욱, 정 용 화, 박 대 희‡
고려대학교

Abnormal Crowd Behavior Detection via H.264 Compression and SVDD in Video Surveillance System*

Seunggeun Oh,† Jonguk Lee, Yongwha Chung, Daihee Park‡
Korea University

요 약

감시카메라 환경에서 군중의 비정상 집단행동 탐지란 감시카메라로부터 유입되는 영상에서 다중 객체가 위험에 처한 상황을 신속하고 정확하게 탐지하는 분야를 말한다. 본 논문에서는 CCTV 등과 같은 감시카메라 환경에서 움직임 벡터와 SVDD를 이용하여 집단내의 비정상 상황을 탐지하는 프로토타입 시스템을 제안한다. 제안된 시스템은 H.264 압축과정에서의 움직임 벡터 정보를 이용하여 영상내의 움직임 정보를 추출·표현하였으며, 비정상 집단행동의 판별 문제를 실용적 차원의 단일 클래스 분류 문제로 재해석하여 단일 클래스 SVM의 대표적 모델인 SVDD를 탐지기로 설계하였다. 제안된 시스템은 H.264 압축 과정에서 얻어지는 움직임 벡터를 이용함으로써, 실시간성을 보장하며 SVDD의 점증적 갱신 학습 능력으로 인하여 비정상 집단행동 데이터베이스의 변화에도 능동적으로 적응할 수 있다. 공개적으로 사용 가능한 벤치마크 데이터 셋인 PETS 2009와 UMN을 이용하여 본 논문에서 제안한 비정상 집단행동 탐지 시스템의 성능을 실험적으로 검증한다.

ABSTRACT

In this paper, we propose a prototype system for abnormal sound detection and identification which detects and recognizes the abnormal situations by means of analyzing audio information coming in real time from CCTV cameras under surveillance environment. The proposed system is composed of two layers: The first layer is an one-class support vector machine, i.e., support vector data description (SVDD) that performs rapid detection of abnormal situations and alerts to the manager. The second layer classifies the detected abnormal sound into predefined class such as 'gun', 'scream', 'siren', 'crash', 'bomb' via a sparse representation classifier (SRC) to cope with emergency situations. The proposed system is designed in a hierarchical manner via a mixture of SVDD and SRC, which has desired characteristics as follows: 1) By fast detecting abnormal sound using SVDD trained with only normal sound, it does not perform the unnecessary classification for normal sound. 2) It ensures a reliable system performance via a SRC that has been successfully applied in the field of face recognition. 3) With the intrinsic incremental learning capability of SRC, it can actively adapt itself to the change of a sound database. The experimental results with the qualitative analysis illustrate the efficiency of the proposed method.

Keywords: SVDD, SRC, Abnormal audio detection and identification, Surveillance system

접수일(2011년 9월 29일), 게재확정일(2011년 12월 14일)

* 본 연구는 교육과학기술부와 한국연구재단의 지역혁신인력 양성사업으로 수행된 연구결과임

† 주저자, gmo85@korea.ac.kr

‡ 교신저자, dhpark@korea.ac.kr

I. 서 론

지난 2001년의 9.11 테러 발생 후, 전 세계적으로 보안 감시 분야에 대한 지속적인 연구와 투자가 증가되고 있다[1]. 최근의 연구동향에 따르면, IT관련 보안 감시 분야의 연구는 감시 카메라로부터 획득한 멀티미디어 데이터로부터 원하는 장면 및 정보를 검색하는 단순한 보안 검색 시스템에서부터, 동영상에서 객체의 움직임을 적극적으로 검출하고 위협 상황을 미리 인지하여 알람(alert)을 통해 관리자에게 알리거나, 실제 발생한 사건 또는 이벤트들을 이해하기 위해 자동으로 객체를 인식 및 식별하고 추적하는 등 고수준의 의미 정보를 찾기 위한 연구들까지 다양하게 진행되고 있다[1-3]. 이러한 연구들 중 본 논문의 연구 주제인 감시카메라 환경에서의 비정상 집단행동 탐지는 감시카메라로부터 유입되는 감시 영상에서 단일 객체들의 세부 움직임 추적이 아닌 다중 객체가 위협에 처한 상황을 신속하고 정확하게 탐지하는 분야이다[2-3].

감시카메라 환경에서 비정상 집단행동을 탐지하는 연구의 가장 중요한 두 가지 요소 기술은 영상내의 움직임을 표현하는 방법과 비정상상황을 탐지하는 알고리즘이다. 먼저, 영상내의 움직임을 표현하는 방법은 최근 움직임 벡터를 선호하는 추세로 진행되고 있다[4-6]. 실제 감시카메라 환경에서는 동영상을 H.264와 같은 동영상 부호화 표준에 의해 압축된 비트 스트림(bit stream)의 형태로 전송 및 저장하고 있다. 따라서 원 동영상 전체를 대상으로 비정상상황을 탐지하는 기존의 방법론들[7-9] 보다는 압축된 비트 스트림 내에서 필요한 정보인 움직임 벡터만을 추출하여 비정상상황을 탐지하는 방법이 보다 설득력이 있다. 본 논문에서는 실제 감시카메라 환경을 고려하여 H.264 압축 과정에서 자연스럽게 얻어지는 움직임 벡터를 이용하여 영상내의 움직임 정보를 경제적으로 추출·표현하고자 한다. 반면, 탐지기와 관련된 선행 연구들은 GMM(Gaussian Mixture Model)이나 LDA(Latent Dirichlet Allocation)와 같은 전통적인 확률론적 패턴인식 알고리즘[7-8]을 비롯하여, 최근에는 패턴 분류 및 함수 근사 등의 문제에서 우수한 성능을 보이는 SVM(Support Vector Machine)을 집단행동 기반의 감시시스템에 적용하는 연구들이 흥미롭게 발견된다[9]. 특히, Pathan 등[9]은 이진 클래스 분류기인 SVM을 사용하여 집단행동내의 비정상상황을 탐지하였다. 일반적으로 이진클래스 분류를 위해서는

각 클래스의 훈련 집합에 사용하는 데이터의 양이 일정해야 한다. 그러나 비정상 집단행동 판별 문제에서는 입력되는 대부분의 데이터가 정상 데이터인 반면 극소수의 데이터만이 비정상 데이터이다. 따라서 훈련 집합의 스펙트럼 및 크기에 많은 차이가 난다. 결국, 이진 분류기인 SVM을 이용하여 비정상 집단행동을 분류할 경우, 관측되지 않은 영역을 포함하여 결정 경계면을 생성하므로 새로운 테스트 데이터에 대해서 오분류를 할 가능성이 높다. 그러므로 본 논문에서 다루고자 하는 비정상 집단행동을 탐지하는 문제에는 이진 분류기인 SVM보다는 해당 클래스만을 독립적으로 표현하는 단일 클래스 SVM으로 결정 경계면을 결정하는 것이 효율적이다.

본 논문에서는 위에서 언급한 바와 같이, 감시카메라 환경에서 비정상 집단행동을 탐지하는 연구의 가장 중요한 두 가지 요소 기술 중, 영상내의 움직임을 표현하는 방법으로는 실제 감시카메라 환경의 H.264 압축과정에서 자연스럽게 얻어지는 움직임 벡터를 이용하여 영상내의 움직임 정보를 경제적으로 추출·표현하였으며, 비정상 집단행동의 판별 문제를 실용적 차원의 단일 클래스 분류 문제로 재해석하여 단일 클래스 SVM의 대표적 모델인 SVDD(Support Vector Data Description)를 탐지기로 설계하였다. 본 논문에서 제안하는 감시카메라 환경에서의 비정상 집단행동 탐지 시스템은 다음과 같은 특성을 갖는다. 첫째, H.264 영상 압축 과정에서 자연스럽게 얻어지는 움직임 벡터를 이용함으로써, 추가적인 특징 추출 모듈 및 영상 분석을 위한 디코딩 모듈을 필요로 하지 않는 경제적인 시스템 구축이 가능하다. 둘째, 영상의 크기가 다르더라도 데이터의 재구성을 필요로 하지 않으며, H.264 영상 압축 모듈을 포함하고 있다면 어떠한 감시카메라 환경에서도 시스템 구축이 가능한 융통성을 갖는다. 셋째, 탐지기로 SVDD를 사용하여 정상 데이터만을 학습함으로써, 감시카메라 환경에서의 비정상 데이터의 구성문제를 해결함과 동시에 정상 데이터가 아닌 데이터는 추가 학습을 하지 않음으로 불필요한 연산을 줄인다. 넷째, 상황에 따라 새로운 클래스의 비정상 데이터가 추가되더라도 이미 정상 데이터만으로 학습을 마친 SVDD를 통해 추가 학습과정 없이 자동으로 비정상상황 탐지를 수행함으로써, 시스템의 점증적 갱신 및 확장을 보장한다.

본 논문의 구성은 다음과 같다. 2장에서는 비정상 집단행동 탐지와 관련된 연구방법론 및 관련 연구의 문헌고찰을 간략하게 서술하고, 3장에서는 본 논문

서 제안하는 비정상 집단행동 탐지 프로토타입 시스템에 대해 상세히 기술한다. 4장에서는 실험 결과 및 성능 분석을, 마지막으로 5장에서는 결론 및 향후 연구 과제에 대해 논한다.

II. 관련 연구

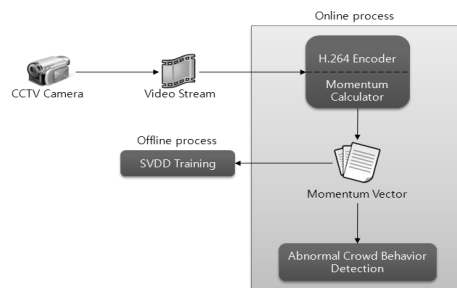
감시카메라 환경에서 집단 내의 비정상상황을 탐지하는 연구는 집단 내의 밀집도를 이용하는 정적(static) 방법과 집단 내의 움직임 정보에 기초한 동적(dynamic) 방법으로 분류할 수 있다[10]. 정적 방법은 하나의 영상 장면에 일정 이상의 사람이 유입되어 미리 정해진 임계값과 차이가 날 때를 비정상상황으로 결정하는 방법이며, 동적 방법은 하나의 영상 장면에서 사람들의 움직임이 급변하게 발생할 때를 비정상상황으로 결정하는 방법이다. 반면, 움직임을 표현하는 방법으로는 객체 추적 방법과 움직임 기반 접근 방법이 있다 [11-12]. 객체 추적방법은 움직이는 객체의 궤도 분석에 초점을 맞추어 비정상 집단행동을 탐지하기 때문에 실제계 상황의 복잡한 장면에서 움직이는 모든 객체에 대한 추적을 해야만 한다. 이는 과도한 계산량을 필요로 할 뿐만 아니라 정확한 결과 값을 얻기도 힘들다는 단점이 있다. 움직임 기반 접근 방법은 차영상(background subtraction), 옵티컬 플로우(optical flow), 혹은 움직임 벡터를 이용하는 방법으로 트래킹 과정을 요구하지 않기 때문에 신속하게 비정상 집단행동을 탐지할 수 있는 방법이다[13]. 본 논문에서는 영상의 움직임 벡터를 이용하여 감시카메라 환경에서 유입되는 동영상으로부터 움직임 정보를 추출 및 표현(움직임 기반 접근 방법)하고, 이 움직임 정보가 급격하게 변화하는 상황을 비정상상황으로 결정하는 동적 연구방법론을 취한다.

본 논문과 관련된 대표적 선행 연구이면서 동시에 실험적 비교를 위하여 선택된 선행 연구들을 정리하면 다음과 같다. Mehran 등[7]은 집단 에너지(social force)를 통해 집단행동을 모델링하였으며 옵티컬 플로우를 사용하여 움직임 정보를 측정하고, LDA 모델로 비정상 집단행동을 탐지하는 시스템을 제안하였다. Shi 등[8]은 움직임 벡터와 위상 상관도(phase correlation)를 결합하여 움직임 정보를 측정하고, 확률적 모델인 GMM을 이용한 모델링을 통해 실시간으로 비정상상황을 탐지하는 시스템을 제안하였다. 이와 같은 GMM이나 LDA등은 학습 이론 중 학습오류(empirical error)를 최소화하는 경험적 위험 최소화

화 원칙(empirical risk minimization)을 적용한 것이기 때문에 새로운 데이터가 나타나면 시스템의 성능이 갑자기 나빠진다는 문제점을 지니고 있다. 반면 SVM은 구조적 위험 최소화 원칙(structural risk minimization)의 개념을 기초로 하기 때문에 일반화 성능이 우수하며, 비선형 분리 문제와 같은 복잡한 분류문제를 해결하기 위한 다양한 커널들을 제공한다[14]. Pathan 등[9]은 사회과학 분야에서 시작한 사회 엔트로피 이론(social entropy theory)을 토대로 집단행동을 정의하였으며 이진 클래스 분류기인 SVM을 사용하여 집단행동내의 비정상상황을 탐지하였다.

III. 감시카메라 환경에서의 비정상 집단행동 탐지 시스템

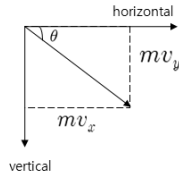
본 논문에서 제안하는 감시 카메라 환경에서의 비정상 집단행동 탐지 시스템의 구조는 [그림 1]과 같다. 이는 1개의 오프라인 처리 모듈인 SVDD 학습 모듈과 2개의 온라인 처리 모듈인 H.264 Encoder/Momentum Calculator, 그리고 비정상 집단행동 탐지 모듈로 구성된다. 각 모듈의 기능은 다음과 같다. 1) SVDD 학습 모듈에서는 정상 데이터 훈련 집합만으로 오프라인 상에서 학습을 실시한다. 2) H.264 Encoder에서는 CCTV에서 실시간으로 유입되는 영상을 H.264 포맷으로 압축하고, 압축 시 발생하는 움직임 벡터를 Momentum Calculator에서 움직임 정보의 량으로 변환한다. 3) 비정상 집단행동 탐지 모듈에서는 학습이 완료된 SVDD를 이용하여 유입되는 데이터가 정상 상황의 범주에서 벗어나는지를 실시간으로 판단하여, 정상 상황의 범주에서 벗어나면 이를 CCTV 관리자에게 통보함으로써 사건·사고에 신속하게 대처한다.



(그림 1) 비정상 집단행동 탐지 시스템의 구성도

3.1 H.264 Encoder/Momentum Calculator

본 논문에서는 영상 압축 인코더(encoder)로써, 현재까지의 영상 압축 기술 중 가장 압축률이 뛰어나며 원본 화질의 열화가 크게 발생하지 않는 H.264/AVC(Advanced Video Coding)[15]를 이용하여 영상 압축을 수행한다. H.264/AVC 영상 압축의 움직임 추정 과정에서 발생하는 움직임 벡터의 크기와 방향은 [그림 2]와 같이 각각 m 과 θ 로 표현된다(식 (1)참조)[15]. 여기서, mv_x 와 mv_y 는 움직임 벡터의 각각 가로축과 세로축 성분이다[15].



(그림 2) 움직임 벡터의 크기와 방향

$$m = \sqrt{(mv_x^2 + mv_y^2)} \quad (1)$$

Momentum Calculator에서는 식 (2)와 같이 I-frame(Infra frame) 사이에서 발생하는 P-frame(Previous frame) 또는 B-frame(Bidirection frame)의 각 모션 블록별 벡터 크기의 평균값을 계산하여 영상의 움직임량을 표현한다. 여기서, m_b 는 b 번째 블록의 벡터 크기, N_b 는 현재 프레임까지의 b 번째 블록의 개수를 의미한다.

$$\text{Momentum} = \frac{1}{N_b} \sum_{b=0}^{N_b-1} m_b \quad (2)$$

3.2 SVDD 기반의 비정상 집단행동 탐지

감시카메라 환경에서 비정상 집단행동 탐지 시스템의 가장 중요한 목적은 유입되는 영상으로부터 신속하게 비정상 상황여부를 판별하는 것이다. 여기서 비정상 상황이라는 것은 군중들이 한가롭게 걷고 있는 정상상황에서 갑자기 군중들이 동일한 방향으로 함께 뛰거나 황급히 분산하여 흩어지는 장면을 뜻한다. 이 같은 비정상 집단행동 여부의 판별 문제는 집단행동의 식별 문제와는 달리 유입된 영상이 정상 집단행동의 범주에 속하는지 아닌지를 판별하는 이진 클래스 분류

(binary class classification) 문제로 볼 수 있다. 그러나 비정상 집단행동 탐지를 수행하기 위해서는 정상 집단행동만으로 기계학습을 수행한 후, 실시간으로 유입된 영상이 정상 집단행동 범주에 속하는지 벗어나는지를 확인하는 것이 보다 실용적이다. 결국 이는 단순히 실시간으로 유입되는 영상의 이상 여부를 확인하는 과정이므로, 이진 클래스 분류 문제가 아닌 단일 클래스 분류(one class classification) 문제로 보는 것이 합리적이다. 따라서 본 논문에서는 단일 클래스 SVM의 가장 대표적인 방법론인 SVDD를 이용하여 비정상 집단행동 여부를 신속하게 판단하고자 한다.

SVDD를 이용한 단일 클래스 분류 방법은 다음과 같다[16]. d -차원 입력공간에 존재하는 N 개의 데이터로 구성되는 학습 데이터의 집합 $D = \{x_i | i = 1, \dots, N\}$ 에 대해서 R^d 위에 정의되는 중심이 a 이고 반경이 R 인 원형체 B 를 이용하여 학습 클래스의 존재 영역을 표현한다. 이 경우 각 학습 데이터 x_i 와 중심 a 사이의 거리가 R 을 초과하는 경우에는 적절한 벌점(penalty)을 부과하는 전략을 사용하여 다음과 같은 최적화 문제를 정의한다.

$$\begin{aligned} \min L_0(R^2, a, \xi) &= R^2 + C \sum_{i=1}^N \xi_i \\ \text{s.t. } \|x_i - a\|^2 &\leq R^2 + \xi_i, \xi_i \geq 0, \forall i. \end{aligned} \quad (3)$$

여기에서, ξ_i 는 i -번째 학습 데이터 x_i 가 원형체에서 벗어나는 벌점이며, C 는 반지름과 벌점항의 상대적 중요성을 조정하는 상수(trade-off constant)이다. 라그랑제 함수(Lagrange function) L 을 이용하여 위 문제에 대응하는 쌍대 문제(dual problem)를 구한다. 쌍대 문제는 다음과 같은 QP(Quadratic Programming) 문제로 정리된다.

$$\begin{aligned} \min_a \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j \langle x_i, x_j \rangle - \sum_{i=1}^N \alpha_i \langle x_i, x_j \rangle \\ \text{s.t. } \sum_{i=1}^N \alpha_i = 1, \alpha_i \in [0, C], \forall i. \end{aligned} \quad (4)$$

학습이 끝난 후 적용 과정에서 결정함수는 다음과 같이 정의된다.

$$\begin{aligned} f(x) &= R^2 - \|x - a\|^2 \\ &= R^2 - (\langle x, x \rangle - 2 \sum_{i=1}^N \alpha_i \langle x_i, x \rangle \\ &\quad + \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j \langle x_i, x_j \rangle) \geq 0 \end{aligned} \quad (5)$$

가우시안 커널을 사용할 경우 최적화 문제 식(5)는

다음과 같이 보다 간단한 형태로 단순화 될 수 있다.

$$\begin{aligned} \min_a \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j \langle x_i, x_j \rangle \\ \text{s.t. } \sum_{i=1}^N \alpha_i = 1, \alpha_i \in [0, C], \forall i. \end{aligned} \quad (6)$$

이 경우 커널 함수 ϕ 도입후의 결정함수 식(6)은 아래의 식으로 대체된다.

$$\begin{aligned} f(x) &= R^2 - \| \phi(x) - a \|^2 \\ &= R^2 - (1 - 2 \sum_{i=1}^N \alpha_i k(x_i, x) \\ &\quad + \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j k(x_i, x_j)) \geq 0 \end{aligned} \quad (7)$$

IV. 실험 및 결과분석

본 논문에서 제안한 감시카메라 환경에서의 비정상 집단행동 탐지 시스템의 정량적 성능평가 및 기존 방법론과의 비교 실험을 위하여 공개적으로 사용이 가능한 PETS 2009와 UMN(University of Minnesota) 데이터 셋[17-18]을 사용하였다(그림 3 참조). PETS 2009는 영상 감시 환경에서 사용하는 벤치마크 데이터로써 동일한 영상을 시점 및 카메라 구성에 따라 다양하게 표현하고 있으며, 군중들이 걷고 있는 정상상황에서 갑자기 군중들이 동일한 방향으로 함께 뛰는 비정상상황을 보여준다. UMN 데이터는 3종류의 실내와 실외로 구분되는 11가지의 다양한 시나리오를 포함하는 탈출(escape) 이벤트에 관한 실험 데이터로써, PETS 2009 데이터와 같이 군중들이 걷고 있는 정상상황에서 갑자기 군중들이 동일한 방향으로 함께 뛰거나 황급히 분산하여 흩어지는 비정

[표 1] UMN 데이터 셋의 시나리오 별 프레임 정보

	정상 상황의 프레임	비상 상황의 프레임
시나리오 1	1,015	324
시나리오 2	2,830	1,260
시나리오 3	1,570	560

상상황을 보여준다. 한편, 본 실험에서는 비상 상황 탐지를 위한 특징값을 추출하기 위하여 Matlab 환경에서 H.264/AVC 부호화가 수행되는 JM8.6[19]버전을 이용하여 인코딩을 수행하여 움직임 추정 과정에서 발생하는 움직임 벡터를 추출하였다.

첫 번째 실험은 UMN 데이터 셋[18]을 이용하여 Shi 등[8] 그리고 Pathan 등[9]의 방법론과 성능을 비교하는 실험이다. 먼저, UMN 데이터 셋으로부터 H.264의 압축 프레임 포맷인 4 : 2 : 0 YUV 포맷 프레임을 추출한 후 원본 데이터에 손실을 주지 않기 위해서 크기 변환 없이 초당 30프레임으로 압축을 수행하여 얻어지는 실제 움직임 벡터로 실험을 수행하였다. 전체 7,559 프레임은 정상 상황에 속하는 5,415개의 프레임과 비상 상황에 속하는 2,144개의 프레임으로 구성되었으며, 각 시나리오별 정상 상황의 프레임과 비상 상황의 프레임은 [표 1]과 같다. 각 시나리오 별로 정상 상황 프레임의 1/2을 학습 데이터로 사용하였으며, 나머지 학습에 참여하지 않은 정상 상황의 프레임과 비상 상황의 프레임으로 테스트를 수행하였다. 성능 측정을 위한 지표로는 precision과 recall[20]을 사용하였으며, 실험 결과는 [표 2]에 정리하였다.

$$\text{recall} = \frac{TP}{TP+FN} \quad (8)$$

$$\text{precision} = \frac{TP}{TP+FP} \quad (9)$$



[그림 3] PETS 2009 및 UMN 데이터 셋

[표 2] UMN 데이터 셋을 이용한 비정상 탐지 성능 비교

Performance		Precision	Recall
Shi et al.[8]	scene-1	0.98	0.95
	scene-2	0.86	0.95
	scene-3	0.99	0.92
Pathan et al.[9]	Overall Scene	0.93	0.97
Proposed Method	scene-1	0.94	0.99
	scene-2	0.99	0.97
	scene-3	0.97	0.98

위 식에서 TP(True Positive)는 정상 클래스의 입력 데이터를 정상 클래스로 정확히 분류한 데이터의 수, FP(False Positive)는 비정상 클래스의 입력 데이터를 정상 클래스로 오 분류한 데이터의 수, 그리고 FN(False Negative)은 비정상 클래스의 입력 데이터를 정상 클래스로 잘못 예측한 데이터의 수를 의미한다[20]. 결과적으로 precision은 전체 데이터 중 정확하게 비정상 군중행동이라고 판별한 데이터의 비율을 뜻하며, recall은 비정상 군중행동이라고 판단한 데이터 중 실제로 비정상 군중행동인 데이터의 비율을 뜻한다.

[표 2]에서 보듯이, 세 가지 방법론 모두 비정상 집단행동 탐지의 정확도 평가에서 매우 우수한 성능을 보이고 있음을 확인할 수 있다. 특이한 것 중 하나는 Shi 등[8]의 GMM은 첫 번째와 세 번째 장면에서는 우수한 성능을 보이나, 두 번째 장면의 실험에서는 상대적으로 낮은 성능의 precision 값을 보였으며, Pathan 등[9]의 SVM은 전체 비정상상황의 precision 값이 다른 방법론에 비해 상대적으로 낮은 것을 알 수 있다. 그러나 본 논문의 방법론은 모든 경우의 장면에서 90%이상의 좋은 성능을 보임을 확인하였다.

두 번째 실험은 PETS 2009 데이터를 이용하여 Pathan 등[9]의 방법론과 비교한 실험이다. PETS 2009 데이터[17]는 동영상 데이터가 아닌 프레임 데이터이기 때문에 각 프레임을 4 : 2 : 0 YUV 프레임으로 변환한 후 JM8.6[19]을 이용하여 H.264 포맷

으로 압축을 수행하였다. 그 결과 정상 상황 범주에 속하는 6,336개의 프레임과 2,208개의 비정상 상황 프레임으로 구성된 총 8,544개의 프레임을 얻었다. 이 중 3,150개의 정상 프레임으로 학습을 수행하였으며, 학습에 참여하지 않은 나머지 정상 데이터와 비정상 데이터를 테스트 데이터로 사용하였다. 실험 결과는 [표 3]에 정리하였다. [표 3]에서 보듯이, 두 가지 방법론 모두 우수한 성능을 보이고는 있으나 본 논문에서 제안한 시스템이 SVM기반 방법론에 비하여 보다 우수한 성능을 나타내고 있음을 확인하였다. 특히 비정상 상황 탐지 시스템에서의 주요 평가항목인 FN(False Negatives: 비정상 클래스의 입력 데이터가 정상 클래스로 잘못 예측된 데이터의 수)에서 본 논문에서 제안한 시스템이 SVM기반 방법론에 비해 우수한 성능을 나타내고 있음을 확인할 수 있었다.

V. 결론 및 향후 연구

본 논문에서는 CCTV등과 같은 감시 카메라 환경에서 움직임 벡터와 SVDD를 이용하여 집단내의 비정상 상황을 탐지하는 프로토타입 시스템을 제안하였다. 비정상 집단행동을 탐지하기 위하여 본 논문에서는 H.264 압축과정에서의 움직임 벡터 정보를 이용하여 영상내의 움직임 정보를 추출·표현하였으며, 비정상 집단행동의 판별 문제를 실용적 차원의 단일 클래스 분류 문제로 재해석하여 단일 클래스 SVM의 대표적인 모델인 SVDD를 탐지기로 설계하였다. 제안된 시스템은 다음과 같은 특성을 갖는다. 첫째, H.264 영상 압축 과정에서 자연스럽게 얻어지는 움직임 벡터를 이용함으로써, 실시간성을 보장하며 추가적인 특징 추출 모듈 및 영상 분석을 위한 디코딩 모듈을 필요로 하지 않는 경제적인 시스템 구축이 가능하다. 둘째, H.264 영상 압축 과정에서 얻어지는 움직임 벡터를 이용하기 때문에 영상의 크기가 다르더라도 비정상 군중행동 탐지기로 설계한 SVDD의 데이터의 재구성을 필요로 하지 않으며, H.264 영상 압축 모듈을 포함하고 있다면 어떠한 감시카메라 환경에서도 시스템 구축이 가능한 융통성을 갖는다. 셋째, 탐

[표 3] PETS 2009 데이터를 이용한 비정상 탐지 성능 비교

SVM (Pathan et al.[9])			SVDD (Proposed Method)		
Event	Normal	Abnormal	Event	Normal	Abnormal
Normal	95.8	4.2	Normal	98.6	1.4
Abnormal	0.8	99.2	Abnormal	0	100

지기로 SVDD를 사용하여 정상 데이터만을 학습함으로써, 감시카메라 환경에서의 비정상 데이터의 구성문제를 해결함과 동시에 비정상 군중행동에 대한 데이터는 추가 학습을 하지 않음으로 불필요한 연산을 줄인다. 넷째, 상황에 따라 새로운 클래스의 비정상 데이터가 추가되더라도 이미 정상 데이터만으로 학습을 마친 SVDD를 통해 추가 학습과정 없이 자동으로 비정상 상황 탐지를 수행함으로써, 시스템의 점증적 갱신 및 확장을 보장한다. 공개 데이터 셋인 UMN 데이터와 PETS 2009 데이터를 대상으로 제안된 시스템의 성능을 실험적으로도 검증하였다. 향후 연구 과제로는 재난 피해 방지에 중요한 정보로 사용될 수 있는 비정상 군중 행동의 주요 이동 방향 및 지역별 혼잡도 등과 같은 정보를 동영상으로부터 추출하는 후속 연구를 수행하고자 한다.

참고문헌

- [1] 전지혜, 박중화, 정철준, 강인구, 안태기, 박구만, "실시간 지능형 감시 시스템을 위한 방치, 제거된 객체 검출에 관한 연구", *한국통신학회 논문지*, 35(1), pp. 24-32, 2010년 1월.
- [2] T. Zho, R. Nevatia, and B. Wu, "Segmentation and Tracking of Multiple Human in Crowded Environments," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 7, pp. 1198-1211, July 2008.
- [3] J. Feng, C. Zhang, and P. Hao, "Online Learning with Self-Organizing Maps for Anomaly Detection in Crowd Scenes," *ICPR 2010*, pp. 3599-3602, Oct. 2010.
- [4] W. Fei and S. Zhu, "Mean Shift Clustering-based Moving Object Segmentation in the H.264 Compressed Domain," *IET Image Processing*, vol. 4, no. 1, pp. 11-18, Jan. 2010.
- [5] S. K. Kapotas and A. N. Skodras, "Moving Object Detection in the H.264 Compressed Domain," *Image System and Technique (IST) 2010 IEEE International Conference*, pp. 325-328, July 2010.
- [6] N. Kiryati, T. Raviv, Y. Invanchenko, and S. Rochel, "Real-time Abnormal Motion Detection in Surveillance Video," *The 19th International Conference on Pattern Recognition (ICPR 2008)*, pp. 1-4, Dec. 2008.
- [7] R. Mehran, A. Oyama, and M. Shah, "Abnormal Crowd Behavior Detection Using Social Force Model," *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR 2009)*, pp. 935-942, Jun. 2009.
- [8] Y. Shi, Y. Gao, and R. Wang, "Real-Time Abnormal Event Detection in Complicated Scenes," *The 20th International Conference on Pattern Recognition*, pp. 3653-3656, Oct. 2010.
- [9] S. Pathan, A. Al-Hamadi, and B. Michaelis, "Incorporating Social Entropy for Crowd Behavior Detection Using SVM," *ISVC 2010*, pp. 153-162, 2010.
- [10] T. Cao, X. Wu, J. Guo, S. Yu, and Y. Xu, "Abnormal Crowd Motion Analysis," *International Conference on Robotics and Biomimetics*, pp. 1709-1714, Dec. 2009.
- [11] X. Zhang, H. Liu, Y. Gao, and D. Hu, "Detecting Abnormal Events via Hierarchical Dirichlet Process," *PAKDD*, pp. 278-289, Apr. 2009.
- [12] V. Mahadevan, W. Li, V. Bhalodia, and N. Vasconcelos, "Anomaly Detection in Crowded Scenes," *Computer Vision and Pattern Recognition*, pp. 1975-1981, Jun. 2010.
- [13] L. Kratz and K. Nishino, "Anomaly Detection in Extremely Crowded Scenes Using Spatio-temporal Motion Pattern Models," *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR 2009)*, pp. 1446-1453, Jun. 2009.
- [14] J. Ramirez, J. Gorriz, J. Segura, and C. Puntonet, A. Rubio, "Speech/Non-speech Discrimination Combining Advanced Fea-

- ture Extraction and SVM Learning.” *INTERSPEECH 2006*, pp. 1662-1665, Sep. 2006.
- [15] 변주완, 김성민, 원치선, “H.264/AVC 압축 영역에서의 동영상 검색”, *전자공학회논문지-SP편*, 43(5), pp. 485-491, 2006년 9월.
- [16] D. Tax and R. Duin, “Uniform Object Generation for Optimizing One-Class Classifiers,” *Journal of Machine Learning Research*, vol. 2, pp. 155-173, 2001.
- [17] <http://mha.cs.umn.edu/Movies./Crowd-Activity-All.avi>.
- [18] <http://www.cvg.rdg.ac.uk/PETS2009>
- [19] <http://iphone.hhi.de/suehring/tml/>
- [20] J. Davis and M. Goadrich, “The Relationship between Precision-Recall and ROC Curves,” *In Proceedings of the 23rd International Conference on Machine Learning*, pp. 233-240, 2006.

〈 著 者 紹 介 〉



오 승 근 (Seunggeun Oh) 학생회원
 2008년: 고려대학교 컴퓨터정보학과 학사
 2010년: 고려대학교 컴퓨터정보학과 석사
 2010년 ~ 현재: 고려대학교 컴퓨터정보학과 박사과정
 <관심분야> 패턴인식, 데이터마이닝, 상황인지, 컴퓨터 기반의 미술치료



이 종 욱 (Jonguk Lee) 학생회원
 2003년: 고려대학교 전산학과 학사
 2005년: 고려대학교 전산학과 석사
 2005년 ~ 현재: 고려대학교 전산학과 박사과정
 <관심분야> 멀티미디어 마이닝, 얼굴 탐지 및 인식, 기계학습, 지능 데이터베이스



정 용 화 (Yongwha Chung) 종신회원
 1984년: 한양대학교 전자통신공학과 학사
 1986년: 한양대학교 전자통신공학과 석사
 1997년: 미국 Univ. of Southern California 전기·공학과(컴퓨터공학 전공) 박사
 1986년 ~ 2003년: 한국전자통신연구원 생체인식기술연구팀 팀장
 2003년 ~ 현재: 고려대학교 컴퓨터정보학과 교수
 <관심분야> 바이오인식, 정보보호, 바이오정보보호



박 대 희 (Daihee Park) 정회원
 1982년: 고려대학교 수학과 학사
 1984년: 고려대학교 수학과 석사
 1989년: 플로리다 주립대학 전산학과 석사
 1992년: 플로리다 주립대학 전산학과 박사
 1993년 ~ 현재: 고려대학교 컴퓨터정보학과 교수
 <관심분야> 지능 데이터베이스, 데이터마이닝, 인공지능, 퍼지이론