

논문 2017-54-3-9

배경모델링과 CNN을 이용한 실시간 피플 카운팅 알고리즘

(A Real-time People Counting Algorithm Using Background Modeling and CNN)

양 훈 준*, 장 혁**, 정 재 협*, 이 보 원***, 정 동 석***

(HunJun Yang, Hyeok Jang, JaeHyup Jeong, Bowon Lee[©], and DongSeok Jeong)

요 약

최근 IoT 및 딥러닝 관련 기술요소들이 영상보안감시시스템에서도 다양하게 응용되고 있다. 그 중 CCTV를 통해 촬영된 동영상에서 자동으로 특정 객체를 검출, 추적, 분류 하는 감시 기능이 점점 지능화되고 있다. 본 논문에서는 보급형 CPU만 사용하는 PC 환경에서도 실시간 처리가 가능한 알고리즘을 목표로 하였다. GMM(Gaussian Mixture Model)을 이용한 배경 모델링과 헝가리안 알고리즘, 그리고 칼만 필터를 조합한 추적 알고리즘은 전통적이며 복잡도가 비교적 적지만 검출 오류가 높다. 이를 보강하기 위해 대용량 데이터 학습에 적합한 딥러닝을 기술을 적용하였다. 특히 움직임이 있는 사람의 특징을 강조하기 위해 추적된 객체에 대해 SRGB-3 Layer CNN을 사용하였다. 성능 평가를 위해 기존의 HOG와 SVM을 이용한 시스템과 비교했을 때 Move-in은 7.6%, Move-out은 9.0%의 오류율 감소가 있었다.

Abstract

Recently, Internet of Things (IoT) and deep learning techniques have affected video surveillance systems in various ways. The surveillance features that perform detection, tracking, and classification of specific objects in Closed Circuit Television (CCTV) video are becoming more intelligent. This paper presents real-time algorithm that can run in a PC environment using only a low power CPU. Traditional tracking algorithms combine background modeling using the Gaussian Mixture Model (GMM), Hungarian algorithm, and a Kalman filter; they have relatively low complexity but high detection errors. To supplement this, deep learning technology was used, which can be trained from a large amounts of data. In particular, an SRGB(Sequential RGB)-3 Layer CNN was used on tracked objects to emphasize the features of moving people. Performance evaluation comparing the proposed algorithm with existing ones using HOG and SVM showed move-in and move-out error rate reductions by 7.6 % and 9.0 %, respectively.

Keywords : Background Modeling, Object tracking, Convolutional Neural Network, People Counting

* 학생회원, *** 정회원, 인하대학교 전자공학과 (Dept. of Electronic Engineering, Inha University, Korea)

**정회원, ETRI (Electronics and Telecommunications Research Institute, Korea)

© Corresponding Author (E-mail : bowon.lee@inha.ac.kr)

※ 이 논문은 2015년도 정부(미래창조과학부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임 (NRF-2015R1C1A1A01055914)

※ 이 논문은 2010년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임 (2010-0020163)

Received ; September 29, 2016 Revised ; February 9, 2017

Accepted ; February 28, 2017

I. 서 론

CCTV를 이용한 영상보안시스템은 범죄예방, 재난재해 방지 등 안전에 대한 요구가 높아지면서 여러 분야에 널리 확산되어 사용되고 있다. 기존 아날로그 CCTV에서 최근 네트워크 기반의 지능형 CCTV로 발전하고 있으며 많은 카메라를 통합 관리할 수 있도록 그림 1과 같이 통합 관제 시스템이 등장하였다. 통합관제 시스템은 크게 저장서버, 영상분석 서버, 이벤트 서버, DB 서버로 나눌 수 있다. 이러한 서버들은 클라이언트가 라

이브 영상, 분석 이벤트 등을 실시간으로 조회 할 수 있게 해주며, 과거의 영상을 리플레이 할 수 있게 한다. 최근 영상 분석은 GPU를 통해 보다 효율적으로 이루어지며 Hadoop^[1]과 같은 빅데이터 분산 처리 시스템을 통해 대용량 분석 처리가 가능하다.

지능형 CCTV 시스템에서 단순한 상황 감시만이 아닌 특정 목적을 위한 솔루션들이 개발되고 있다. 상황 감시, 피플 카운팅, 배회 감지, 화재 감지 등의 영상 데이터를 분석하고 처리하는 방법들은 많은 복잡도가 필요하다. 분석 솔루션 구현 형태를 보면 OpenVX를 이용한 임베디드 소프트웨어로 카메라 내부에 분석 기능을 넣거나 전용 영상 분석 서버를 이용하여 여러 카메라 영상을 동시에 분석하고 이벤트를 발생 시킨다. 그래서 이러한 영상 자동 분석 시스템을 사용할 때는 효과적이고 빠른 알고리즘의 적용이 요구된다.

본 논문에서는 외부 환경 변화를 고려한 객체 검출을 위해 Stauffer가 제안한 GMM (Gaussian Mixture Model)^[2]을 사용한다. GMM 모델링을 통해 검출된 객체를 추적하여 특정 위치를 지나는 사건이 발생할 경우 CNN (Convolutional Neural Network)를 이용하여 사람과 배경을 구분한다. 이를 위해 CCTV 영상에 대한 사람, 배경 DB를 구성하고 훈련하여 피플 카운팅을 위한 모델로 사용하였다.

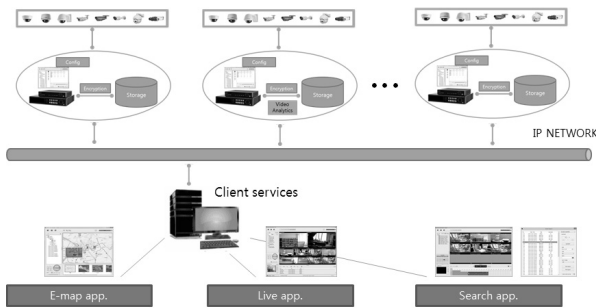


그림 1. CCTV 통합 관제 시스템
Fig. 1. CCTV Integrated control system.

사람을 배경에서 분리하는 연구가 많은 주목을 받고 있다. 이 연구가 어려운 이유는 사람과 배경이 모양과 크기가 고정된 객체가 아니라 그림 2와 같이 매우 다양한 변형을 가지고 있기 때문이다. 이를 분류하기 위해 현재 사용되는 방법들은 객체의 글로벌 템플릿에 Haar^[3], HOG^[4], HOG-LBP^[5]등의 사람이 직접 고안한 특징들을 추출한 후, SVM, boosting classifiers^[6]를 이용하여 훈련하는 방법이 있다. 또는 DPM^[7]을 이용하여 사람에 대한 로컬 템플릿을 구성하고 사람의 자세 변화를 감지

할 수도 있다. 하지만 이런 특징들은 변형이 크고 다양한 사람과 배경을 분류 위해 최적화 된 특징이 아니기 때문에 실제로 사용하기 어려운 문제를 갖는다.

최근 객체 분류를 위해 빅 데이터에 기반한 딥러닝을 사용하는 방법들이 주목받고 있다. Yann Lecun은 1990년대에 최초로 뉴럴 네트워크에 컨볼루션을 성공적으로 적용한 LeNet^[8]을 제안하였다. 이 후, Alex Krizhevsky 등은 컴퓨터 비전에서 컨볼루션 네트워크를 사용한 AlexNet^[9]을 제안하였다. ImageNet ILSVRC challenge 2012에 제출된 Alexnet은 그 대회에서 우승하면서 뛰어난 성능을 보였다. CNN 기반의 딥러닝 방법들은 비지도 사전 훈련과 드랍 아웃(drop-out) 같은 방법들이 개발되면서 레이어를 계속 쌓아나가 더 깊어지고 커지는 방향으로 발전되었다. Szegedy 등은 GoogLeNet^[10]을 발표했으며 ISLVRC 2014에서 우승을 차지했다. GoogLeNet은 인셉션 모듈(Inception Module)을 이용해서 네트워크의 파라미터 수를 상당히 감소시킬 수 있었다. Kaiming He 등이 발표한 Residual Net^[11]은 top 5 오류율에서 3.57%의 성능을 나타내며 ISLVRC 2015에서 우승하였다.

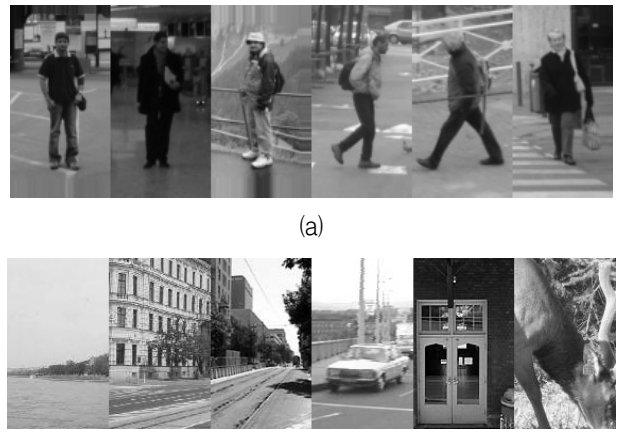


그림 2. 사람과 배경 영상의 비교 (a) 사람 영상, (b) 배경 영상
Fig. 2. Comparison of people image and background image (a) people images, (b)background images.

II. 본 론

그림 3과 같이 제안된 알고리즘에서는 GMM을 이용한 배경 모델링을 사용하여 움직이는 객체를 분리한다. 그 후 검출 된 객체를 대상으로 헝가리안 알고리즘(Hungarian Algorithm)^[12-13]과 칼만 필터(Kalman filter)^[14]를 사용하여 다중 객체를 추적한다. 그리고 추적된 객체의 궤적이 설정된 관심영역을 지나갈 때 SRGB-3 Layer CNN

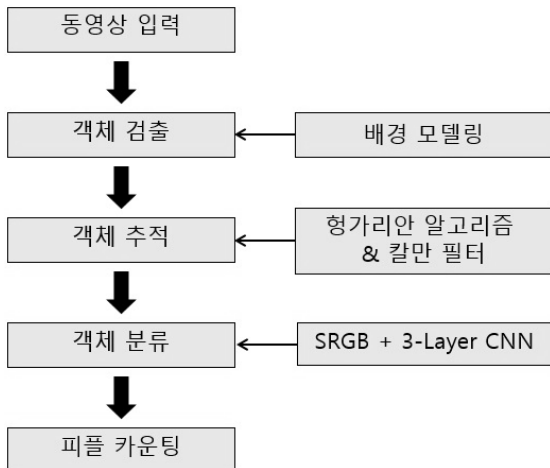


그림 3. 피플 카운팅 알고리즘 흐름도

Fig. 3. Algorithm flowchart of people counting.

을 사용하여 객체 분류를 수행하고 해당 범주가 사람이면 카운팅 한다.

1. 배경 모델링을 이용한 객체 검출

CCTV와 같이 고정된 카메라 입력에 대한 객체 검출을 위해서는 배경과 전경을 분리하는 방법이 사용되는데 배경 차분 방법과 GMM 방법이 있다. 본 논문에서는 외부 환경의 변화를 고려한 객체 검출을 위하여 Stauffer의 GMM 배경 모델링을 사용하였다. 만약 배경이 전체적으로 일정한 밝기를 가질 경우에는 배경을 하나의 가우시안 분포만으로도 표현 할 수 있다. 하지만 시간의 변화와 외부 환경에 따라 실제 배경의 밝기가 변할 수 있다. 이 같은 변화를 모델링하기 위해서는 하나가 아닌 다수의 가우시안 분포를 가지는 적응형 가우시안 분포를 사용한다. 모델링한 가우시안 분포에 포함되지 않는 화소 값이 존재하는 경우 이를 검출하게 된다.

배경의 화소들의 집합은 시간 t 에 대해서 $\{X_1, \dots, X_t\}$ 이다. k 개의 가우시안 분포가 있을 경우, 현재 화소값에 해당하는 확률 함수는 다음과 같다.

$$P(X_t) = \sum_{i=1}^K w_{i,t} \eta(X_t | \mu_{i,t}, \Sigma_{i,t}) \quad (1)$$

식 (1)에서 $w_{i,t}$, $\mu_{i,t}$, $\Sigma_{i,t}$ 는 각각 시간 t 일 때 i 번째 가우시안 모델의 가중치, 평균값, 공분산 행렬을 의미한다. η 는 확률 밀도 함수를 의미하며 식 (2)와 같이 표현할 수 있다.

$$\eta(X_t | \mu, \Sigma) = \frac{1}{(2\pi)^{\frac{1}{2}} |\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(X_t - \mu)^T \Sigma^{-1} (X_t - \mu)} \quad (2)$$

현재의 화소가 k 개의 가우시안 모델 중 어떤 분포에 가장 근접한지 확인한다. 만약 k 개 가우시안 모델의 분포 범위 내에서 현재 화소가 포함되지 않는다면, 가장 확률이 적은 분포는 현재의 화소를 평균으로 갖고 분산은 매우 큰 초기값을 갖는 가우시안 분포로 대체되며, 가중치는 식 (3)과 같이 500 프레임마다 한번씩 업데이트 된다. 여기서 α 는 learning rate을 나타내며, 0.002로 설정하였다. $M_{k,t}$ 는 현재의 화소가 속한 분포에서는 1, 나머지에서는 0의 값을 갖는다.

$$w_{i,t} = (1 - \alpha)w_{i,t-1} + \alpha(M_{k,t}) \quad (3)$$

공분산을 현재 화소의 분산으로 근사화하면, 평균과 분산의 업데이트 방법은 수식 (4)-(6)과 같이 표현된다.

$$\mu_t = (1 - \rho)\mu_{t-1} + \rho X_t \quad (4)$$

$$\sigma_t^2 = (1 - \rho)\sigma_{t-1}^2 + \rho(X_t - \mu_t)^T (X_t - \mu_t) \quad (5)$$

$$\rho = \alpha \eta(X_t | \mu_k, \sigma_k) \quad (6)$$

배경 모델링의 후처리 방법으로 미디언 필터와 모폴로지를 사용하여 노이즈 화소들을 제거한다.

2. 객체 추적



(a)



(b)

그림 4. GMM을 이용한 배경 모델링 (a) CCTV 영상, (b) CCTV 영상에 대한 배경 모델링 결과

Fig. 4. Background modeling using GMM (a) CCTV, (b) Background modeling result of CCTV.

객체 추적을 위해서 그림 4와 같이 검출 된 블롭(blob)의 바운드 박스(bound box)를 구하고 중앙점의 위치를 검출한다. 그리고 이전 프레임의 중앙점과 현재 프레임의 중앙점은 유클리디언 거리를 비용 함수로 하는 헝가리안 알고리즘을 통해 추적 ID를 부여하고 객체를 추적한다. 추적 과정 중에 다양한 문제로 추적이 방해 받을 수 있다. 예를 들어 가려짐(occlusion)이나 그림자로 인한 문제점, 사람이 이동하면서 생기는 모양의 변화 등이다. 이를 해결하기 위해 칼만 필터를 사용한다. 객체 추적은 각 프레임 단위로 칼만 추적 보정 및 예측을 수행하여 보다 정확하게 움직임을 감지한다.

다수의 객체를 추적하기 위해 헝가리안 알고리즘을 사용한다. 완전 탐색(brute force search) 방법이 $O(n!)$ 의 수행 시간을 갖는 반면에 헝가리안 알고리즘은 $O(n^3)$ 을 갖는다. 그림 5에서 T 는 이미 검출 된 객체들을 의미하며 각 객체는 t_i 로 표기한다. N 은 다음 프레임에서 새롭게 검출 된 객체들이며, 각 객체는 n_j 로 표기한다. 그림에서 검은 실선으로 연결된 검은 점과 파란 점은 매칭 된 t_i 와 n_j 이다. 그리고 회색 점과 하늘색 점은 매칭 되지 못한 객체들을 의미한다.

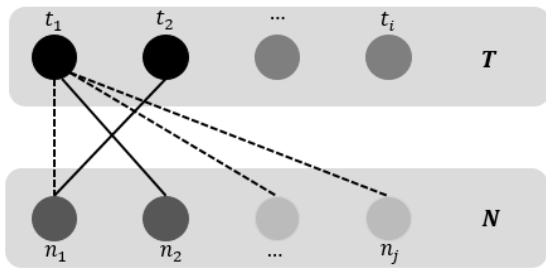


그림 5. 헝가리안 알고리즘의 메커니즘
Fig. 5. Mechanism of Hungarian algorithm.

헝가리안 알고리즘은 행렬 C 를 통해 수행된다. 행렬의 행은 i , 열은 j 라고 할 때, 행렬 C 의 요소 $c_{i,j}$ 는 식 (7)과 같다.

$$c_{i,j} = d_{i,j} \quad (7)$$

$d_{i,j}$ 는 T 와 N 의 유클리디언 거리이며 오류 비용(error cost)을 의미한다. 이 때 T 의 객체 개수와 N 의 객체 개수가 다를 수 있기 때문에 행렬을 식 (8)과 같이 정의한다.

$$C = \begin{cases} C_0 & i = j \\ [C_0 0_{i \times (j-i)}] & i < j \\ \begin{bmatrix} C_0 \\ 0_{(i-j) \times j} \end{bmatrix} & i > j \end{cases} \quad (8)$$

T 와 N 의 객체가 매칭 될 경우 할당 값은 1이 되고 식 (9)와 같이 행렬이 생성된다.

$$\sum_{i=1}^n \sum_{j=1}^n c_{ij} x_{ij} \quad (9)$$

where

$$x_{ij} = \begin{cases} 1 & \text{if } t_i \text{ is assigned } n_j \\ 0 & \text{otherwise} \end{cases} \quad (10)$$

추적 알고리즘을 통해 같은 추적 ID를 갖는 객체 바운드의 중앙점이 누적되어 궤적이 생성된다. 이동한 궤적을 사용하여 입구 영역과의 그레디언트(gradient)를 계산한다. 그리고 식 (11)을 이용하여 객체의 Move-in과 Move-out을 판단한다. 이동 궤적의 x 방향은 g_x 이고 y 방향은 g_y 이다.

$$\theta = \tan^{-1} \left[\frac{g_y}{g_x} \right] \quad (11)$$

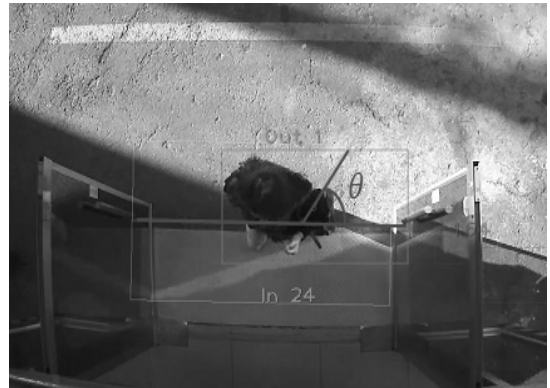


그림 6. 궤적의 평균 그레디언트와 출입문 라인과의 사이각을 이용한 Move-in, Move-out 판정

Fig. 6. Determining the Move-in and Move-out using the angle of the trajectory's mean gradient and the virtual door line.

3. SRGB-3 Layer CNN을 이용한 객체 분류

객체 추적을 이용하여 입구를 지나가는 객체가 존재한다고 판단되면 그 객체의 패치(patch)를 가져온다. 그리고 패치에 슬라이딩 윈도우(sliding window) 방법을 사용한다. 이 방법은 배경 모델링을 통해 얻은 블롭에 대해서 윈도우를 일정한 간격으로 이동시키면서 해당 윈도우 내에 우리가 원하는 객체가 있는지 확인하는 방법이다. 이 때, 카메라가 고정된 위치에 설치되어 있기 때문에 영상 안의 객체 크기의 변화가 적다. 그러므로 고정된 윈도우의 크기로 슬라이딩하여 불필요한 반복

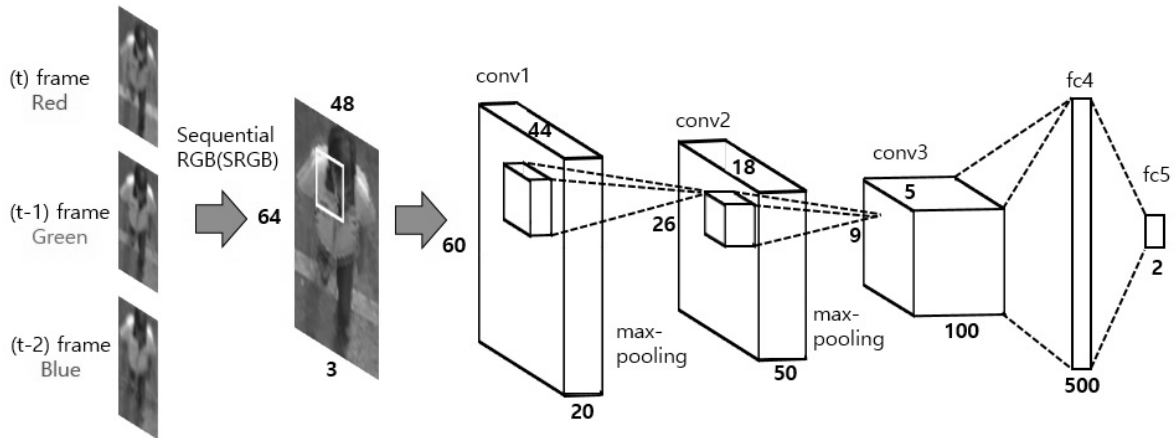


그림 7. SRGB-3 Layer CNN의 아키텍처
Fig. 7. Architecture of SRGB-3 Layer CNN.

검사를 제거하여 복잡도를 줄인다. 그리고 블록의 크기가 너무 작거나 큰 경우도 계산에서 제외한다.

사람과 배경을 분리하기 위한 분류 알고리즘으로 SRGB (Sequential RGB)와 3 Layer CNN을 사용한다. SRGB 영상은 (t) 프레임의 R채널, $(t-1)$ 프레임의 G채널, $(t-2)$ 의 B채널을 합쳐서 만들어진다. SRGB 영상은 움직이는 객체의 시간적인 특징을 얻기 위해 사용된다. 영상에서 배경은 움직임이 거의 없기 때문에 (t) , $(t-1)$, $(t-2)$ 프레임에서 RGB 성분의 변화가 거의 없으나 움직이는 객체의 경우 프레임마다 RGB 성분이 변한다. SRGB 영상의 크기는 $64 \times 48 \times 3$ 이다.

SRGB-3 Layer CNN의 각 레이어 출력은 식 (12)와 같이 표현할 수 있다. 레이어 l 의 노드를 $\{1, 2, \dots, j\}$ 라고 하고 layer $l-1$ 의 노드를 $\{1, 2, \dots, i\}$ 라고 할 때, 각 노드의 가중치 W 와 바이어스 b 에 대해서 노드의 출력은 식 (12)와 같은 형태를 갖는다.

$$O = \sigma \left(\sum_i x_i^l W_{i,j}^l + b_j^l \right) \quad (12)$$

where $\sigma(x) = \max(0, x)$.

레이어 $l+1$ 의 노드를 $\{1, 2, \dots, k\}$ 라고 할 때 O_k 와 실제 라벨 t_k 를 비교하여 식 (13)과 같이 비용 함수를 정의한다. 비용 함수를 이용하여 $\partial E / \partial W_{ik}$, $\partial E / \partial b_k$, $\partial E / \partial x_k$ 를 계산하고 최적의 파라미터를 찾기 위해 역전파(back-propagation)를 수행하여 파라미터를 업데이트한다.

$$E = \frac{1}{2} \sum_k (O_k - t_k)^2 \quad (13)$$

그림 7의 SRGB-3 Layer CNN의 아키텍처를 보면

SRGB를 생성한 후, 3개의 컨볼루션 레이어(convolution layer)와 2개의 완전 연결 레이어(fully connected layer)로 구성된다. 사용되는 커널은 5×5 크기의 컨볼루션 커널을 사용하고 max-pooling은 2×2 크기의 커널을 사용한다. 그리고 히든 레이어의 구성을 위해 conv1에서 conv3까지 컨볼루션과 max-pooling을 반복적으로 수행한다. 그 결과를 완전 연결 레이어에 연결하여 fc4, fc5로 구성한다. 모든 레이어의 활성화 함수(activation function)로 ReLU(Rectified Linear Unit)을 사용한다.

III. 실험

제안된 알고리즘을 적용하기 위해서 건물 입구 상단에 CCTV 감시 시스템을 설치하였다. 이 시스템은 입구에 출입하는 객체를 검출하고 실시간으로 이를 분석한다.

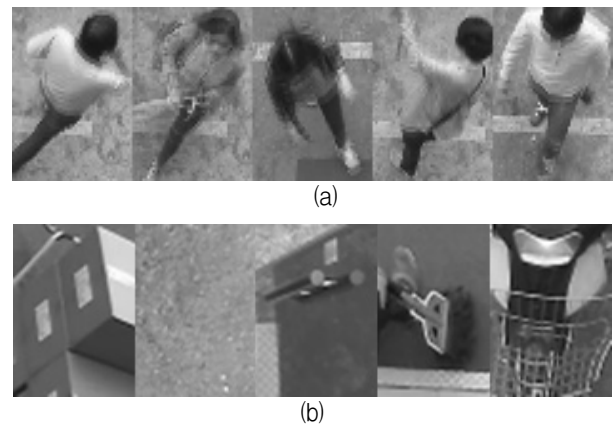


그림 8. 훈련 영상 (a) 사람 영상, (b) 배경 영상
Fig. 8. Training image (a) people image, (b) background image.

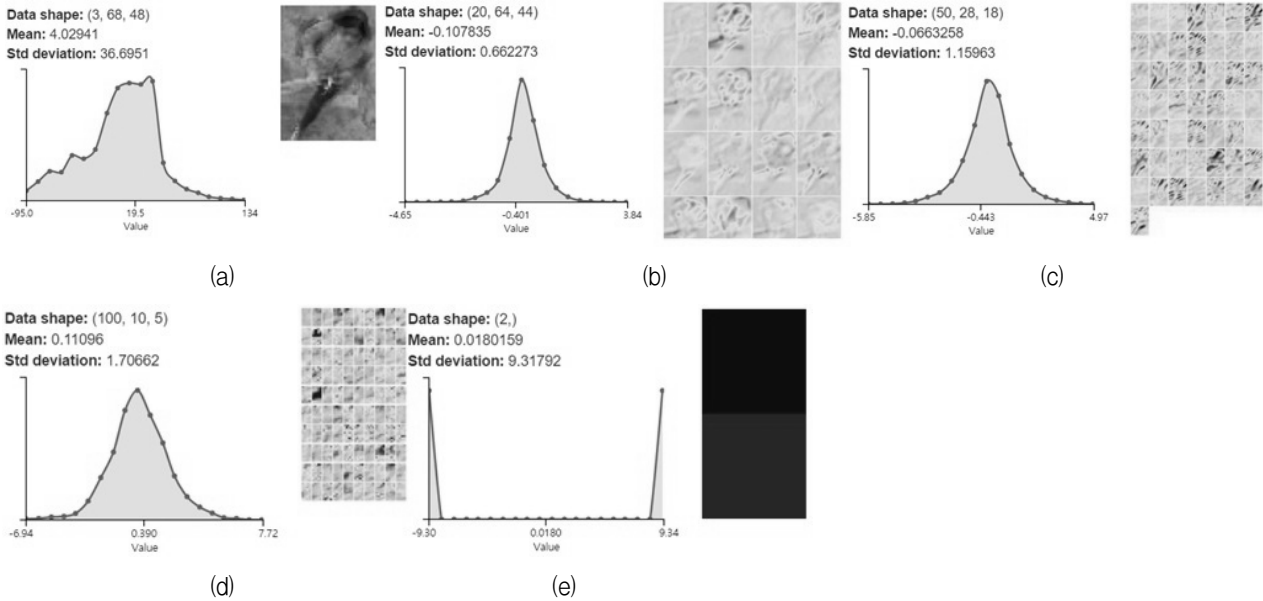


그림 9. SRGB 영상의 3 레이어 CNN 결과 (a) SRGB 영상의 분포, (b) 첫 번째 컨볼루션 레이어의 활성화 결과, (c) 두 번째 컨볼루션 레이어의 활성화 결과, (d) 세 번째 컨볼루션 레이어의 활성화 결과, (e) 완전 연결 레이어의 활성화 결과

Fig. 9. The 3-Layer CNN result of an SRGB image, (a) distribution of an SRGB image, (b) activation result of the first convolution layer, (c) activation result of the second convolution layer, (d) activation result of the third convolution layer, (e) activation result of the fully connected layer.

영상에서 사람과 배경을 구분하기 위해서 CCTV 감시 시스템을 통해 획득된 각 DB를 훈련하였다. 사람 영상은 사람의 전체 모습이 드러난 영상을 대상으로 훈련하였으며 배경 영상은 바닥, 박스, 문, 오토바이 등 특정한 객체가 아닌 다양한 크기와 모양의 변형이 존재하는 영상을 사용하였다. 학습 DB는 그림 8과 같이 $64 \times 48 \times 3$ 크기의 SRGB 영상을 사용했으며 실제 입구를 지나서 사람을 카운트하여 시스템의 성능 평가를 하였다.

그림 9에서 제안된 알고리즘에서 사용된 SRGB에 대한 CNN의 결과를 보여주고 있다. 이를 통해서 각 컨볼루션 레이어와 완전 연결 레이어의 데이터 분포를 알 수 있다.

성능 비교를 위하여 기존의 널리 사용되는 HOG를 이용하여 특징을 추출하고 SVM으로 분류하는 방법, LeNet을 이용한 분류 방법과 제안된 알고리즘을 비교하였다. 각 알고리즘은 사람이 가상의 선을 지나 건물 안으로 들어가는 Move-in과 건물에서 가상의 선을 지나 밖으로 나가는 Move-out으로 구분된다. 오류 카운트를 위해서 각 DB 영상에서 GT(Ground Truth)와 각 알고리즘을 통해 나온 결과를 비교한다.

피플 카운팅 시스템의 성능을 비교하기 위한 오류율 ER 은 식 (14)와 같이 정의한다.

$$ER = \frac{\sum_k |G_T - P_C|}{G_T} \quad (14)$$

G_T 는 k 개의 DB 영상의 GT를 의미하고 P_C 는 예측 값을 의미한다. 그림 10에서 보이듯이 SRGB-3 Layer CNN 방법이 SVM-HOG와 LeNet 보다 상대적으로 Move-in, Move-out 정확도가 높은 것을 알 수 있다.



그림 10. SVM_HOG, LeNet, SRGB-3 Layer CNN 오류율 결과 비교

Fig. 10. Comparison of SVM-HOG, LeNet, and an SRGB-3 Layer CNN.

그림 11과 12에서는 GT와 제안된 알고리즘을 사용하여 카운트한 사람의 누적값을 그래프로 표현했다.

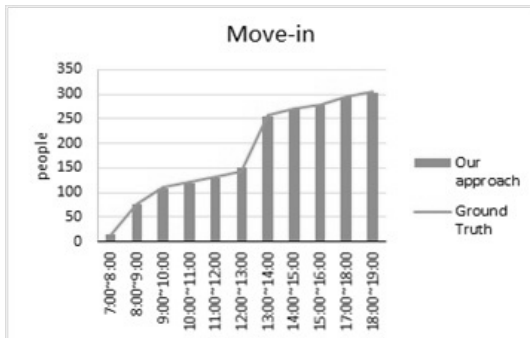


그림 11. GT와 제안된 알고리즘 결과의 Move-in 결과 비교

Fig. 11. Move-in count comparison of the ground truth and the proposed approach.

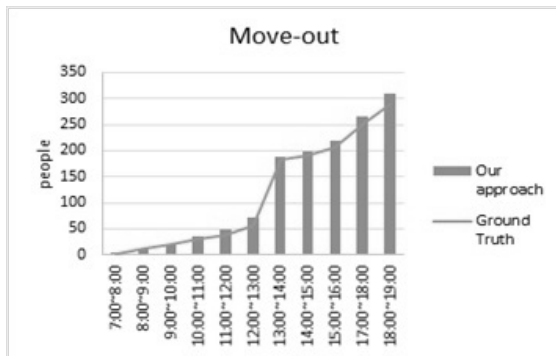


그림 12. GT와 제안된 알고리즘 결과의 Move-out 결과 비교

Fig. 12. Move-out count comparison of the ground truth and the proposed approach.

IV. 결 론

본 논문에서는 GMM을 이용한 배경 모델링을 통해 객체를 검출하고 헵가리안 알고리즘과 칼만 필터를 통해 객체를 추적한 후, SRGB-3 Layer CNN으로 사람배경을 구분하는 실시간 피플 카운팅 시스템을 제안하였다. 실시간 시스템에 적용하기 위해 복잡도를 줄일 수 있는 방법을 객체 검출과 추적에 사용하였다. 그림 10에서 확인할 수 있듯이 제안된 알고리즘을 통해서 Move-in의 경우는 SVM+HOG 보다 7.6%, LeNet보다 3.2%의 성능 향상이 있었다. 그리고 Move-out의 경우는 SVM+HOG 보다 9.0%, LeNet보다 0.9%의 성능 향상이 있었다.

오류 내용을 확인해보면 입구에서 배회하는 객체가 있을 경우 GT보다 더 많이 카운트 되는 경우가 있다. 이를 방지하기 위해서 배회 감지 알고리즘을 추가하여 오류율을 줄이는 방법이 연구되어야 할 것으로 보인다.

REFERENCES

- [1] K. Shvachko, H. Kuang, S. Radia, and R. Chansler, "The Hadoop Distributed File System," IEEE 26th Symposium on Mass Storage Systems and Technologies(MSST), pp. 1-10, 3-7 May 2010.
- [2] C. Stauffer and W. Grimson, "Learning Patterns of Activity Using Real-Time Tracking," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 22, pp. 747-757, 2000.
- [3] P. Viola, M.J. Jones, and D. Snow, "Detecting pedestrians using patterns of motion and appearance," IJCV, vol. 63, pp. 153-161, 2005.
- [4] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," In CVPR, pp. 886-893, 2005.
- [5] X. Wang, T.X. Han, S. Yan, "An HOG-LBP human detector with partial occlusion handling," ICCV, pp. 32-39, 2009.
- [6] P. Doll'ar, Z. Tu, P. Perona, and S. Belongie, "Integral channel features," In BMVC, pp. 1-11, 2009.
- [7] P.F. Felzenszwalb, R.B. Girshick, D. McAllester, and D. Ramanan, "Object detection with discriminatively trained part-based models," TPAMI, vol. 32, pp. 1627-1645, 2010.
- [8] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," Proceedings of the IEEE, vol. 86, pp. 2278-2324, 1998.
- [9] A. Krizhevsky, I. Sutskever, and G.E. Hinton, "Imagenet classification with deep convolutional neural networks," NIPS 2012, pp. 1097-1105, 2012.
- [10] C. Szegedy et al, "Going deeper with convolutions," CoRR, vol. abs/1409.4842, 2014.
- [11] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," arXiv preprint arXiv:1512.03385, 2015.
- [12] J. Munkres, "Algorithms for the Assignment and Transportation Problems," Journal of the Society for Industrial and Applied Mathematics, vol. 5, no. 1, pp. 32-38, 1957.
- [13] F. Lutteke, X. Zhang, and J. Franke, "Implementation of the Hungarian Method for object tracking on a camera monitored transportation system," ROBOTIK 2012, pp. 1-6, 2012.
- [14] R. Rad and M. Jamzad, "Real-time classification and tracking of multiple vehicles in highways," Pattern Recognition Letters, vol. 26, pp. 1597-1607, 2005.

저 자 소 개



양 훈 준(학생회원)
2011년 인하대학교 전자공학과 학사 졸업
2013년 인하대학교 전자공학과 석사 졸업
2013년~현재 인하대학교 전자공학과 박사 과정

<주관심분야: 패턴인식, 영상처리, 컴퓨터 비전, 딥러닝>



장 혁(정회원)
1997년 인하대학교 전자공학과 학사 졸업
1999년 인하대학교 전자공학과 석사 졸업
2000년~현재 인하대학교 전자공학과 박사 과정

2000년~2001년 메디슨 x-ray 사업부 연구원
2002년~2009년 인피니트헬스케어 연구소 팀장
2010년~현재 한국전자통신연구원 사업화부문 중소기업협력부 기업지원선임연구원
<주관심분야: 영상처리, 컴퓨터 비전, 의료영상 시스템, ITS(지능형교통시스템)>



정 재 협(학생회원)
2009년 인하대학교 전자공학과 학사 졸업
2011년 인하대학교 전자공학과 석사 졸업
2011년~현재 인하대학교 박사 과정

<관심분야: 영상신호처리, 영상 검색, 패턴인식, 컴퓨터 비전, 머신러닝>



이 보 원(정회원)-교신저자
2000년 서울대학교 전기공학부 학사
2003년 University of Illinois at Urbana-Champaign 석사
2006년 University of Illinois at Urbana-Champaign 박사

2007년~2014년 Hewlett-Packard Laboratories 선임연구원
2011년~현재 IEEE Senior Member(Signal Processing Society)
2014년~현재 인하대학교 전자공학과 조교수
<주관심분야: 통신, 컴퓨터, 신호처리, 반도체>



정 동 석(정회원)
1977년 서울대학교 전기공학과 학사 졸업
1985년 Virginia Tech 전자공학과 공학 석사
1988년 Virginia Tech 전자공학과 공학 박사

1988년~현재 인하대학교 전자공학부 교수
1990년~1994년 전자공학회 논문지 편집위원
1990년~1994년 통신학회 논문지 편집위원
2000년~2004년 정보전자공동연구소 소장
2010년~2011년 인하대학교 IT공대학장
2012년~2013년 인하공업전문대학 총장
<주관심분야: 영상처리, 컴퓨터 비전, 패턴인식, 내용기반 멀티미디어검색>