

딥 러닝과 마르코프 랜덤필드를 이용한 동영상 내 그림자 검출

이종택[†], 강현우[†], 임길택[†]

Moving Shadow Detection using Deep Learning and Markov Random Field

Jong Taek Lee[†], Hyunwoo Kang[†], Kil-Taek Lim[†]

ABSTRACT

We present a methodology to detect moving shadows in video sequences, which is considered as a challenging and critical problem in the most visual surveillance systems since 1980s. While most previous moving shadow detection methods used hand-crafted features such as chromaticity, physical properties, geometry, or combination thereof, our method can automatically learn features to classify whether image segments are shadow or foreground by using a deep learning architecture. Furthermore, applying Markov Random Field enables our system to refine our shadow detection results to improve its performance. Our algorithm is applied to five different challenging datasets of moving shadow detection, and its performance is comparable to that of state-of-the-art approaches.

Key words: Moving Shadow Detection, Convolutional Neural Network, Markov Random Field, Surveillance System, Object Detection

1. 서 론

최근 컴퓨터 비전 혹은 인공 지능 분야에서 딥 러닝(Deep Learning)에 관련된 연구가 활발하게 진행되고 있다. Convolutional neural network(CNN)이라고 불리는 딥 러닝은 사실 인공 신경망에 기반하여 설계된 개념으로 이미 1980년대부터 연구가 시작된 기계 학습 알고리즘의 하나이다. 1989년 Yann LeCun[1]이 오류 역전파(back propagation) 알고리즘을 사용하여 필기체 숫자 인식에 성공함으로써 알고리즘이 성공적으로 동작하는 것을 증명했으나, 느린 학습의 속도, 학습 데이터 세트에 과적용 되도록

학습이 되는 문제(over fitting) 등으로 인해 크게 각광받지 못했다.

이러한 딥 러닝이 2000년대에 들어 다시 각광을 받게 되고, 특히나 오늘 날에 이르러서는 인공 지능 분야의 대세로 부활하게 된 데에는 우선 몇 가지 중요한 연구(non-linear 한 convolution layer 사용, layer의 Deep Boltzmann Machine으로서의 해석, stochastic gradient descent의 사용 등)들이 진행되었다는 점과, 무엇보다 기존 신경망의 단점을 극복시킬 수 있는 방대한 양의 데이터(Big Data)와 이를 학습시킬 수 있는 컴퓨터 처리 성능의 향상이 있었다[2].

※ Corresponding Author: Jong Taek Lee, Address: (42994) 1 Techno Sunhwan-ro 10-gil, Yuga-meon, Dalseong-gun, Daegu, Korea, TEL: +82-53-670-8023, FAX: +82-53-670-8008, E-mail: jongtaeklee@etri.re.kr
Receipt date: Nov. 2, 2015, Approval date: Nov. 20, 2015

[†]Regional Industry IT Convergence Research Section, Daegu-Gyeongbuk Research Center, IT Convergence Technology Research Laboratory, Electronics and Telecommunications Research Institute (ETRI)
(E-mail: hwkang@etri.re.kr)
(E-mail: ktl@etri.re.kr)

딥 러닝이 각광받기 시작한 이후 다양한 분야에서 응용되기 시작하였으며, 특히 음성 인식(speech recognition) 분야에서 탁월한 성능을 보여주었으며, 컴퓨터 비전 분야에서도 굉장한 성공을 보였다. ImageNet challenge에서 백만 개 이상의 이미지 데이터를 학습에 사용하여 1000가지의 물체를 인식하는데 성공한 것은 대표적인 사례라고 할 수 있다[3].

딥 러닝의 주요한 특징 중 하나는 raw data에서 feature를 자동으로 학습한다는 데 있다. 기존의 기계 학습에서 전형적으로 사용되는 방법은 사람이 영상을 분석하여 잘 설계된 feature를 추출하고 이를 학습에 사용하였다. 그러나 딥 러닝에서는 여러 개의 convolutional layer를 사용하여 자동적으로 학습된 feature를 추출하게 된다. 어떤 문제는 사람이 좋은 feature를 설계하는 데 어려움이 있을 수 있으며, 이러한 경우 feature를 자동으로 학습하는 deep learning은 좋은 대안이 될 수 있다. 우리는 이러한 딥 러닝의 특성에 주목한다.

한편, 본 논문의 연구 주제인 그림자 검출(shadows detection) 문제는 컴퓨터 비전 분야에서 중요한 연구 주제 중의 하나이다. 그림자는 실내 환경 이든 실외 환경이든 언제나 존재하고 있으며, 광원의 종류에 따라 다양한 형태로 나타난다. 이러한 그림자에 의한 영향은 영상을 인식 혹은 이해하는데 있어서 도움이 되기보다는 방해 요소로 존재하는 경우가 더 많다 [4,5]. 따라서 그림자를 찾아내고 제거하는 것은 영상을 해석하는데 많은 도움을 줄 수 있다.

그림자 검출에 관한 연구는 이미 여러 편이 발표되었지만, 그림자 검출은 여전히 어려운 문제이다. 앞서 언급한 바와 같이, 그림자는 광원에 의해서 다양한 형태로 나타날 수 있기 때문이다. 우리는 이러한 문제점을 극복하기 위해서 영상에서 배경 영상을 분리하고, 입력 영상과 배경 영상에서 그림자 영역에 관한 feature를 CNN을 사용하여 학습하는 방법을 사용하였다[6]. 앞서 언급한 바와 같이 딥 러닝을 사용할 경우 데이터를 통해서 feature를 스스로 학습하므로, 그림자와 같이 다양한 형태로 나타날 수 있는 객체에서 특징을 찾아내는데 유리하다고 판단된다.

딥 러닝을 그림자 검출에 응용한 연구는 기존에 존재하지 않았던 것은 아니다. 그러나 본 연구에서처럼 동영상에서 그림자 검출에 응용한 사례는 아직까지 발견하지 못했다. 특히 실제 현장에 설치된 CCTV

영상에서 조명 및 환경에 대한 별도의 설정이나 조종이 없이 딥 러닝 학습만을 사용하여 그림자 검출에 성공한 것이 의미 있다고 할 수 있다.

본 연구에서는 딥 러닝을 이용하여 동영상에서 그림자를 검출하는데 보다 일반적이고 간편한 방법으로서 다른 최신 방법과 비견할만한 의미 있는 결과를 얻었으며, 추가적으로 마르코프 랜덤필드(Markov Random Field)를 사용하여 그림자 검출의 성능을 더욱 향상시킬 수 있는 방법을 제시한다.

본 논문의 구성은 다음과 같다. 2장에서 본 논문의 제안하는 방법에 대한 전체적인 흐름과 개요를 제공한다. 3장에서는 feature 학습과 그림자 영역 분류를 위한 딥 러닝 기반 네트워크를 서술하며, 4장에서 제안하는 마르코프 랜덤필드 기반 그림자 영역 분류 결과 정제 알고리즘을 설명한다. 그리고 5장에서는 제안한 알고리즘으로 실험한 결과를 기존의 알고리즘들과 공개된 데이터 셋을 이용하여 성능을 비교한다. 끝으로, 6장에서 결론을 맺으며 향후 연구 방향을 제시한다.

2. 시스템 요약

본 장에서는 동영상 내 그림자 검출을 위한 시스템을 Fig. 1의 플로우 차트와 함께 요약하여 설명한다. 그림자 검출 시스템을 구성하는데 있어 유의할

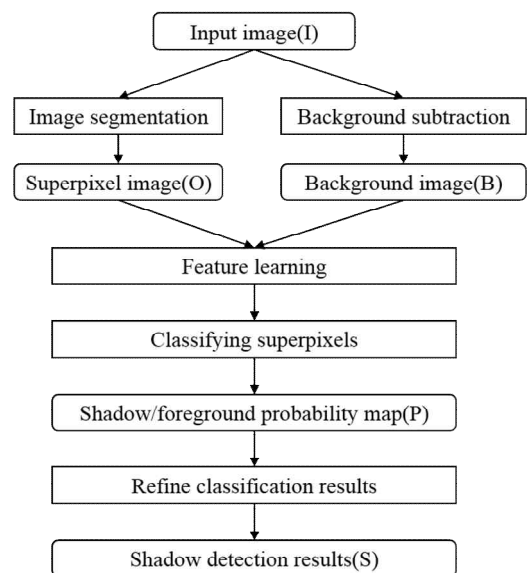


Fig. 1. Flowchart for moving shadow detection.

사항은 다음과 같다. 일반적으로 동영상 내에서 그림자 검출을 시도하는 경우 그림자 검출 자체에 목적이 있는 것이 아니라, 이를 관심 객체 검출 또는 추적의 성능을 향상 시키는 데 목적이 있는 경우가 많다. 따라서 그림자 검출이 일종의 전처리와 같이 여겨질 수 있으며, 그러므로 속도가 느려서는 실제 영상 보안 시스템에 활용되기가 어렵다. 또한 우리가 학습과 테스트를 위해 이용하는 데이터 셋의 경우 모든 프레임에 대한 입력 원영상은 주어지지만, 이 영상에 대한 그라운드 트루스(ground truth)는 훨씬 간헐적으로 주어지기에(10%이하) 연속적 분할 결과를 학습에 이용하기 어렵다는 점 역시 참고해야 한다. 반면, 촬영 된 동영상은 모두 고정식 카메라에서 얻어졌기 때문에 백그라운드 모델링이 용이하다.

입력 영상에 대해 처음 적용하는 프로세스는 영상 과분할이다. 여러 개의 비슷한 특징을 가지는 화소를 모은 슈퍼픽셀(superpixel)을 얻어낸 후, 주어진 영상 내에 있는 모든 화소에 대해 그림자(shadow)/전경(foreground) 구별을 하는 대신 슈퍼픽셀 영역 별로 그림자/전경 구별을 한다. 이와 같은 방법을 통해, 시스템의 처리 속도를 향상 시킬 수 있을 뿐 아니라, 같은 클래스를 가질 가능성이 높은 화소를 미리 묶어 이후 CNN 구조에서 이 영상에 대해 더 좋은 특징을 추출할 수 있도록 한다. 슈퍼픽셀을 구하는 방법들 중 Simple Linear Iterative Clustering(SLIC)[7]과 Superpixels Extracted via Energy-Driven Sampling(SEEDS)[8]가 가장 많이 이용되며, 본 연구에서는 시스템의 처리 속도가 중요하므로 더 빠른 방법인 SEEDS를 이용하여 슈퍼픽셀을 구한다. 슈퍼픽셀의 개수를 적절히 고르는 것이 또한 중요한데, 본 연구에서는 320×240 해상도의 영상에 대해 슈퍼픽셀의 개수를 60개로 정하였다.

백그라운드 모델링을 통해 배경으로 확실히 분별되는 슈퍼픽셀을 고를 수 있다. 슈퍼픽셀 내의 화소 중 배경으로 분별되는 화소 수가 90%이상인 경우 해당 슈퍼픽셀을 그림자/전경 분류 학습과 테스트에서 제외한다. 슈퍼픽셀은 다양한 모양과 크기를 가지지만, 우리가 가지고 있는 CNN 구조의 한계 상 같은 모양과 크기를 가지는 것이 좋다. 따라서 얻어낸 슈퍼픽셀을 둘러싸는 직사각형 (MBR: Minimum Bounding Rectangle) 영역의 영상을 받아 32×32의 정사각형 영상으로 크기를 제조정한다.

32×32 입력 영상과 마찬가지로 방법으로 얻을 수 있는 32×32 백그라운드 영상을 3장에서 소개하는 그림자/전경 분류를 위한 CNN 구조에 입력으로 넣는다. CNN 구조를 학습할 때에는 슈퍼픽셀 별 그림자/전경 분류 결과를 그라운드 트루스를 통해 알아내어 CNN 학습에서의 label로서 이용하고, CNN 구조가 학습이 끝난 후 테스트 할 때는 슈퍼픽셀 별로 그림자/전경 분류 확률분포 값을 얻을 수 있다. 해당 슈퍼픽셀의 확률분포 값을 모두 모아 확률분포 맵을 만든 후, 이 확률분포 맵을 4장에서 설명하는 것과 같이 마르코프 랜덤필드의 입력으로 넣어 분류 결과를 정제하여, 최종 그림자 검출 결과 영상을 얻는다.

3. Convolutional Neural Network (CNN)

본 연구에서 구성된 Deep CNN 구조는 Fig. 2에서 보는 것과 같이 32(가로) × 32(세로) × 6(채널)의 입력을 받아 3개의 클래스에 대한 확률 값을 출력으로 내보낸다. 이 구조는 총 7개의 층(Layer)으로 이루어져 있으며, 앞의 4개의 층은 2개의 convolutional 층과 2개의 subsampling 층으로 뒤의 3개의 층은 fully connected 층으로 구성되어 있다. 처음 convolutional 층에서는 32×32×6의 입력을 20개의 5×5×6 크기 kernel로 필터링하고, 두 번째 convolutional 층에서는 앞선 convolutional 층에서 얻어진 값에 max pooling과 정규화를 거친 결과(14×14×20)를 입력으로 넣어 60개의 5×5×20 크기 kernel로 필터링한다. 마찬가지로, 두 번째 convolutional 층에서 얻어진 값에 max pooling과 정규화를 거친 결과(5×5×60)를 2개의 동일한 개수의 뉴런을 가진 은닉층(500)을 fully connected 층으로 거쳐 최종 결과를 내보낸다.

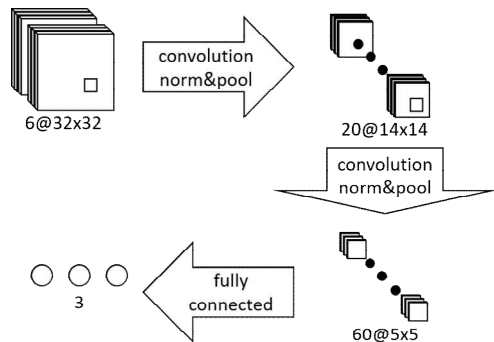


Fig. 2. An illustration of our CNN architecture.

최종 결과는 입력 영상에 대한 분류 결과로서 그림자, 전경, 그 외 세 가지에 대한 확률 분포값이다.

본 CNN 구조의 초기 입력은 32×32×3의 RGB 입력영상(I)과 32×32×3의 RGB 배경영상(B)을 중첩시켜 얻어지는 32×32×6의 크기를 가진다. 입력영상(I)만을 가지고 본 CNN 구조를 학습했을 때 결과는 배경영상을 중첩시켜 학습한 결과에 비해 성능이 크게 떨어졌다. 그림자와 전경 사이의 경계선을 학습하는 경우 배경영상을 사용하지 않더라도 충분한 feature를 얻어낼 수 있으나, 그림자 영역을 바로 학습하는 방법에서는 부분 입력영상만으로 그림자인지 전경인지 여부를 판별하는 것은 실용적이지 못하다.

입력 영상의 크기는 본 네트워크의 성능에 영향을 미치는 중요한 요인이다. 입력영상이 작은 경우 학습에 걸리는 시간이 크게 줄어 들지만, 분류 정확도가 떨어지게 된다. 반면, 입력영상이 큰 경우에는 분류 정확도가 향상될 수 있으나, 학습에 걸리는 시간이 크게 늘어날 수 있다. 입력영상을 작은 것부터 점점 크게 만들었을 때, 가로 32화소, 세로 32화소의 입력을 받을 때가 성능 향상이 가장 두드러졌고, 그보다 큰 크기의 입력의 경우 처리속도가 느려지는 것에 비해 성능향상은 미미했다.

본 네트워크를 standard stochastic gradient descent 알고리즘을 통해 학습하며, 초기 학습률은 0.0001, momentum은 0.9, weight decay는 0.0005로 할당하였다. 처음 두 개의 fully connected 층에 대해서는 더 강한 feature를 학습하기 위해 drop-out 층을 넣었으며, 주어진 데이터 셋이 그림자 영역이 전경 영역에 비해 비교적 작은 불균형한 상황이지만, 큰 문제는 되지 않아 특별히 클래스 내의 분포 균형을 맞추기 위한 작업은 하지 않는다. 본 연구에서는 딥 러닝 전문 라이브러리인 Caffe를 이용하여 그림자 검출 시스템을 구현했다[9].

4. 마르코프 랜덤필드 (MRF: Markov Random Field)

화소 단위로 클래스를 분류하는 영상 분할(segmentation) 문제는 그림자 화소의 최대사후(Maximum a Posteriori, MAP)를 추정하여 얻을 수 있다. 최대사후를 추정하기 위해 마르코프 랜덤필드 모델링을 통하여 에너지 함수를 정의하고, 이를 통해 앞선 장에서 얻어지는 그림자/전경에 대한 확률분포

맵을 가지고 분류 성능을 향상시키고자 한다. 에너지 함수는 두 개의 항을 가지는데 하나는 분류결과가 실제 영상과 불일치하는 경우에 대해 페널티를 부여하는 것과 다른 하나는 지역적으로 일관성을 가지도록 하는 것이다. 그리고 에너지 최소화(energy minimization) 방법으로 화소 별 최대사후를 추정하게 된다[10].

에너지 함수는 다음과 같이 정의한다.

$$E(I) = \sum_{p \in I} C(p) + \sum_{p \in I, q \in N} C(p, q) \tag{1}$$

$C(p)$ 는 unary 항목으로 화소 p 가 현재 할당된 클래스에 머무를 수 있기 위해 필요한 코스트이며, $C(p, q)$ 는 pairwise 항목으로 p 와 p 주변(4방향, N 으로 표현)의 q 화소간의 구조를 유지하기 위한 코스트이다. 입력으로는 영상 I 를 받는다. $C(p)$ 값을 구할 때는 3장에서 구한 확률 분포값을 이용하여 그림자/전경 분별이 불명확한 경우 그에 비례하여 높은 코스트를 가지도록 하고, $C(p, q)$ 값을 구할 때는 p 와 q 가 다른 클래스일 때 두 위치에서의 색상 값을 비교하여 그 차이가 작은 경우 그에 비례하여 높은 코스트를 가지도록 한다. 이를 수식으로 표현하면 다음과 같다.

$$C(p, q) = \beta e^{-\frac{|I_p - I_q|}{\alpha}}, \text{ if } class(p) \neq class(q) \tag{2}$$

α 는 색상 값 차이에 대한 민감도를 조정해주는 파라미터이며(클수록 둔감해짐), β 는 unary 항목과 pairwise 항목의 weight를 조정해주는 파라미터이다(클수록 pairwise의 영향력이 커짐). 상기 에너지 함수의 값을 가장 낮춰주는 최적화 문제를 푸는 방법은 여러 가지가 이용되어 왔으나, 본 연구에서는 graph cut 방식을 이용하여 에너지 최소화 문제를 푼다. 화소마다 주어진 클래스는 그림자/전경/배경 세 가지로, 그림자와 전경 사이의 클래스 변경을 위한 코스트는 낮게, 그 외의 경우는 코스트를 높게 설정한다. 그리하여 배경 화소에 대해서는 에너지 함수를 최적화하는 데 미치는 영향을 최소화한다.

5. 실험 결과 및 고찰

제안하는 시스템의 성능 평가를 위해 동영상 내 그림자 검출을 위해 많이 이용되고 있는 5개의 다른 환경에서 얻어진 공개된 데이터 셋[4, 11, 12]을 이용하였다. 이들 영상은 다른 환경에서 얻어진 만큼 각

기 다른 특징이 있는데, Campus[4] 영상의 경우 53개의 프레임에 대하여 그림자 그라운드 트루스가 제공되며 실외 아스팔트 위에서 촬영된 영상으로 노이즈가 심하고 사람과 차 모두 등장하며 그림자도 길게 나타난다. Hallway[11], Lab[4], Room[4] 영상의 경우 모두 실내 영상이며 노이즈가 심하지 않고 객체로는 사람이 주로 등장하며 그림자 그라운드 트루스가 제공되는 프레임 수도 각각 13, 14, 22개로 작은 편이다. CAVIAR[12] 영상의 경우, 앞선 영상들이 1개의 동영상으로 구성되어 있는 것과 달리 25개의 동영상으로 구성되어 있고 동영상 별 평균 44.5개의 프레임에 대하여 그림자 그라운드 트루스 정보가 제공된다. 또한 촬영 현장 바닥이 비교적 빛 반사가 심한 편이다. 제공된 영상의 해상도는 CAVIAR 영상의 경우 384×288 이고, 그 외의 영상은 모두 320×240의 해상도를 가진다.

상기 다양한 환경에서 얻어진 영상에 대하여 우리는 영상 별로 별도의 파라미터 값 조절이 없이 동일한 파라미터를 사용하여 그림자 검출 실험을 시도하였다. 우리가 제안한 방법의 정확성을 다른 최신 연구들과 비교하기 위해, 앞서 설명한 두 가지 평가방법을 이용한다. 이는 그림자 검출률(η)과 그림자 판별률(ζ)로 다음과 같이 정의된다.

$$\eta = \frac{TP_S}{TP_S + FN_S}, \zeta = \frac{TP_F}{TP_F + FN_F} \quad (3)$$

여기서 TP 는 true positive, FN 은 false negative 를 의미하며, 첨자 S 는 그림자, 첨자 F 는 전경 화소를 말한다. 그림자 검출률은 옳게 검출된 그림자 화소 수 대비 영상 내의 전체 그림자 화소수의 비율로 구해지며, 그림자 판별률은 옳게 검출된 전경 화소 수 대비 영상 내의 전체 전경 화소수의 비율로 구해진다. 임계치 값을 어떻게 조정하느냐에 따라 한 값은 올라가고 다른 한 값은 내려갈 수 있기 때문에 이

두 값을 합치거나 하나만 보기보다는 같이 보는 것이 실제 시스템 성능을 확인하기에 유리하다.

우리가 제안한 방법 중 딥 러닝만을 이용해서 얻어낸 결과와 마르코프 랜덤필드도 이용하여 얻어낸 결과, 그리고 다른 최근 방법들에 대한 그림자 검출률과 그림자 판별률을 Table 1에 기록하였다. DNM은 Prati 등[3]이 제안한 Deterministic Nonmodel-Based 방식이고, RMSDAP는 affinity propagation을 이용한 영역 기반 그림자 검출 방법[13]이다. DNN은 딥 러닝만을 이용한 방법이며[6], DNN+MRF는 우리가 제안하는 방법인 딥러닝과 마르코프 랜덤필드를 이용한 방법이다. 우리의 방법은 다른 방법과 비슷하거나 더 좋은 성능을 보이며, 특히 딥 러닝만을 이용한 방법 대비 약 5%의 오검출률을 줄일 수 있었다.

Fig. 3은 세 가지 환경에서 제안된 그림자 검출 방법이 활용된 결과 영상을 보여준다. 첫 열의 영상은 데이터 셋에서 주어진 입력 영상이며, 두 번째 열의 영상은 CNN결과로 나온 확률분포 맵 영상이다. 배경 영역에 대해서는 녹색으로 표시하였으며, 그 외의 영역은 그림자/전경 영역으로 밝기값이 클수록(밝을수록) 전경일 확률이 높은 것을 의미한다. 세 번째 열의 영상은 MRF결과로 나온 최종 그림자 검출 영상이다. 전경은 흰색, 그림자는 회색으로 표시했다. 마지막 열의 영상은 그라운드 트루스 영상이며 세 번째 열의 영상과 네 번째 열의 영상이 유사할수록 그림자를 정확하게 검출했다고 볼 수 있다. (정확한 성능 수치는 Table 1을 참조)

6. 결론 및 향후 연구

우리는 다양한 환경에서 동영상 내의 그림자 검출을 위해 딥 러닝을 사용한 그림자 영역 학습 방법과 마르코프 랜덤필드를 사용한 분류 결과의 정제 방법

Table 1. The shadow detection rates (η) and shadow discrimination rates (ζ) for proposed algorithms(DNN, DNN+MRF) and conventional algorithms(DNM, RMSDAP)

| Method | Campus η / ζ [%] | Hallway η / ζ [%] | Lab η / ζ [%] | Room η / ζ [%] | CAVIAR η / ζ [%] |
|---------|------------------------------|-------------------------------|---------------------------|----------------------------|------------------------------|
| DNM | 82.9 / 86.7 | - / - | 76.3 / 89.9 | 78.6 / 90.3 | - / - |
| RMSDAP | 33.8 / 91.0 | 88.3 / 96.4 | - / - | - / - | 89.7 / 96.4 |
| DNN | 74.3 / 84.9 | 89.6 / 92.2 | 86.2 / 84.3 | 80.3 / 93.7 | 84.0 / 93.7 |
| DNN+MRF | 74.1 / 85.8 | 90.3 / 91.6 | 87.3 / 84.3 | 82.4 / 93.8 | 84.7 / 93.9 |

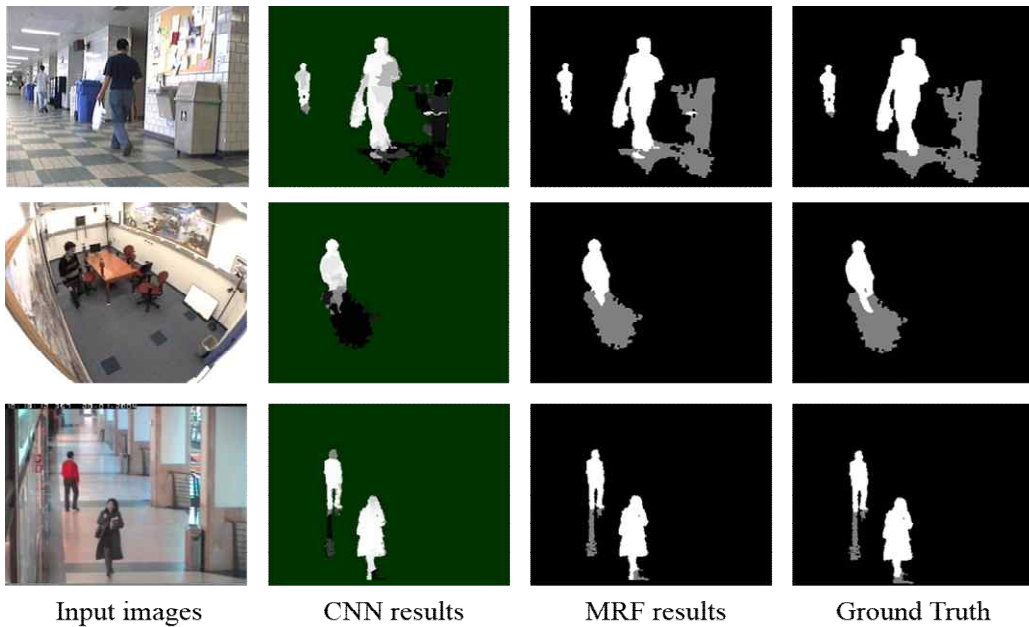


Fig. 1. Qualitative shadow detection results of the proposed method.

을 제안하였다. 제안된 방법은 간단하고 범용적이기 때문에, 확장이 용이하여 다른 영역 분류 문제에 적용시킬 수 있다. 딥 러닝과 배경영상을 이용하여 그림자/전경 영역의 특징을 학습하고 또한 마르코프 랜덤필드를 통해 얻어진 영역 구별 결과를 더 향상시킬 수 있었다.

반면, 본 연구에서는 동영상이 시간에 대해 연속적이라는 좋은 특성을 배경영상을 이용하는 것 외에는 활용하지 못하고 있다. 마르코프 랜덤필드를 이용하여 지역적인 연관성을 고려할 수 있었으나, 시간적인 연관성에 대해서는 아직 연구하지 못한 상황이다. 공개된 대부분의 데이터 셋에서 원영상은 모두 제공되지만, 그림자 검출 그라운드 트루스는 훨씬 제한적으로 제공되기 때 시간적 연관성을 이용한 학습 자료를 만드는 데 어려움이 있다. 따라서 그라운드 트루스가 제공된 학습 자료는 부족할 수밖에 없다는 것을 인지하고, 현 supervised 학습 방식이 아닌 unsupervised 학습 방식이나 semi-supervised 학습방식을 활용한 문제 해결 방법을 생각해 볼 필요가 있다.

REFERENCE

[1] Y. LeCun, B. Boser, J.S. Denker, D. Hender-son, R.E. Howard, W. Hubbard, and L.D.

Jackel, "Backpropagation Applied to Hand-written Zip Code Recognition," *Neural Computation*, Vol. 1, No. 4, pp. 541-551, 1989.
 [2] M.S. Ryoo, "How the Computer Vision Researchers survive in an era of Deep Learning," *Journal of the Institute of Electronics and Information Engineers*, Vol. 42, No. 5, pp. 29-33, 2015.
 [3] A. Krizhevsky, I. Sutskever, and G.E. Hinton, "Imagenet Classification with Deep Convolutional Neural Networks," *Proceedings of Advances in Neural Information Processing Systems*, pp. 1097-1105, 2012.
 [4] A. Prati, I. Mikic, M.M. Trivedi, and R. Cucchiara, "Detecting Moving Shadows: Algorithms and Evaluation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 25, No. 7, pp. 918-923, 2003.
 [5] S. Yeon and J. Kim, "Robust Illumination Change Detection Using Image Intensity and Texture," *Journal of Korea Multimedia Society*, Vol. 16, No. 2, pp. 169-179, 2013
 [5] J.T. Lee, K.T. Lim, and Y. Chung, "Moving Shadow Detection from Background Image

and Deep Learning,” *Proceedings of PSIVT 2015 Workshop on Video Surveillance*, 2015.

[6] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, “SLIC Superpixels Compared to State-of-the-Art Superpixel Methods,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 34, No. 11, pp. 2274–2282, 2012.

[7] M.V. den Bergh, X. Boix, G. Roig, B. de Capitani, and L.V. Gool, “SEEDS: Superpixels Extracted via Energy-Driven Sampling,” *Proceeding of European Conference on Computer Vision*, pp. 13–26, 2012.

[8] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, et al., “Caffe: Convolutional Architecture for Fast Feature Embedding,” *Proceeding of the ACM International Conference on Multimedia*, pp. 675–678, 2014.

[9] R. Qin, S. Liao, Z. Lei, and S.Z. Li, “Moving Cast Shadow Removal based on Local Descriptors,” *Proceeding of 20th International Conference on Pattern Recognition*, pp. 1377–1380, 2010.

[10] N. Martel-Brisson and A. Zaccarin, “Kernel-based Learning of Cast Shadows from a Physical Model of Light Sources and Surfaces for Low-level Segmentation,” *Proceeding of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8, 2008.

[11] CAVIAR: Context Aware Vision using Image-based Active Recognition, (2005) <http://homepages.inf.ed.ac.uk/rbf/CAVIAR> (accessed Oct. 1, 2015).

[12] J. Dai and D. Han, “Region-based Moving Shadow Detection using Affinity Propagation,” *International Journal of Signal Processing, Image Processing and Pattern Recognition*, Vol. 8, No. 3, pp. 65–74, 2015.



이 종 택

2005년 한국과학기술원(KAIST) 전기및전자공학과 공학사
 2007년 텍사스주립대(University of Texas at Austin) 전자 컴퓨터공학 공학석사
 2012년 텍사스주립대(University of Texas at Austin) 전자 컴퓨터공학 공학박사

현재 한국전자통신연구원 대경권연구센터 연구원.
 관심분야 : 컴퓨터비전, 행동인식, 패턴인식



강 현 우

2005년 경북대 컴퓨터공학과 공학사
 2009년 경북대 전자전기컴퓨터학부 공학석사
 현재 한국전자통신연구원 대경권 연구센터 선임연구원.

관심분야 : 컴퓨터통신, 임베디드 소프트웨어, 컴퓨터 비전



임 길 택

1993년 경북대학교 전자공학 공학사
 1995년 경북대학교 전자공학 공학석사
 1999년 경북대학교 전자공학 공학박사

현재 한국전자통신연구원 대경권연구센터 책임연구원
 관심분야 : 컴퓨터비전, 패턴인식, 비디오보안