

일반논문 (Regular Paper)

방송공학회논문지 제25권 제3호, 2020년 5월 (JBE Vol. 25, No. 3, May 2020)

<https://doi.org/10.5909/JBE.2020.25.3.362>

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

적응적 쌍선형 보간 이미지 피라미드를 이용한 DPM 기반 고속 객체 인식 기법

한 규 동^{a)}, 김 응 태^{a)†}

Fast Object Detection with DPM using Adaptive Bilinear Interpolated Image Pyramid

Gyu-Dong Han^{a)} and Eung-Tae Kim^{a)†}

요 약

최근 자율 주행 자동차와 지능형 CCTV에 대한 관심이 높아지면서 효율적인 객체 검출의 중요성은 필수적인 요소이다. 본 논문의 기반이 되는 DPM(Deformable Part Models)은 객체에 대한 변형 가능한 부분의 혼합을 사용하여 가변적인 객체를 나타낼 수 있는 대표적인 검출기로 다양한 분야에서 많이 연구 되고 있다. 객체 모델의 파트 모양과 구성을 잡아내는 기법으로 높은 검출 성능을 보여주지만 복잡한 알고리즘으로 인해 실제 어플리케이션에서 사용하기에는 한계가 있다. 이를 개선하기 위해 본 논문에서는 DPM에서 많은 연산을 필요로 하는 이미지 특징 피라미드(feature pyramid)를 구성하는 과정 대신, 특정 스케일에서 구해진 소수의 특징(feature) 맵에 적응적인 쌍선형(bilinear) 보간법을 이용하여 이미지 특징 피라미드를 재구성해 연산 속도를 줄이는 방법을 제안한다. 모의실험 결과, 제안된 방식의 DPM은 기존 DPM 방식 대비 검출 성능은 2.82%가 낮아졌지만 평균 연산 시간 10%를 향상시킴을 알 수 있었다.

Abstract

Recently, as autonomous vehicles and intelligent CCTV are growing more interest, the efficient object detection is essential technique. The DPM(Deformable Part Models) which is basis of this paper have used a typical object system that represents highly variable objects using mixtures of deformable part for object. Although it shows high detection performance by capturing part shape and configuration of object model, but it is limited to use in real application due to the complicated algorithm. In this paper, instead of image feature pyramid that takes up a large amount of computation in one part of the detector, we propose a method to reduce the computation speed by reconstructing a new image feature pyramid that uses adaptive bilinear interpolation of feature maps obtained on a specific image scale. As a result, the detection performance for object was lowered a little by 2.82%, however, the proposed detection method improved the speed performance by 10% in comparison with original DPM.

Keyword : object detection, deformable part models, image pyramid, bilinear interpolation

I. 서론

최근 지능형 차량 시스템인 자율 주행 자동차와 범죄나 재난 상황을 인식하는 지능형 CCTV에 대한 관심이 높아지면서 효율적인 객체 검출은 필수적인 요소이다. 객체 검출 분야에서 방법론적으로 중요한 의미를 가지는 방법들이 몇 가지 있는데, 고전적인 방법으로는 웨이블릿의 한 종류인 Haar feature와 유사한 매우 단순한 특징들을 조합하여 물체를 찾아내는 P. Viola et al.의 방식^[1]이 있다. 처음에는 단순한 검출기를 적용하고 점진적으로 복잡한 검출기를 적용하는 방법이다. 속도는 빠르지만 검출 성능을 높이기 위해서는 많은 샘플 데이터와 오랜 학습과정을 필요로 한다. 그리고 어떤 데이터를 사용하고 어떻게 학습 시키느냐에 따라서 성능에 많은 차이를 야기한다. 마찬가지로 속도 개선에 중점을 맞춘 다른 방법으로는 이미지 에지의 방향에 중점을 두어 물체 영역을 고정된 크기(16×16 픽셀)로 분할한 후 각 분할 영역마다 지역적으로 HOG를 계산하여 객체 모델을 만드는 N. Dalal et al. 방식^[2]이 있다. 또한 이 방식에 cascade 기법을 적용한 S. Avidan et al.의 방식^[3]은 다양한 크기와 수많은 위치의 블록들로부터 HOG를 계산하고 유의미한 블록들을 부스팅(boosting) 학습 과정을 통해 선별한다. 그리고 단일 채널의 영상 특징을 이용하여 객체를 검출하는 대신 객체 검출 대상의 다양한 채널(RGB, gray, edge, magnitude, histogram)을 계산하고 이 채널들을 종합적으로 이용해서 객체를 검출하는 P. Dollar et al.의 방식^[4]이 있다.

이러한 방식들의 객체 검출은 조명이나 복잡한 배경 등의 외적 요인이나 객체의 외형, 색상 등 내적 요인에 의해 검출 성능이 크게 영향을 받기 때문에 객체를 정확성 있게 검출하는데 있어서 상대적으로 어려움이 있다. 특히 단일 물체 모델을 사용하기 때문에 객체의 전체적인 모습이 나와야만 검출할 수 있고 보행자라는 객체가 앉아 있거나 누

워있는 경우 검출 성능이 크게 낮아진다. 이에 객체 검출에 있어서 객체의 모양이 가지는 고유의 기하학적인 특징을 이용하여 높은 정확성을 가지는 방법으로 Felzenszwalb et al.가 제안한 DPM(Deformable Part Models) 방식^[5]은 크게 입력 영상의 이미지 특징 피라미드, 필터 응답성, 검출 판단으로 나뉜다. 이미지 특징 피라미드는 입력 영상을 여러 스케일로 만들어 필요한 분석 작업을 통해 HOG 특징 맵을 만든다. 객체마다 학습된 모델 필터를 모든 특징 맵에 슬라이딩 윈도우 방식으로 처리하여 필터의 응답성을 구하고 이를 종합해서 객체 검출을 판단한다. 객체 모델의 파트 모양과 구성을 잡아내는 강력한 표현 기법을 사용하여 높은 정확성으로 영상 검출 방법들 중 뛰어난 성능을 보여주지만 복잡한 알고리즘으로 인해 높은 연산을 가지고 있다.

따라서 본 논문에서는 DPM의 많은 연산 문제를 해결하는 방법으로 이미지 특징 피라미드를 구성하는 과정에서 모든 스케일에 대한 특징 맵 처리 대신 특정 스케일에서 구해진 소수의 특징 맵에 보간법을 적용하여 새로운 이미지 특징 피라미드를 재구성하여 연산 속도를 줄이는 방법을 제안한다. 또한 소수의 특징 맵만으로 전체적인 이미지 특징 피라미드를 구성하는데 있어서 생기는 검출 성능 문제를 특정 조건에 따라 균일 영역과 에지 영역으로 나누어 에지 영역에 대한 특정 처리로 최대한 보상한다.

2장은 기존 DPM 방식에 대해서 간략히 설명하고, 3장에서는 제안하는 고속 검출 구조 및 알고리즘을 설명한다. 그리고 4장에서 모의실험을 통하여 성능의 비교 및 분석 결과를 보인다. 마지막으로 5장에서 본 논문의 결론을 맺는다.

II. Deformable Part Models(DPM)

DPM의 파트 기반 모델^[5]은 객체가 가지고 있는 각각의 파트들이 유기적으로 연결되어 하나의 모델을 이룬다고 가정한다. 이러한 모델의 장점은 여러 파트들 사이의 연관성을 고려하기 때문에 좀 더 강인하게 객체를 검출할 수 있다. 보행자를 예를 들면, 보행자의 경우 팔다리의 위치가 항상 고정되어 있지 않아서 어느 정도의 변화가 생기게

a) 한국산업기술대학교 전자공학부(Department of Electronics Engineering, Korea Polytechnic University)

✉ Corresponding Author : 김응태(Eung-Tae Kim)
E-mail: etkim@kpu.ac.kr
Tel: +82-31-8041-0488

ORCID: <https://orcid.org/0000-0001-5984-0045>

· Manuscript received January 16, 2019; Revised March 10, 2020;
Accepted March 10, 2020.

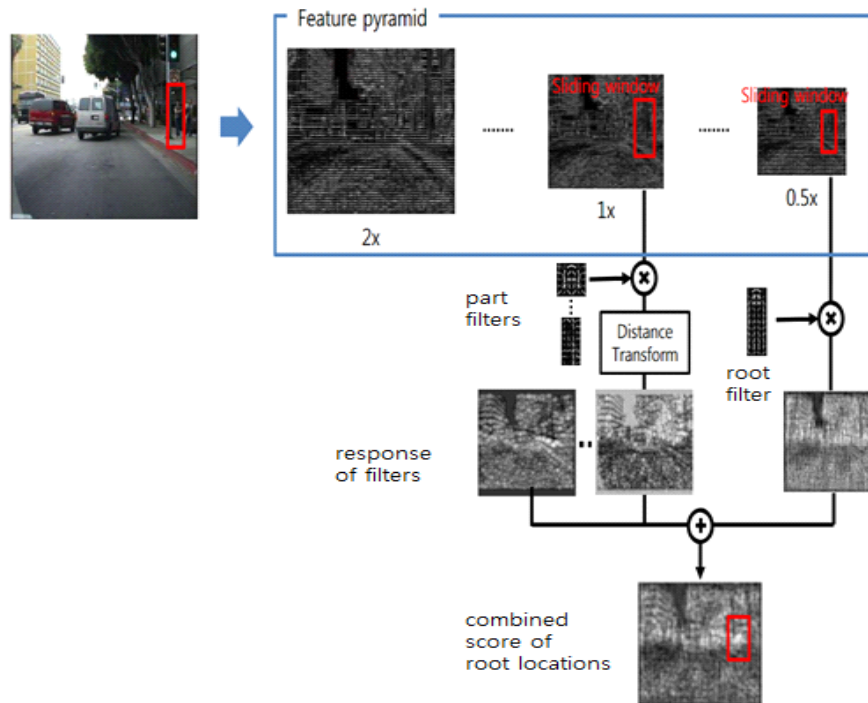


그림 1. DPM 알고리즘
Fig. 1. DPM Algorithm

되는데 단일 모델로는 이러한 변화를 모두 고려하기 힘들지만 파트 기반 모델로 접근하게 된다면 이러한 변화를 고려하여 더 강인한 성능을 보인다.

그림 1은 DPM의 전체적인 흐름도를 보여준다.

그림에서 보는 바와 같이 우선 입력영상의 HOG(Histogram of Oriented Gradient) 특징 피라미드를 구성한다. 일반적인 영상 피라미드의 각 레벨에서 HOG 특징을 계산한다. 피라미드 탑 레벨(0.5x)의 특징들은 입력영상의 큰 영역들에 대한 coarse gradients를 얻게되고, 반면에 피라미드 바텀 레벨(2x)에서는 작은 영역에 대한 fine gradient를 특징 추출하게 된다.

다양한 스케일의 HOG 피라미드 영상들을 통해 각 레벨들의 영상내 일정크기의 후보 영역을 설정하고 슬라이딩 윈도우 방식으로 한 픽셀씩 옆으로 움직이면서 영역내 객체를 인식하기 위한 선형 서포트 벡터 머신을 수행하게 된다. 이때 미리 학습된 객체들의 모델들인 루트 필터와 파트 필터들을 이용하여 슬라이딩 윈도우내 영역과 컨볼루션을 수행한다. 물체 전체의 모양정보를 담는 낮은 해상도의 루

트 필터(root filter)와 객체의 자세한 부분 모양 정보를 담는 더 높은 해상도의 파트 필터(part filter)들로 구성되는 스타 모델(Star Model)을 사용한다. 학습된 필터들과의 컨볼루션 값이 높을수록 물체의 모양과 일치하게 되는 특징을 이용하게 된다. 그리고 학습에 사용되는 필터들은 얻고자 하는 객체에 따라 다양한 형태의 필터들로 구성되어지는데, 예를 들어 사람의 모델을 만드는 것을 고려해 보면 루트 필터는 사람 경계와 같은 거친 해상도 가장자리를 캡처할 수 있고, 파트 필터들은 목, 팔, 다리, 몸통과 같은 디테일을 캡처할 수 있다. 또한 VOC2010에서 제공하는 데이터 셋의 학습된 승용차의 경우 6개의 뿌리 필터(root filter)와 48개의 파트 필터(part filter)들로 총 54개의 필터들을 사용하게 된다.

이때 영상피라미드 해상도의 차이로 인해 파트 필터는 피라미드 아래 여러 레벨들에 위치하게 되며, 이로 인해 그 레벨의 HOG 셀들은 루트 필터 레벨의 절반 크기의 셀들을 가지게 된다. HOG 피라미드내 한 모델의 배치는 루트필터의 위치와 n개의 파트필터들의 위치로 정의되

어진다. DPM에서는 다이내믹 프로그래밍과 거리변환 기법(distance transform technique)^{[8][9]}을 사용하여 루트 위치의 함수로서 모델의 파트에 가장 적합한 위치를 계산하게 된다. 학습된 모델에서 각각의 파트들은 루트 필터와의 상대적인 위치로 정의된 기준 위치(Anchor Position)를 가지며, 각 파트는 이 기준 위치로부터 멀어질수록 벌칙을 받는다. 즉, 이진 분류기에 의해 학습된 필터들에 의한 점수들의 합에서 파트들의 위치 변형에 의한 벌칙을 뺀 점수가 된다. 전체 점수를 계산하고 이 값이 임계값 이상일 경우 객체라고 판단하게 된다.

학습된 모델을 이용하여 다양한 스케일에서 객체 검출 과정은 1) 입력 이미지의 HOG 특징 피라미드를 계산하고 2) 이에 대해 모든 필터는 컨볼루션을 수행하여 3) 파트 모델에 대한 거리 변환을 통해 변형에 대한 벌칙을 추가 한 후 4) 전체 점수를 더해줌으로써 슬라이딩 윈도우와 이진

분류를 효과적으로 수행할 수 있다. 성능이 좋은 반면에 복잡한 알고리즘으로 연산이 많아 실시간 적용하기 위해서는 계산량을 줄여야하는 문제점이 생긴다.

III. 제안된 고속 검출 및 알고리즘

이에 본 논문은 생성된 이미지 특징 피라미드에 많은 필터들이 슬라이딩 윈도우 방식으로 계산하는 방식으로 인해 생겨나는 연산을 줄임과 동시에 객체에 대한 정확성을 크게 손실하지 않는 새로운 구조의 이미지 특징 피라미드를 제안한다. 그리고 이미지 특징 피라미드를 구성하는데 있어서 적응적 쌍선형 보간 알고리즘을 제안한다.

기존 방식의 경우 이미지 피라미드 계산은 P. Burt et al.^[10]에 따라 식 1을 적용한다. 스케일 s 에서의 이미지 I 를

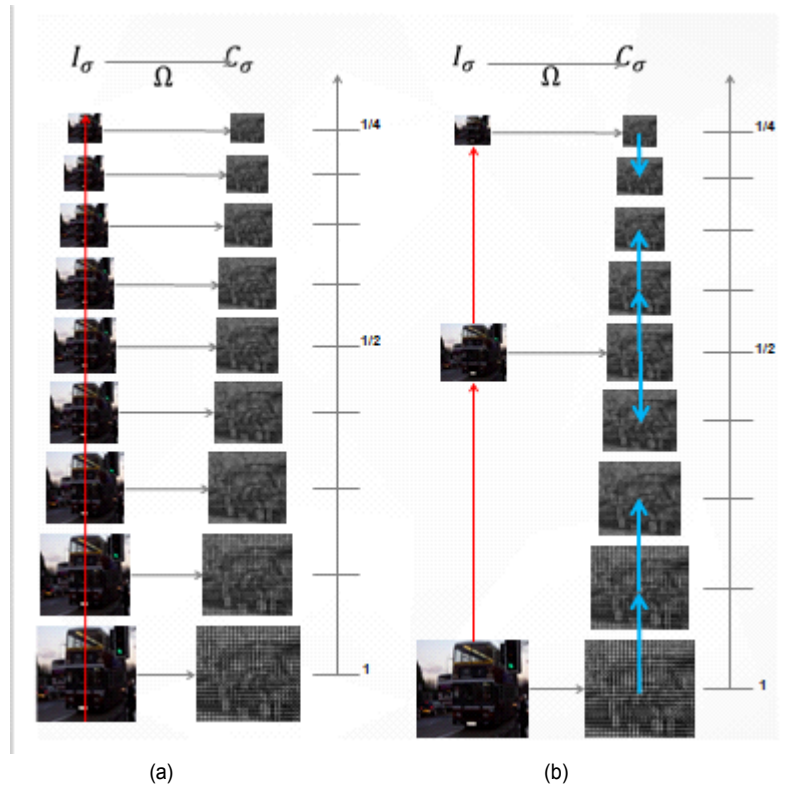


그림 2. (a) 기존 이미지 특징 피라미드 구조, (b) 제안하는 이미지 특징 피라미드 구조
Fig. 2. (a) Original Image Feature Pyramid, (b) Proposed Image Feature Pyramid

I_s 라 정의하고 s 에 의해 재구성된 I 를 $R(I, s)$ 라 한다. 즉, 이미지 스케일 s 에 의해 크기가 조정된 이미지 I_s 를 HOG 특징인 Ω 에 적용하면 C_s 로 특징 맵을 출력한다.

$$C_s = \Omega(I_s), I_s = R(I, s) \quad (1)$$

그림 2의 (a)는 스케일 s 에 따른 이미지들의 전체적인 특징 맵 피라미드를 나타내고 있다. 처음 입력 이미지의 크기를 여러 스케일에 맞춰 조정 후에 각각에 대하여 Ω 인 HOG 특징을 계산한다. 그러므로 전체 스케일에 따른 HOG 특징 맵을 구하므로 매우 많은 계산량이 요구된다.

이를 개선하기 위해 제안하는 이미지 특징 피라미드는 그림 2의 (b)에서 보는 바와 같은 전체적인 구조를 가진다. 제안하는 이미지 특징 피라미드 방식은 전체 스케일 단계 중 5개의 기준 스케일 $s \in (2, 1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8})$ 만을 사용한다. 여기서 기준이 되는 스케일 단계의 분해능을 높이는 것도 한 가지 방법일 수 있지만 최대한 넓은 범위로 잡은 이유는 HOG 특징 계산의 복잡성이 상대적으로 크기 때문에 이를 최대한 줄이기 위해 최소한의 스케일을 사용한다. 그리고 5개의 기준 스케일에 맞게 이미지 크기를 재조정 한 뒤 이에 대해서만 HOG 특징 맵을 계산한다. 제안하는 그림 2 (b)의 구조처럼 만들어진 5개의 HOG 특징 맵들로 부터 전체 이미지 HOG 특징 피라미드를 구성하기 위해 쌍선형 보간법을 적용한다.

쌍선형 보간법은 1차원에서의 선형 보간법을 2차원으로 확장한 것이다. 쌍선형 보간법을 이용하면 어떤 직사각형의 네 꼭지점의 값으로 이 사각형의 변 및 내부의 임의의 점에서의 값을 추정 할 수 있다. 어떤 객체의 HOG 특징은 각 스케일마다 지역적으로 비슷한 성분들로 구성되며 대부분의 균일한 영역의 경우는 스케일마다 큰 변화가 없고, 특정 복잡한 구조의 에지에서만 변화가 생기기 때문에 주어진 기준 특징 맵을 이용하여 다른 스케일의 특징 맵을 예측할 수 있다.

그림 3의 구조처럼 스케일 $s=1(x_1)$ 인 경우 입력 이미지의 특징 맵을 구한 후 그 다음 큰 스케일과 작은 스케일에서의 특징 맵을 구성하는데 있어 쌍선형 보간법을 적용한다. 이

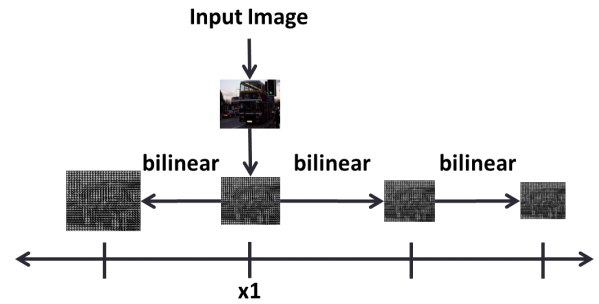


그림 3. 쌍선형 보간법을 이용한 특징 맵 보간

Fig. 3. Feature Map Interpolation using Bilinear Interpolation

렇게 만들어진 특징 맵에 다시 보간법을 적용하여 다음 단계의 특징 맵을 만드는 반복적인 과정을 통해 전체 이미지 특징 피라미드를 구성한다. 일반적인 쌍선형 보간법의 경우 기준 스케일 특징 맵의 값 4개를 이용하여 만들어야 할 스케일이 다른 특징 맵의 값 1개를 보간 한다. 이를 슬라이딩 윈도우(sliding window) 방식으로 전체의 특징 맵을 구성하고 구성된 특징 맵을 다시 이용하여 그 다음 특징 맵을 반복해서 구성하게 된다.

이렇게 쌍선형 보간법만을 이용해 특징 맵을 구성하면 기준 스케일에서 존재하던 특정 에지의 패턴이 사라지거나 손실이 일어날 수 있다. 이렇게 쌍선형 보간법만을 이용해 특징 맵을 구성하면 기준 스케일에서 존재하던 특정 에지의 패턴이 사라지거나 손실이 일어날 수 있다. 그림 4와 같이 A, B, C, D 중 한 개의 값이 나머지 값들에 비하여 상대적으로 크거나 작은 경우, A, B, C, D 중 두 개의 값이 비슷하면서 나머지 값들에 비하여 상대적으로 크거나 작은 경우 같은 특정한 패턴에서 에지에 대한 경사도 크기의 손실이 커진다. 이는 최종 스코어 계산에 대한 손실로 이어져 객체 검출의 정확성을 떨어뜨리게 된다.

A	B
C	D

i) $A \gg B = C = D$ or $A \ll B = C = D$

ii) $A = B \gg C = D$ or $A = B \ll C = D$

그림 4. 에지 손실을 야기하는 특정 특징 패턴들

Fig. 4. Specific Feature Patterns invoking The Loss of The Edge

따라서 이러한 보간법에 의한 잘못된 손실을 줄이고자 보간 처리가 잘 되는 균일한 영역과 특정한 패턴을 갖는 에지 영역을 나누는 개념을 도입한 알고리즘을 제안한다. 그림 5에서 제안하는 알고리즘은 크게 두 단계로 나누어져 있다. 첫 번째 단계에서는 먼저 입력되는 특징 맵의 균일한 영역을 판별해 내고 균일한 영역에 대해서는 쌍선형 보간법을 적용한다. 두 번째 단계에서는 균일한 영역으로 판별되지 않은 에지로 판별된 부분들에 대해서 특정 처리한다.

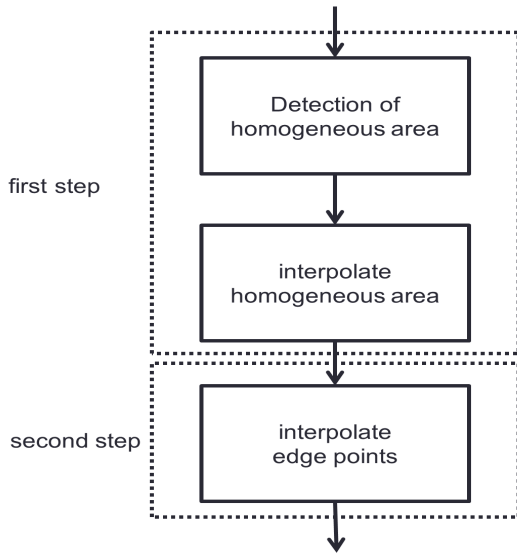


그림 5. 제안하는 영역 판단 알고리즘
Fig. 5. Proposed Region Decision Algorithm

첫 번째 단계에서 균일한 영역을 판별해 내기 위해 그림 6에서의 커널을 이용한다. 이 3x3 커널이 특징 맵에 슬라이딩 윈도우 방식으로 진행하면서 흰색 네 개의 포인트를 가

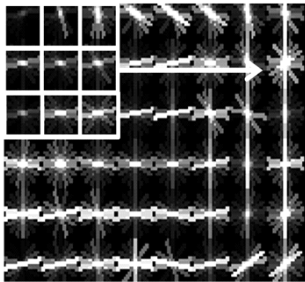


그림 6. 균일한 영역 판단을 위한 3x3 커널
Fig. 6. 3x3 kernel for uniform area decision

지고 검정색 세 개의 포인트에 대해서 에지인지 에지가 아닌지를 판별한다.

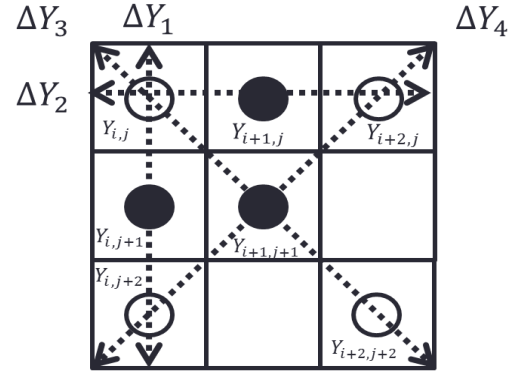


그림 7. 3x3 커널에서 네 방향에 대한 에지 정보들
Fig. 7. The Edge Information for Four Direction in 3x3 Kernel

에지 포인트들의 판별 여부를 위해 각 수평적, 수직적, 대각선, 역방향 대각선에 대한 에지 정보인 $\Delta Y_1, \Delta Y_2, \Delta Y_3, \Delta Y_4$ 를 구한다.

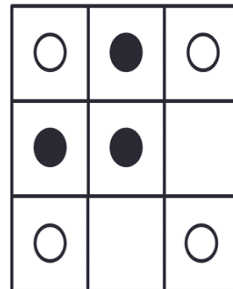
$$\Delta Y_1 = |Y_{i,j} - Y_{i+2,j}| \quad (2)$$

$$\Delta Y_2 = |Y_{i,j} - Y_{i,j+2}| \quad (3)$$

$$\Delta Y_3 = |Y_{i,j} - Y_{i+2,j+2}| \quad (4)$$

$$\Delta Y_4 = |Y_{i+2,j} - Y_{i,j+1}| \quad (5)$$

여기서 $Y_{i,j}$ 는 (i,j) 위치의 특징맵 값이고, ΔY_i for $i =$



● Point to determine
If it is an edge or not

○ Point used for
Edge determination

1,...,4는 각 방향에 대한 특징 값의 절대 차이값을 나타낸다. 이 값의 크기가 해당 방향에 대한 에지의 강인함을 나타낸다. 즉, 그 값이 작을수록 해당 주변 특징들과의 상관관계가 높다.

균일한 영역인지 에지 영역인지를 판별은 적정 임계값에 기반하며 그림 8의 알고리즘을 따른다. 판별해야 할 포인트, 그림 7의 $Y_{i+1,j}$, $Y_{i,j+1}$, $Y_{i+1,j+1}$ 들에 대해서 다음과 같은 판단을 한다. 먼저 수평적 판단 에지 정보를 가지고 있는 ΔY_1 이 적정 임계값보다 작을 경우 수평적 근접 특징 값들의 평균으로 취한다. 만약 적정 임계값보다 클 경우 이를 에지 포인트라 판단하고 두 번째 단계에서 필요하기 때문에 이에 대한 좌표를 따로 저장해 놓는다. 그리고 수평적 에지 판단 또한 수직적 에지 판단과 동일하게 진행한다.

대각선과 역방향 대각선 판단의 경우 $Y_{i+1,j+1}$ 의 포인트가 겹치기 때문에 좀 더 알고리즘을 세분화 할 필요가 있다. 만약 ΔY_3 , ΔY_4 가 모두 적정 임계값보다 작을 경우에는 변화량이 더 작은 값으로 에지 판단의 기준을 잡는다. 이 외의 경우는 수직적, 수평적 판단과 동일하다.

```

if  $\Delta Y_1 < threshold$ 
     $Y_{i+1,j} = \frac{Y_{i,j} + Y_{i+2,j}}{2}$ 
else
     $Y_{i+1,j}$  is edge point

if  $\Delta Y_2 < threshold$ 
     $Y_{i,j+1} = \frac{Y_{i,j} + Y_{i,j+2}}{2}$ 
else
     $Y_{i,j+1}$  is edge point

if  $\Delta Y_3$  and  $\Delta Y_4 < threshold$ 
    {
         $\Delta Y_{min} = \min\{\Delta Y_3, \Delta Y_4\}$ 
        if  $\Delta Y_{min} == \Delta Y_3$ 
             $Y_{i+1,j+1} = \frac{Y_{i,j} + Y_{i+2,j+2}}{2}$ 
        else ( $\Delta Y_{min} == \Delta Y_4$ )
             $Y_{i+1,j+1} = \frac{Y_{i+2,j} + Y_{i,j+2}}{2}$ 
    }
else if  $\Delta Y_3 < threshold$ 
     $Y_{i+1,j+1} = \frac{Y_{i,j} + Y_{i+2,j+2}}{2}$ 
else if  $\Delta Y_4 < threshold$ 
     $Y_{i+1,j+1} = \frac{Y_{i+2,j} + Y_{i,j+2}}{2}$ 
else
     $Y_{i+1,j+1}$  is edge point

```

그림 8. 에지 판별을 위한 알고리즘

Fig. 8. Algorithm for Edge Region Decision

위의 과정으로 균일한 영역과 에지 포인트들을 판단해내고 균일한 영역은 쌍선형 보간법으로 새로운 특징 맵을 구성한다. 그리고 에지 포인트는 두 번째 단계에서 별도 처리

한다.

두 번째 단계에서는 첫 번째 단계에서 적정 임계값을 넘어 판별된 에지 포인트들에 대해 적응적인 보간법을 적용한다. 크게 존재할 수 있는 경우의 수는 그림 9에서처럼 가운데 에지 포인트를 중심으로 그 주변에 에지 포인트가 하나라도 있는 경우 (a), 그리고 그 주변에 에지 포인트가 하나도 없는 경우 (b)로 나눌 수 있다. 여기서 검은색 포인트는 미리 판별된 에지 포인트들이고, 흰색 포인트는 에지 포인트가 아닌 지점으로써 첫 번째 단계에서 일반적인 쌍선형 보간법이 적용된 특징 포인트들이다.

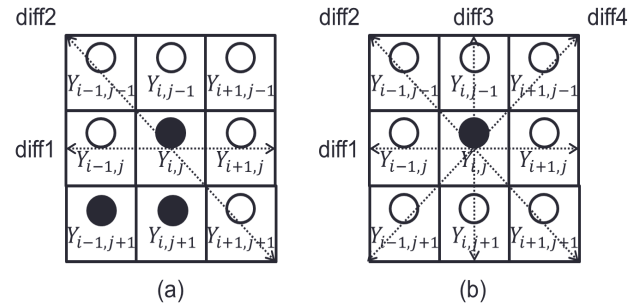


그림 9. 에지 포인트가 존재할 수 있는 경우의 수

Fig. 9. Cases where an Edge Points

두 번째 단계에서는 첫 번째 단계에서 적정 임계값을 넘어 판별된 에지 포인트들에 대해 적응적인 보간법을 적용한다. 크게 존재할 수 있는 경우의 수는 그림 9에서처럼 가운데 에지 포인트를 중심으로 그 주변에 에지 포인트가 하나라도 있는 경우 (a), 그리고 그 주변에 에지 포인트가 하나도 없는 경우 (b)로 나눌 수 있다. 여기서 검은색 포인트는 미리 판별된 에지 포인트들이고, 흰색 포인트는 에지 포인트가 아닌 지점으로써 첫 번째 단계에서 일반적인 쌍선형 보간법이 적용된 특징 포인트들이다.

그림 9의 (a)와 같이 중심 에지 포인트를 주변으로 하나 이상의 에지 포인트가 존재할 경우 $Y_{i-1,j+1}$, $Y_{i+1,j+1}$ 을 에지 포인트라고 가정하고 있다. 처음의 경우와 다르게 에지 포인트가 있는 경우, 이 에지 포인트들을 모두 피해야 할 필요가 있다. 이미 에지 처리된 포인트들에게 더 이상 에지에 대한 정보를 얻을 필요는 없다. 따라서 그림 9의 (a) 경우 식(6)과 식(7)을 통해 $diff_1, diff_2$ 를 계산하고 이 방향들에 대해서만 감안한다. 즉, 찾아야 할 방향이 정 가운데는 지나

지만 주변 에지 포인트들을 지나지 않는 방향성들을 찾는 것이다. 그리고 마찬가지로 가장 작은 차이를 가지는 방향에 대한 주변 특징 값으로 $Y_{i,j}$ 를 계산한다. 만약 식(8)에 의해 구해진 $diff_{\min}$ 가 $diff_1$ 이라면 $Y_{i-1,j}$ 와 $Y_{i+1,j}$ 의 상관관계가 $diff_2$ 방향에 해당하는 각 주변 특징 값들 보다 가장 가깝다. 따라서 식(9)에 의해 $Y_{i,j}$ 는 $Y_{i-1,j}$ 와 $Y_{i+1,j}$ 로 구해질 수 있다.

$$diff_1 = |Y_{i-1,j} - Y_{i+1,j}| \quad (6)$$

$$diff_2 = |Y_{i-1,j-1} - Y_{i+1,j+1}| \quad (7)$$

$$diff_{\min} = \min \{diff_1, diff_2\} \quad (8)$$

$$Y_{i,j} = \frac{Y_{i-1,j} + Y_{i+1,j}}{2} \quad \text{if } diff_{\min} = diff_1 \quad (9)$$

주변에 에지 포인트가 한 개라도 존재하지 않는 그림 9의 (b)의 경우는 모든 방향에 대해 고려하여 상관관계가 가장 높은 에지를 찾아야 한다. 따라서 가장 적합한 특징을 찾기 위해 수직, 수평, 대각선, 역방향 대각선에 대한 차이 $diff_1, diff_2, diff_3, diff_4$ 를 모두 구한다. 그리고 그 차이가 가장 작은 값에 해당하는 방향의 주변 특징 값 두 개를 이용하여 에지 포인트 $Y_{i,j}$ 를 계산한다.

$$diff_1 = |Y_{i-1,j} - Y_{i+1,j}| \quad (10)$$

$$diff_2 = |Y_{i-1,j-1} - Y_{i+1,j+1}| \quad (11)$$

$$diff_3 = |Y_{i,j-1} - Y_{i,j+1}| \quad (12)$$

$$diff_4 = |Y_{i+1,j-1} - Y_{i-1,j+1}| \quad (13)$$

$$diff_{\min} = \min \{diff_1, diff_2, diff_3, diff_4\} \quad (14)$$

$$Y_{i,j} = \frac{Y_{i-1,j} + Y_{i+1,j}}{2} \quad \text{if } diff_{\min} = diff_1 \quad (15)$$

만약 식(14)에 의해 구해진 $diff_{\min}$ 가 $diff_1$ 이라면 $Y_{i-1,j}$ 와 $Y_{i+1,j}$ 의 상관관계가 $diff_2, diff_3, \dots$ 방향에 해당하는 각 주변 특징 값들 보다 가장 가깝다. 따라서 식(15)에 의해 $Y_{i,j}$ 는 $Y_{i-1,j}$ 와 $Y_{i+1,j}$ 로 구해질 수 있다.

IV. 실험 결과 및 분석

본 논문의 실험에서는 The PASCAL Visual Object Classes(VOC) Challenge^[7]에서 제공하는 데이터 셋을 이용하였다. 전체적인 실험은 일반적인 쌍선형 보간을 적용한 DPM, 적응적 쌍선형 보간을 적용한 DPM 그리고 모든 스케일에 특징 피라미드를 갖는 기존 DPM을 비교 분석하였고 균일한 이미지 데이터 셋과 에지 이미지 데이터 셋을 세분화 하여 어떤 성향의 이미지에 대해 얼마나 더 강인함을 보이는지에 대해 평가하였다.

실험에는 윈도우 기반의 가상 머신 환경에서 3.4 GHz Intel Core i7 CPU, 16 GB, 8 processors, Ubuntu 14.04 64bit를 사용했으며, 매트랩에서 실험을 진행하였다. 성능 평가는 검출 성능 AP(Average Precision)와 검출 처리 시간을 지표로 정했다. 기존 이미지 특징 피라미드의 스케일은 입력 이미지의 크기에 따라 좌우되는데, 일반적인 기존 DPM의 경우 한 옥타브 당 8~12개의 스케일을 가진다. 본 실험에서는 입력 이미지 640x480일 경우 총 스케일은 46단계 특징맵을 구성하며, 각 옥타브 스케일당 10단계씩 스케일을 가지고 있으며 마지막 옥타브는 6개의 스케일만 가지고 있다. 본 실험에서는 $s \in (2, 1, 1/2, 1/4, 1/8)$ 에서의 HOG 특징맵을 구성하여 보간하도록 설계되었으며, 5개 기준 스케일로 부터 내부 보간을 통해 46단계의 HOG 특징 영상들이 만들어진다.

그림 10의 그래프는 자동차 객체에 대한 기존 DPM과 일

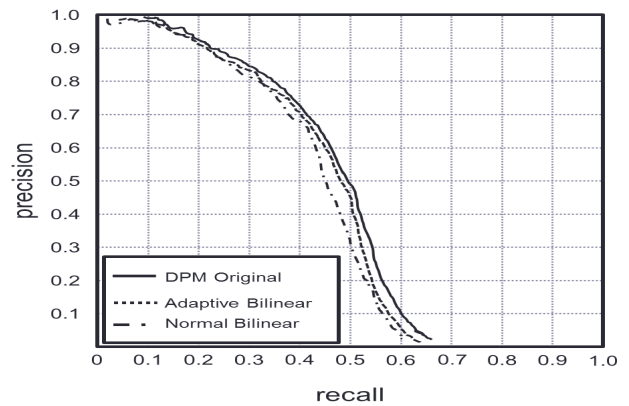


그림 10. 기존 DPM과 일반적인 쌍선형 보간법 DPM 그리고 적응적 쌍선형 보간법 DPM의 정확성 비교

Fig. 10. Average Precision Comparison(Original DPM vs Bilinear Interpolation DPM vs Adaptive Interpolation DPM)

반적인 쌍선형 보간법이 적용된 DPM, 그리고 적응적 쌍선형 보간법이 적용된 DPM의 검출 성능 비교를 나타내고 있다. 제안하는 알고리즘의 검출 성능을 비교하기 위해 AP (Average Precision)를 지표로 잡는다. AP를 계산하기 위해 선 정밀도(precision)와 검출율(recall)이 필요한데, 식(16)의 정밀도(precision)는 검출된 결과들 중 원하는 객체가 얼마나 포함되어 있는가, 식(17)의 검출율(recall)은 객체들을 빠뜨리지 않고 얼마나 잘 잡아내는가를 뜻한다. 그림 10에서 보는 바와 같이 이 두 값이 이루는 그래프에서 그래프 선 아래쪽의 면적으로 AP가 계산된다.

$$precision = \frac{TRUE\ detections}{whole\ detections\ of\ an\ algorithm} \quad (16)$$

$$recall = \frac{detected\ TRUE}{total\ number\ of\ existing\ TRUE} \quad (17)$$

기존 DPM의 경우 AP가 0.457, 일반적인 쌍선형 보간법을 적용한 DPM은 0.428로 6.34%의 검출 성능이 낮아졌다. 제안하는 적응적 쌍선형 보간법을 적용한 DPM의 경우는 AP가 0.441로 기존 DPM 대비 3.5%의 검출 성능이 낮아졌다. 표 1은 데이터 셋에서 제공하는 승용차를 비롯한 추가적인 19가지 객체(5105장)에 대해서 AP를 비교한 것이다. 20가지 물체에 대한 일반적인 쌍선형 보간을 적용한 이미지 피라미드 DPM의 경우는 평균 AP가 0.251로 기존 DPM의 AP 0.262와 비교하였을 때 0.011의 성능저하를 보였다. 하지만 적응적인 쌍선형 보간을 적용한 경우는 AP가 0.255로 기존 DPM 대비 0.007의 성능 저하를 보여 일반적인 쌍선형 보간을 적용하는 방식보다 적응적인 쌍선형 보간을 사용하

는 것이 검출 성능을 1.25% 향상시킴을 알 수 있었다.

연산 속도 측정은 20가지 물체에 대한 5105장을 검출해 내는데 걸리는 시간을 종합하여 평균으로 계산하였다. 표 2처럼 기존 DPM의 경우 평균 3.37초가 걸리고 제안하는 알고리즘의 경우 3초로 10.8%의 연산 속도가 향상된 것을 알 수 있었다.

표 2. 20가지 물체에 대한 기존 DPM과 적응적 쌍선형 보간법 DPM의 연산 시간 비교

Table 2. Computing Time Comparison for 20 Objects(Original DPM vs Adaptive Interpolation DPM)

	Computing Time
Original-DPM	3.369s
Adaptive-DPM	3.004s

제안하는 알고리즘은 균일한 영역을 먼저 처리하고 판단되는 에지 영역들에 대해서 후처리를 진행한다. 균일한 성향이 많고 에지한 성향이 적은 이미지들의 경우 후처리에 대한 연산의 양이 줄어들어 더 빠른 연산 속도를 보여줄 수 있다. 따라서 이미지들을 균일한 이미지, 일반적인 이미지, 그리고 에지한 이미지로 분류하여 실험을 진행하였다.

표 3은 각 50장씩 분류한 세 종류의 이미지들에 대해 연

표 3. 세 종류의 이미지에 따른 연산 시간 비교

Table 3. Computing Time Comparison for 3 Case Images

	Original-DPM	Adaptive-DPM (proposed)
homogeneous-images	2.892s	2.556s
normal-images	3.452s	3.163s
edge-images	3.627s	3.302s

표 1. 20가지 객체에 대한 기존 DPM과 일반적인 쌍선형 DPM, 적응적 쌍선형 DPM의 AP(Average Precision)비교

Table 1. Average Precision Comparison for 20 Objects(Original DPM vs Bilinear Interpolation DPM vs Adaptive Interpolation DPM)

	aero	bike	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	mbik	pers	plant	sheep	sofa	train	tv	Total Average
Original DPM	.290	.546	.006	.134	.262	.394	.457	.161	.163	.165	.245	.050	.436	.378	.350	.088	.173	.216	.340	.390	0.262
Normal bilinear DPM	.279	.523	.006	.128	.251	.377	.428	.152	.153	.157	.232	.049	.412	.362	.339	.086	.164	.206	.326	.374	0.251
Adap-bilinear DPM	.281	.536	.006	.129	.255	.384	.441	.155	.153	.158	.236	.049	.419	.368	.341	.087	.168	.212	.328	.376	0.255

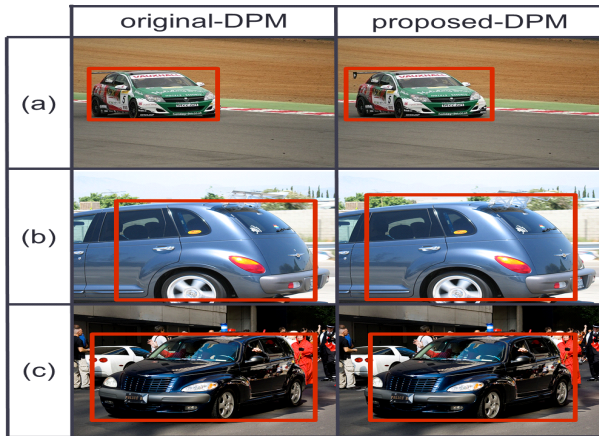


그림 11. 기존 DPM과 제안된 DPM의 객체 검출 결과 (a) 균일한 이미지, (b) 일반적인 이미지, (c) 에지한 이미지
Fig. 11. Object Detection Results(Original DPM vs Proposed DPM)
(a) Homogeneous Images, (b) Normal Images, (c) Edge Images

산 시간을 측정하여 평균으로 계산한 결과이다. 균일한 이미지들에 대한 평균 검출 시간은 2.556초로 일반적인 이미지 보다는 19.2% 만큼의 빠른 연산 속도를, 에지한 이미지 보다는 22.6% 만큼의 빠른 연산 속도를 보여주고 있다. 즉, 제안하는 알고리즘은 특정 패턴이 적게 들어가 있는 균일한 이미지들에 대해 더 효과가 좋음을 확인할 수 있었다.

그림 11은 분류한 세 종류의 이미지들, 그림 12는 다양한 객체에 대한 검출 결과들이다. 제안하는 적응적 쌍선형 보간법을 적용 시 바운딩 박스(Bounding Box)의 위치가 정확하게 기존 DPM과 같지만 객체들을 잘 찾아내는 것을 확인 할 수 있었다. 또한 그림 11에서 (b)의 경우 승용차의 일부가 잘라져 있음에도 불구하고 잘 검출함을 보였다.



그림 12. 다양한 객체에 대한 기존 DPM과 적응적 쌍선형 보간법 DPM의 객체 검출 결과 비교
Fig. 12. Object Detection Results(Original DPM vs Proposed DPM)

V. 결 론

지능형 차량 시스템인 자율 주행 자동차, 심각한 범죄나 위험한 재난 상황을 인식하는 지능형 CCTV에 대한 관심이 높아지면서 효율적인 객체 검출 연구는 획기적인 성능을 보여주고 있다. 그 대표적인 알고리즘으로 물체 부분 기반의 모델 특징을 이용하여 기하학적 강점이 있는 DPM (Deformable Part-based Models)이 있다. 하지만 복잡한 알고리즘으로 인해 정확성 측면에서 높은 성능을 보여주지만 그 만큼 계산 양이 증대되어 처리 속도가 늦다는 단점을 가지고 있다.

이를 개선하기 위해 본 논문에서는 많은 연산을 필요로 하는 특징 피라미드를 구성하는 과정 대신, 특정 스케일에서 구해진 소수의 특징 맵의 적응적인 쌍선형 보간법을 이용하여 새로운 특징 피라미드를 재구성하여 연산 속도를 줄이는 방법을 제안한다. 또한 소수의 특징 맵만으로 전체적인 특징 피라미드를 구성하는데 있어서 생기는 정확성 문제를 특정 조건에 따라 균일한 영역과 에지 영역으로 나누는 알고리즘을 제안하여 최대한 보상하였다.

결과적으로 적응적 쌍선형 보간 기반의 이미지 피라미드를 적용한 DPM은 기존 DPM에 비해 빠른 속도로 객체를 검출 할 수 있음을 보였다. 기존 3.369초에서 3.004초로 연산 시간을 약 10% 더 줄였다. 또한 제안하는 알고리즘은 균일한 이미지들에 대해서는 2.556초로 에지한 이미지들보다 22.6%의 연산 시간을 더 줄였다. 그리고 일반적인 쌍선형 보간을 이미지 피라미드에 적용한 DPM의 경우, 기존 DPM 대비 평균 AP가 4.07% 낮아졌고 에지 영역에 대한 특정 처리로 객체 검출의 에러율을 줄이는 적응적 쌍선형 보간을 적용한 경우 2.82%의 검출 성능이 낮아졌다. 즉, 단순 쌍선형 보간을 적용하는 방식보다 적응적인 쌍선형 보간을 사용하는 것이 검출 성능을 1.25% 향상시킴을 알 수 있었다.

본 논문에서 제안된 적응적 쌍선형 보간을 적용한 이미지 피라미드 기반의 DPM은 실제 어플리케이션에 보다 빠

른 처리 속도로 적용하여 스마트 자동차, 자율주행 자동차, 지능형 CCTV 시스템 등에 널리 이용 될 수 있다.

참 고 문 헌 (References)

- [1] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in Proc. 2001 IEEE Comput. Soc. Conf. Comput. Vision and Pattern Recognition, vol. 1, pp. 511-518, Hawaii, U.S.A., Dec. 2001. <http://dx.doi.org/10.1109/CVPR.2001.990517>
- [2] Dalal, N and Triggs, B, "Histograms of oriented gradients for human detection". Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 886-893, 2005. <http://dx.doi.org/10.1109/CVPR.2005.177>
- [3] Qiang Zhu, Shi Avidan, Mei-Chen Yeh, Kwang-Ting Cheng, "Fast Human Detection Using a Cascade of Histograms of Oriented Gradients", Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol.2, pp.1491-1498, 2006. <http://dx.doi.org/10.1109/CVPR.2006.119>
- [4] P. Dollar and Z. Tu, P. Perona, S. Belongie, "Integral channel features", British Machine Vision Conference, pp. 1-11, 2009. doi:10.5244/C.23.91
- [5] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, D. Ramanan, "Object Detection with Discriminatively Trained Part Based Models", Pattern Analysis and Machine Intelligence, Vol.32, No.9, pp.1627-1645, 2010. <https://doi.org/10.1109/TPAMI.2009.167>
- [6] R. Benenson, M. Mathias, R. Timofte, and L. Van Gool, "Pedestrian detection at 100 frames per second", In IEEE Conference on Computer Vision and Pattern Recognition, pp. 2903-2910, 16-21 June, 2012. <https://doi.org/10.1109/CVPR.2012.6248017>
- [7] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, A. Zisserman, "The PASCAL Visual Object Classes(VOC) Challenge", International Journal of Computer Vision, Vol.88, pp.303-338, 2010. <https://doi.org/10.1007/s11263-009-0275-4>
- [8] P. Felzenszwalb and D. Huttenlocher. Distance transforms of sampled functions. Cornell Computing and Information Science Technical Report TR2004-1963, September 2004. <https://hdl.handle.net/1813/5663>
- [9] P. Felzenszwalb and D. Huttenlocher. Pictorial structures for object recognition. Internat. J. Computer Vision, Vol.61, No.1, pp.55-79, Jan, 2005. <https://doi.org/10.1023/B:VISI.0000042934.15159.49>
- [10] P. Burt and E. Adelson. "The Laplacian Pyramid as a Compact Image Code," IEEE Transaction on Communications, Vol.31, No.4, pp.532-540, April 1983 <https://doi.org/10.1109/TCOM.1983.1095851>

저 자 소 개



한 규 동

- 2017년 2월 : 한국산업기술대학교 전자공학부 공학사
- 2019년 2월 : 한국산업기술대학교 전자공학부 공학석사
- 2019년 4월 ~ 현재 : ㈜ 웨이티즈 주임 연구원
- ORCID : <https://orcid.org/0000-0001-6348-1698>
- 주관심분야 : 딥러닝기반 영상처리, 자율주행차용 영상신호처리



김 응 태

- 1991년 : 인하대학교 전자공학과 공학사
- 1993년 : KAIST 전기및전자공학과 공학석사
- 1999년 : KAIST 전기및전자공학과 공학박사
- 1998년 3월 ~ 2004년 2월 : (주)LG전자 DTV연구소 책임연구원
- 2004년 3월 ~ 현재 : 한국산업기술대학교 전자공학부 교수
- ORCID : <https://orcid.org/0000-0001-5984-0045>
- 주관심분야 : 멀티미디어 신호처리, DTV SOC, 지능형 영상감시 시스템, 딥러닝기반 영상처리