

DNN과 HoG Feature를 이용한 도로 소실점 검출 방법

Method for Road Vanishing Point Detection Using DNN and Hog Feature

윤대은, 최형일
숭실대학교 미디어학과

Dae-Eun Yoon(vhzpt109@naver.com), Hyung-Il Choi(hic@ssu.ac.kr)

요약

소실점이란 실제 공간의 평행한 선들이 영상 내에 투영되면서 한곳에 모이는 점으로, 도로 공간에서의 소실점은 매우 중요한 공간정보이다. 도로 공간에서의 소실점을 이용해 추출된 차선의 위치를 개선하거나, 깊이지도 영상을 생성할 수 있다. 본 논문에서는 자동차의 시점을 기준으로 도로를 촬영한 영상을 Deep Neural Network(DNN)과 Histogram of Oriented Gradient(HoG) Feature를 이용한 소실점 검출 방법을 제안한다. 제안하는 알고리즘에서는 영상을 블록별로 나눠서 주요 에지 방향을 추출하는 HoG Feature 추출 단계와 DNN 학습 단계, 그리고 Test 단계로 나뉜다. 학습단계에서는 자동차 시점으로 기준으로 도로 영상 2300장으로 학습을 진행한다. 그리고 Test 단계에서는 Normalized Euclidean Distance(NormDist) 방법을 사용하여 제안하는 알고리즘의 효율성을 측정한다.

■ 중심어 : | 소실점 | DNN | HoG Feature | NormDist |

Abstract

A vanishing point is a point on an image to which parallel lines projected from a real space gather. A vanishing point in a road space provides important spatial information. It is possible to improve the position of an extracted lane or generate a depth map image using a vanishing point in the road space. In this paper, we propose a method of detecting vanishing points on images taken from a vehicle's point of view using Deep Neural Network (DNN) and Histogram of Oriented Gradient (HoG). The proposed algorithm is divided into a HoG feature extraction step, in which the edge direction is extracted by dividing an image into blocks, a DNN learning step, and a test step. In the learning stage, learning is performed using 2,300 road images taken from a vehicle's point of views. In the test phase, the efficiency of the proposed algorithm using the Normalized Euclidean Distance (NormDist) method is measured.

■ keyword : | Vanishing Point | DNN | HoG Feature | NormDist |

I. 서론

최근 기술의 발달로 인해 4차 산업 혁명에 대한 관심

이 증대되고 있다. 이와 관련하여 사람의 감각중 제일 많은 비중을 차지하는 시각적 역할을 대체 가능한 인공 지능, 즉 컴퓨터 비전에 대한 연구가 활발히 이루어지

* 본 연구는 2017년도 정부의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업입니다.

(NRF-2017R1D1A1B03034114)

접수일자 : 2018년 11월 05일

수정일자 : 2018년 11월 21일

심사완료일 : 2018년 11월 22일

교신저자 : 최형일, e-mail : hic@ssu.ac.kr

고 있다. 예를 들어 운전자가 차량을 운전하지 않아도 스스로 움직이는 자율주행 자동차는 충돌 감지, 전방 이동물체 인식, 전방 환경인지 등의 사물 인식 기술이 필요하다. 환경 인지 기술은 크게 도로 검출기술과 이동물체 인식 기술로 구분된다. 즉, 전방의 도로 영역을 검출하고 영역내외의 이동하는 물체를 인지하고 인식하여 도로 상의 환경을 분석하는 것이다. 이를 위해선 소실점이 중요하다. 미리 정확한 소실점의 위치를 알고 있다면 도로의 경계선을 검출, 개선 할 수 있다. 즉, 정확한 소실점을 추출함으로써 도로 영역의 공간 정보들을 얻을 수 있다. 그러나 도로 환경같이 실외에 존재하는 수많은 잡음들로부터 강건하게 도로 소실점을 검출하는 것은 매우 어렵고 이에 대한 연구는 다양하게 이루어지고 있다. [1]과 [2]는 소실점과 그 소실점을 잇는 에지(edge)를 연결함으로써 도로(road)를 검출하였다. 소실점은 2차원 영상의 3차원 변환에 대해서도 소실점을 검출함으로써 깊이 지도를 생성할 수 있다. [3]은 영상으로부터 소실점 검출을 통해 깊이정보를 예측하였다. [4]는 Deep Neural Network(DNN) 모델에 Convolution 전처리 과정을 추가한 Convolutional Neural Network(CNN)을 이용해 소실점을 예측하였다. [5]은 Histogram of Orented Gradient(HoG)를 이용해 소실점을 검출하였다. 본 논문에서는 HoG Feature들과 DNN을 접목시켜 도로 소실점을 검출하는 방법을 제안하고자 한다. 제안하는 방법은 도로 영상 내에 소실점이 있으면 검출 할 수 있다. 논문의 구성은 다음과 같다. II장에서 HoG Feature 추출과 DNN Architecture에 대해서 설명하고 DNN Training과 소실점 예측에 대해서 설명한다. III장에서는 Test 영상에 대해서 제안하는 알고리즘의 효율성을 측정하고, IV장에서 결론을 논의한다.

II. 제안하는 방법

본 논문에서 제안하는 방법은 영상을 Resize 후 Gray Scale로 변환하는 단계와 HoG Feature 추출 단계, 추출한 HoG Feature를 DNN에 입력하여 학습을 진행 한 후 소실점의 좌표를 예측하는 단계로 나뉜다.

표 1. 알고리즘 순서

Algorithm
1. 원 영상을 Resize 및 GrayScale로 변환
2. Resize 및 GraScale된 영상에서 HoG Feature 추출
3. 추출된 HoG Feature를 Neural Network에 입력
4. 학습 진행
5. 학습된 Network에 Test 영상을 입력하여 소실점 X, Y좌표 예측

1. HoG Feature 추출

Histogram of Oriented Gradient(HoG)는 영상 안의 영역 내의 픽셀의 변화(Intensity)량의 각도와 크기를 고려하여 히스토그램 형태의 Feature를 추출하는 방법이다. 본래 객체 검출의 목적으로 고안된 기술로, [6]에서 HoG를 이용한 사람 검출 알고리즘을 제안하였다. HoG는 다음과 같이 구현된다. 먼저 영상 I에 대해 커널 G를 연산하여 x축, y축 주요 에지만 남게 한다.

$$G = I \times [-1, 0, 1] + I \times [-1, 0, 1]^T \quad (1)$$

그 다음 구해진 x축 에지와 y축 에지를 Arctan 함수를 이용하여 각 블록마다 Orientation을 구한다. 이 단계에서 블록마다 overlap 사용 여부를 선택할 수 있지만 본 논문에서는 overlap을 사용하지 않았다.

$$\theta = \tan^{-1} \left(\frac{dy}{dx} \right) \quad (2)$$

구해진 Orientation을 모두 사용하면 데이터가 너무 많아지므로 N만큼 등분하여 Bin으로 표현한다. 본 논문에서는 HoG Feature를 추출하기 전에 모든 영상의 일관성을 주기위해 먼저 영상을 Gray Scale로 변환한다. 그 후 같은 크기로 Resize한다. 제안하는 방법에서는 160x80의 크기로 Resize하였다. 그 다음 영상에서 블록 단위로 HoG Feature를 추출한다. 제안하는 방법에서는 영상을 8x8 블록 단위로 나누었고, 18등분으로 에지 방향을 분할하였다. 본래 HoG는 n개의 가장 주요한 방향을 선택해서 사용하지만 제안하는 방법에서는 n개의 주요한 방향이 아닌 모든 에지의 방향을 사용한다. 그 후 추출한 HoG Feature는 1D Vector로 변환한 뒤 DNN에 입력하여 사용한다.



그림 1. 원 영상(a)과 원 영상을 8x8 블록으로 나눈 후 18개의 에지 방향으로 분할한 HoG 영상(b)

2. Network Architecture

Deep Neural Network는 지난 몇 십 년 동안 널리 사

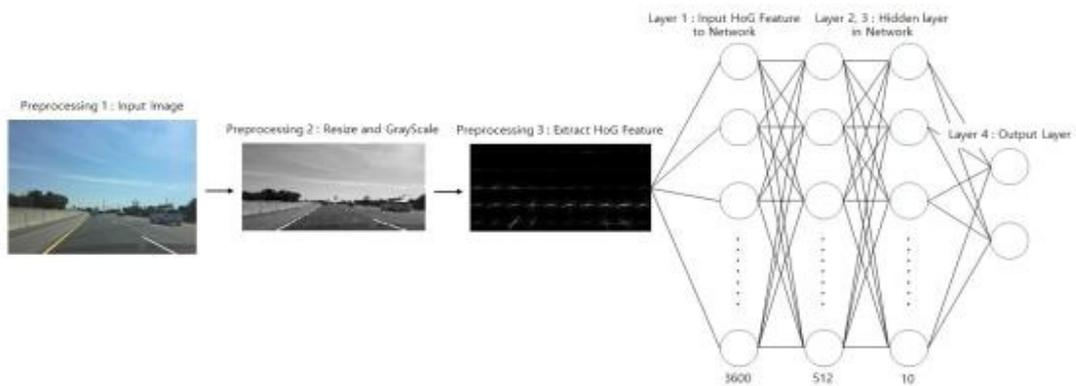


그림 2. Network Architecture

- 첫 번째 Layer는 세 번째 Preprocessing에서 추출된 HoG Feature를 DNN에 입력한다. DNN Input Layer는 3600개의 뉴런(160x80 영상을 8x8 블록으로 나누었을 때 모든 블록의 18-에지 방향에 대한 각각의 확률, 200*18)으로 구성된다.
- 두 번째 Layer는 DNN의 Fully-Connected Layer로 512개의 뉴런으로 구성된다.
- 세 번째 Layer는 Fully-Connected Layer로 10개의 뉴런으로 구성된다.
- 네 번째 Layer는 Fully-Connected Layer의 마지막 출력 Layer로, 2개의 뉴런으로 구성된다.

마지막 Layer에서 소실점 값을 예측한다. 마지막 Layer에서 2개의 뉴런이 예측된 소실점의 값인 X좌표와 Y좌표를 의미한다.

용되어 왔으며, 다양한 분야에 응용되어 사용되어지고 있다. 본 논문에서는 앞 절에서 구한 HoG Feature들을 DNN Architecture에 입력하여 학습을 진행한다. 제안하는 알고리즘의 Architecture는 [그림 2]와 같으며 다음과 같은 단계를 따른다.

- 첫 번째 Preprocessing에서 영상을 입력 받는다.
- 두 번째 Preprocessing는 영상을 160x80으로 Resize, Gray Scale로 변환하는 과정이다.
- 세 번째 Preprocessing는 영상에서 HoG Feature를 추출한다.

3. DNN Training and Data Set

본 논문에서는 100개의 배치사이즈를 사용하여 1000 epoch로 네트워크 학습을 진행했다. 네트워크의 Cost Function은 RMSE(Root Mean Square Error)를 사용하였고, 0.00001의 Learning rate로 Adam Optimizer를 사용하였다. 이미지 Data Set은 총 2700장으로, Training Data와 Test Data로 나누었다. 2300(highway 2000, city 300)장은 Training Data로 사용하였고 400(highway 200, city 200)장은 Test Data로 사용하였다. 본 논문에서 사용한 Data Set은 Image와 Label 2700쌍으로 이루어져있으며 자동차 시점 도로 영상을 업로드 해놓은 [8]에서 다운로드하였고[4] Data Set과 같다. 각 이미지를 160x80으로 Resize했을 때 소실점 위치 X좌표와 Y좌표(Ground Truth)를 txt 파일로 만들어 저장하였다.

4. Vanishing Point Prediction

Traning Data에 대해서 학습을 끝낸 후 Test Data 대해서 효율성을 측정한다. Ground Truth와 비교된 Vanishing Point Prediction의 예시를 [그림 3-5]에서 보여준다.

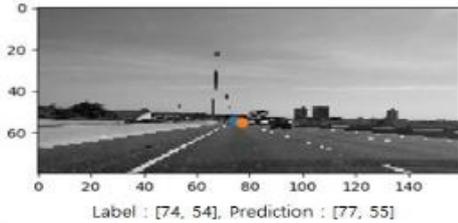


그림 3. 한 장의 영상에 대해 예측된 소실점의 예시. 파란 원은 Ground Truth를 나타내고 주황 원은 예측된 소실점

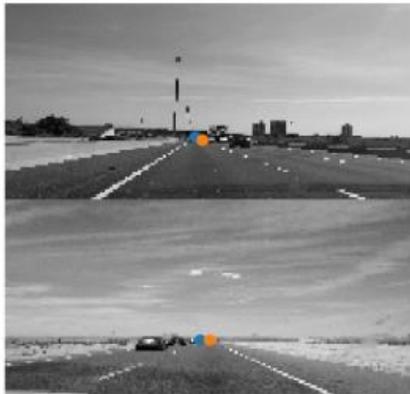


그림 4. 영상에 대해 예측된 소실점의 예시. 파란 원은 Ground Truth를 나타내고 주황 원은 예측된 소실점

III. 정확도 평가

정확도 평가를 위해 앞서 II, 3절에서 설명한 Test Data 이미지 400(highway 230, city 170)장을 사용한다. 소실점 정확도 평가를 위해, 본 논문에서는 Norm Distance 방법을 사용한다. Norm Distance는 Normalized Euclidean Distance라고도 불린다. 일반적으로 두 점의 거리를 구하기 위해선 Euclidean Distance를 많이 사용되지만 Euclidean Distance를 사

용하게 되면 이미지 해상도에 따라 오차가 다르게 나오게 되므로 실제 성능을 나타내는 것은 어렵다. 그러므로 정확한 측정을 위해선 Norm Distance 방법을 사용하는 것이 좋다. Norm Distance는 아래의 식과 같이 정의된다.

$$NormDist = \frac{\| VPP - VPG \|}{diagonal} \quad (3)$$

VPP는 예측한 소실점의 좌표이고 VPG는 실제 소실점의 좌표이다. 'diagonal'은 이미지의 대각선의 길이를 의미한다. 즉 Euclidean Distance를 이미지의 대각선으로 나누는 것이다. [표 2]에서는 Test Data Set에 대해 제안하는 알고리즘의 NormDist Error 평균을 보여준다. NormDist를 사용해 오류를 측정할 때, 1에 가까운 값이 나온다면 실제 소실점의 좌표와 예측된 소실점의 좌표가 멀다는 뜻이고, 반면에 0에 가까운 값이 나온다면 실제 소실점의 좌표와 예측된 소실점의 좌표가 가깝다는 뜻이다. 고속도로 NormDist는 [4]에 비해 성능이 떨어지지만 도시 NormDist와 평균 NormDist 부분에서 기존 도로 소실점 검출 방법에 비해 월등한 효과를 보인다.

표 2. Test Data Set에 대한 NormDist Error

Approach	Average	Highway	City
Our	0.0336	0.0322	0.0354
[4]	0.0445	0.0261	0.0507

[그림 6]은 Test Data Set에 대해 제안하는 알고리즘의 NormDist Error Histogram을 보여준다. x축은 NormDist이고 y축은 NormDist에 해당하는 이미지의 개수이다. Histogram이 왼쪽으로 몰려있을수록 좋은 성능을 나타내고, 오른쪽에 몰려있을수록 좋지 않은 성능을 나타낸다. 제안하는 알고리즘의 NormDist Error Histogram은 왼쪽에 몰려있으므로 좋은 성능을 나타내고 있다. 다음은 알고리즘 수행 시간을 측정한다, HoG Feature를 추출하고 DNN에 입력하여 소실점 좌표 예측까지 이미지 당 평균 250 milliseconds가 걸린다. Test에 사용된 컴퓨터의 성능은 Intel I5-3470 (Desktop CPU), NVIDIA GeForce GTX 1050 Ti, 8gb Ram(DDR3)로 이루어져 있다. 보다 좋은 성능의 CPU에서는 실시간으로 소실점의 좌표를 계산할 수 있다.

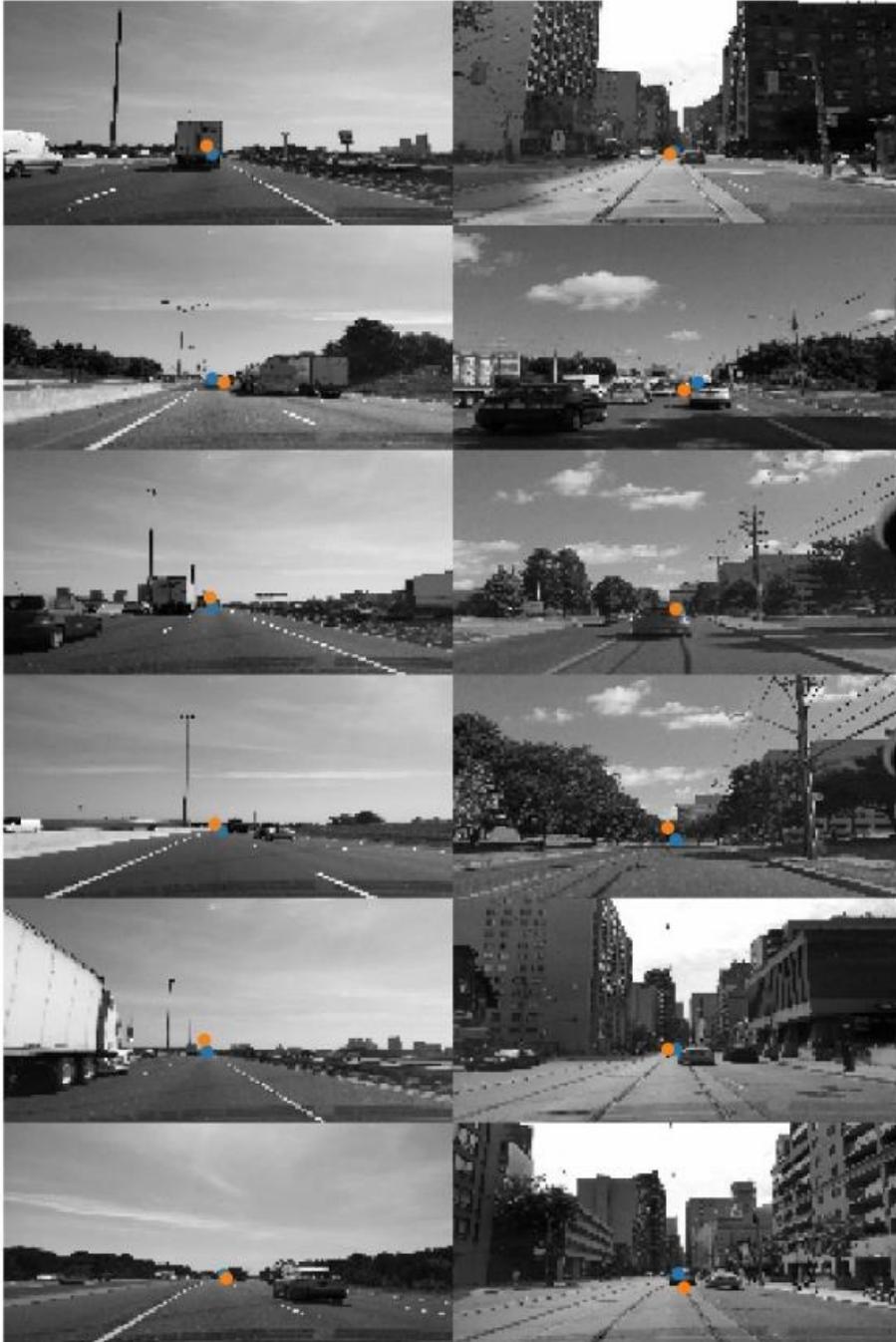


그림 5. 영상에 대해 예측된 소실점의 예시. 파란 원은 Ground Truth를 나타내고 주황 원은 예측된 소실점, 첫 번째 열은 고속도로, 두 번째 열은 도시 영상이다.

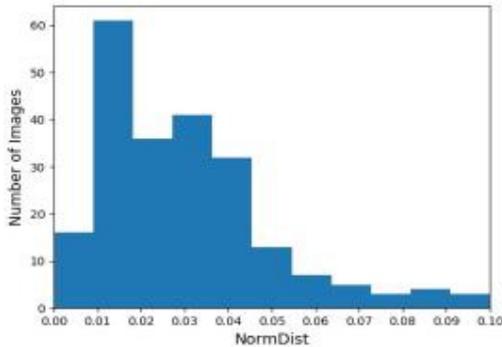


그림 6. Test Data Set에 대한 NormDist Error 11-bins Histogram

IV. 결론

본 논문에서는 DNN과 HoG Feature를 이용하여 도로 공간 영상 내의 소실점의 위치를 예측하는 방법을 제안하였다. 측정 결과가 좋지만 몇 가지 개선 사항이 있다. 하나는 각 이미지에서 필요 없는 객체(표지판, 구름, 가로등 등)를 제거한 후 HoG Feature를 추출하는 것이다. 모든 HoG Feature를 사용하다보니 잡음에 강인하지 않은 단점이 있다. 그 단점을 이미지에서 필요 없는 객체를 제거 한 후 HoG Feature를 추출한다면 네트워크의 성능을 높일 수 있을 것이다. 대부분의 필요하지 않은 객체들은 이미지 상단부분에 존재하며 이 점을 이용한다면 충분히 필요 없는 객체들을 제거한 후 HoG Feature를 추출할 수 있을 것이다. 두 번째로는 Data Set이다. 본 논문에서는 Training Data로 2300장의 영상을 사용하였는데 Training을 위해서는 많지 않은 양의 Data이다. Training Data의 증가는 네트워크의 성능을 높이는 것에 큰 영향을 줄 것이다.

참고 문헌

[1] C. Rasmussen, "Grouping dominant orientations for ill-structured road following," in Proc. of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol.1,

pp.I-470-I-477, 2004.

[2] K. Hui, J. Y. Audibert, and J. Ponce, "Vanishing point detection for road detection," in IEEE Conference on Computer Vision and Pattern Recognition, pp.96-103, 2009.

[3] S. Battiatto, A. Caprab, S. Curtib, and M. La Casciac, "3D Stereoscopic Image Pairs by Depth-Map Generation," Proceedings of the 3D Data Processing, Visualization, and Transmission, 2nd International Symposium, pp.124-131, 2004.

[4] R. Itu, D. Borza, and R. Danescu, "Automatic extrinsic camera parameters calibration using Convolutional Neural Networks," In: 2017 IEEE 13th International Conference on Intelligent Computer Communication and Processing (ICCP 2017), pp.273-278, 2017.

[5] 최지원, 김창익, "Histogram of Gradient를 이 한 실시간 소실점 검출," 전자공학회논문지-SP, 제 48권, 제2호, pp.96-101, 2011.

[6] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol.1, pp.886-893, 2005.

[7] OpenStreetCam, online: <https://www.openstreetcam.org/>, 2018.

저자 소개

윤 대 은(Dae-Eun Yoon)

준회원



- 2017년 8월 : 숭실대학교 평생교육원 컴퓨터공학과(공학사)
- 2017년 9월 ~ 현재 : 숭실대학교 미디어학과 석사과정

<관심분야> : 컴퓨터 비전, 딥 러닝 등

최 형 일(Hyung-II Choi)

정회원



- 1979년 : 연세대학교 전자공학과 (공학사)
 - 1983 : 미시간대학교 전기전산학과(공학석사)
 - 1987년 : 미시간대학 전기전산학과(공학박사)
 - 1989년 ~ 1999년 : 숭실대학교 컴퓨터학부 교수
 - 2000년 ~ 현재 : 숭실대학교 미디어학과 교수
- <관심분야> : 컴퓨터 비전, 패턴인식, 증강현실 등