

DNN과 슈퍼픽셀을 이용한 실내 공간 인식[☆]

Indoor Space Recognition using Super-pixel and DNN

김기상¹ 최형일^{1*}
Kisang Kim Hyung-Il Choi

요약

본 논문은 DNN(Deep Neural Network)와 슈퍼픽셀을 이용한 실내 공간 인식 알고리즘을 제안한다. 영상으로부터 실내 공간 인식을 위해 우선 영상 분할을 위한 세그멘테이션 프로세스가 필요하다. 이를 위해 본 논문에서는 적당한 크기로 나눌 수 있는 슈퍼 픽셀 알고리즘을 이용해 세그멘테이션을 수행한다. 각 세그먼트를 인식하기 위해 세그먼트마다 제안하는 방법을 이용하여 특징을 추출한다. 추출된 특징들을 DNN을 이용하여 학습하고, 학습으로부터 추출된 DNN모델을 이용하여 각 세그먼트를 인식한다. 실험 결과를 통해 제안하는 방법과 기존의 알고리즘과의 성능 비교 분석을 한다.

☞ 주제어 : 딥러닝, 슈퍼픽셀, 실내 공간 인식

ABSTRACT

In this paper, we propose an indoor-space recognition using DNN and super-pixel. In order to recognize the indoor space from the image, segmentation process is required for dividing an image Super-pixel is performed algorithm which can be divided into appropriate sizes. In order to recognize each segment, features are extracted using a proposed method. Extracted features are learned using DNN, and each segment is recognized using the DNN model. Experimental results show the performance comparison between the proposed method and existing algorithms.

☞ keyword : Deep Learning, Super-pixel, Indoor-space recognition

1. 서론

4차 산업 혁명으로 인하여 인공지능에 대한 연구가 다방면 다분야에 걸쳐 연구가 진행되고 있다. [1] 이러한 이유는 이전에는 적용하기 어려웠던 지능형 컴퓨팅이 점차 상용화가 되고 있기 때문이다. 특히, 최근에는 실감 인터랙션인 가상현실(VR), 증강현실(AR), 공존현실(CR)을 위해 컴퓨터 비전, 데이터 마이닝, HCI(Human-Computer Interaction) 분야에 많은 투자 및 연구가 진행되고 있다. 이러한 기술들을 활용하여 3D 스캐너[2], 실내 가상 인터리어 서비스[3], 홀로렌즈[4], 실내 네비게이션[5] 등 삶의 편의성 및 즐거움을 위한 다양한 기술에 대해 연구가 진행되고 있다. 하지만, 현재 이러한 기술들은 제한적인 상

용화에 그치고 있는 실정이다. 객체 검출, 인식, 추적 등의 문제도 있지만, 배경 및 환경에 대해 인식 기술의 한계가 상용화의 어려움으로 발전하고 있다. 따라서, 이러한 문제점을 해결하기 위해 배경 인식에 대해 다양한 연구가 진행되고 있다.

특히 증강현실 및 가상현실과 관련성이 높은 실내 공간 인식은 다양한 분야에서 연구가 진행되고 있다. 실내 공간 인식 방법으로는 크게는 추론을 이용한 공간 인식 방법과 학습을 이용한 방법이 있다. 추론을 이용한 방법으로는 소실점으로부터 기하학적 추론을 이용해 공간 정보 후보를 추출하고 이 후보군들을 지향성 지도를 통해 평가하여 공간 구조를 추출하는 방법 [6], 지향성 지도를 통해 실내 공간 영역과 그 외의 영역을 구분하고, 표면 레이어아웃 정보를 통해 공간 구조를 추출하는 방법 [7] 등이 있다. 이러한 방법들은 통로가 있거나, 클러스터의 상태에 따라 공간 정보가 제대로 추출되지 못하는 단점이 있다. 학습을 이용하는 방법으로는 영상을 세그멘테이션을 통해 영역을 군집화 하고, 각 군집화 된 영역을 Boosted Forest를 통해 인식하는 방법[8]이 있다. 하지만, 이 방법은 상용화 될 정도의 인식률이 확보되지 않은 문제점이 있다.

¹ School of Media, Soongsil University, Seoul, 07027, Korea

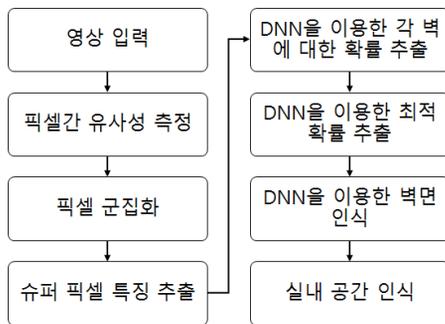
* Corresponding author (hic@ssu.ac.kr)

[Received 24 January 2018, Reviewed 26 January 2018(R2 8 March 2018), Accepted 9 April 2018]

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

(NRF-2017R1D1A1B03034114)

본 논문에서는 기존의 학습 방법을 개선을 통해 인식률을 개선한 방법을 제안한다. 입력 영상이 들어오면 슈퍼픽셀을 이용하여 세그멘테이션 된 영상을 생성한다. 각 세그먼트는 학습을 위한 입력데이터를 추출하기 위해 세그먼트들의 특징을 추출한다. 추출된 특징들을 이용하여 DNN(Deep Neural Network)를 이용해 학습한다. 기존에 사용되었던 Boosted Forest는 하나의 Threshold를 검증하는데 있어 단일 특징을 사용한다. 이러한 점은 분류하는데 있어 다양한 구도에서 볼 수 없는 것이 단점으로 작용한다. 따라서, 본 논문에서는 이러한 문제를 해결할 수 있는 DNN을 통해 학습을 수행하였다. 그림1은 본 논문에서 제안하는 알고리즘의 전체 흐름도를 나타낸다.



(그림 1) 전체 흐름도
(Figure 1) Overall process

본 논문의 2장에서는 기존의 세그멘테이션과 슈퍼픽셀간의 비교하며, 슈퍼픽셀 알고리즘과 생성방법에 대해 설명한다. 3장에서는 학습을 위한 DNN의 구성과 입출력 인터페이스에 대해 설명한다. 4장에서는 기존의 Boosted Forest를 이용한 학습 방법과 DNN간의 성능을 비교 및 실내공간 인식 예제를 나타내며, 마지막으로 5장에서는 추후 연구의 필요성 및 결론에 설명한다.

2. 슈퍼픽셀을 이용한 특징 추출

슈퍼픽셀은 세그멘테이션 기법 중 하나로, 영상분할, 객체검출, 객체추적 등 다양한 분야에서 전처리 과정으로 사용되는 방법이다. 입력된 영상으로부터 의미 있는 객체, 배경들을 분할하기 위해서 세그멘테이션은 필수이다. 대표적인 세그멘테이션 방법으로는 마르코브 랜덤 필드 [9], EM을 이용한 세그멘테이션 [10], k-NN [11] 등 다양한 방법이 있다. 하지만 이들 방법은 세그멘테이션을 위

해 학습이 전제 조건으로 필요하거나, k-NN의 경우 세그먼트의 개수를 입력으로 필요로 하는 한계점이 있다. 이러한 한계점을 극복하기 위해, 최근에는 슈퍼픽셀이라는 세그멘테이션 방법이 연구되고 있다. 하지만 최근에 나온 방법들은 프로세싱 시간을 고려하지 않고 구현되기 때문에, 슈퍼픽셀을 추출하기 위해 시간이 오래 걸리는 문제가 있다.

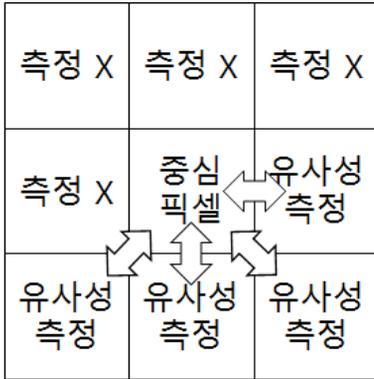
벽면 검출은 형태가 없는 것으로서 기존의 객체 검출 알고리즘을 적용하기엔 어려움이 많다. 벽면에 대한 영상 생성시, 카메라의 위치, 각도, 방향 등에 따라 생성되는 영상의 벽면은 전혀 다른 형태를 갖게 된다. 따라서, 이러한 문제를 해결하기 위해, 형태가 없는 것은 그림 2에서와 같이 영역 분할을 하고 각 영역이 어느 부분에 해당하는지 인식한 후 이를 하나로 묶어 인식하는 방식을 사용한다. [12] 따라서 본 논문에서는 슈퍼픽셀을 이용하여 각 세그먼트에 대해 학습하고 이에 대한 인식을 하는 방법을 제안하며, 이에 가장 적합한 슈퍼픽셀은 2.1장에서 소개하는 슈퍼픽셀 생성 알고리즘을 사용하였다.



(그림 2) 배경영상 인식 예
(Figure 2) Example of Background recognition

2.1 슈퍼픽셀 생성

슈퍼픽셀은 픽셀들의 군집화를 통해 생성되는 하나의 세그먼트를 나타낸다. 이러한 군집화 된 세그먼트를 생성하기 위해 영상에서 각 픽셀과 주변픽셀간의 관계 혹은 유사성을 통해 군집화 여부를 결정한다. 본 논문에서는 군집화 하는 과정에 있어 주변 픽셀을 그림 3에서와 같이 4방향만을 통해 생성한다. 픽셀의 유사성 측정은 방향성이 존재하지 않는다. 다시 말하면 기준이 다르지만 두 픽셀의 관계는 동일한 유사성을 추출하게 된다. 따라서 이러한 중복을 제거하기 위해 그림 3에서와 같이 4방향에 대해서만 유사성을 측정한다.



(그림 3) 픽셀간의 관계성 측정 예시

(Figure 3) Example of pixel relation measurement

그림 3에서와 같이 유사성을 검출하기 위해 수식 (1)을 통해 유사성을 산출한다. 수식 (1)에서 W 는 유사성 측정 결과이다. i 와 j 는 영상의 좌표를 나타내며, R, G, B 는 입력 영상의 색상인 빨강, 초록, 파랑을 나타낸다. 수식 (1)에서는 중심 픽셀과 주변 픽셀간의 차를 통해 두 픽셀간의 유사성을 산출하며, W 값이 작을수록 두 픽셀이 유사하다고 판단한다.

$$W(i,j) = \sqrt{(R(i)-R(j))^2 + (G(i)-G(j))^2 + (B(i)-B(j))^2} \quad (1)$$

각 픽셀간의 유사성을 추출하였으면, 다음으로는 유사성을 이용한 픽셀들의 군집화하는 알고리즘을 수행한다. 슈퍼픽셀의 군집화 알고리즘에 있어 가장 중요한 것은 각 세그먼트의 최대 크기를 설정할 수 있다. 수식 (2)는 세그먼트의 최대 크기인 임계값 업데이트 수식이며, 이 임계값을 통해 수식 (1)에서의 W 와 비교하여 픽셀간의 군집 여부를 결정하게 된다.

$$T(i) = \begin{cases} T(i) > W(i,j) \wedge T(j) > W(i,j), & T(i) - W(i,j) \\ otherwise, & T(i) \end{cases} \quad (2)$$

수식 (2)에서 T 는 임계값을 나타낸다. 비교는 i 와 j 위치 두 군데에서 하는데 임계값 업데이트를 $T(i)$ 만 하는 이유는 두 조건이 만족할 경우 i 와 j 픽셀은 하나의 군집으로 할당하기 때문에 이후 비교 시 $T(j)$ 는 사용되지 않고, $T(i)$ 를 통해 비교하게 된다.

표 1은 슈퍼픽셀 생성 알고리즘의 의사코드를 나타낸다. 슈퍼픽셀은 각 픽셀간의 유사성을 추출한 후, 이 유사성을 이용하여 군집화를 진행한다.

(표 1) 슈퍼픽셀 의사코드

(Table 1) Super-pixel pseudo-code

```

for( int x = 0; x < width; x++ )
for( int y = 0; y < height; y++ ) {
    수식 (1)을 사용한  $W(x,y),(x+1,y)$  산출
    수식 (1)을 사용한  $W(x,y),(x,y+1)$  산출
    수식 (1)을 사용한  $W(x,y),(x-1,y+1)$  산출
    수식 (1)을 사용한  $W(x,y),(x+1,y+1)$  산출
}
임계치 맵 생성 및 최대 임계치 설정
for( int x = 0; x < width; x++ )
for( int y = 0; y < height; y++ ) {
    수식 (2)를 사용한  $(x,y)$ 와  $(x+1,y)$ 간의 군집화
    수식 (2)를 사용한  $(x,y)$ 와  $(x,y+1)$ 간의 군집화
    수식 (2)를 사용한  $(x,y)$ 와  $(x-1,y+1)$ 간의 군집화
    수식 (2)를 사용한  $(x,y)$ 와  $(x+1,y+1)$ 간의 군집화
}
    
```

2.2 슈퍼픽셀 특징

슈퍼픽셀들은 각기 다른 픽셀수를 가지고 있으며, 슈퍼픽셀마다 크기도 다르기 때문에, 이 자체를 DNN에 접목시키기는 어렵다. 따라서 각 슈퍼픽셀에 대해서 특징을 추출해야한다. 표 2는 각 슈퍼픽셀들로부터 26개의 특징 추출에 대한 방법을 나타낸다.

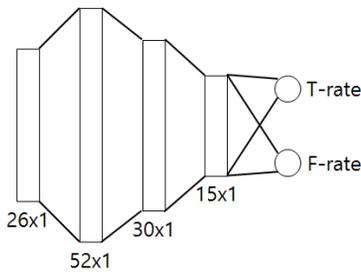
(표 2) 슈퍼 픽셀을 이용한 특징 추출

(Table 2) Feature extraction using Super-pixel

특징	형태	개수
슈퍼픽셀의 상대적 크기	위치	3
슈퍼픽셀의 평균 위치	위치	2
슈퍼픽셀에서 RGB 평균값	색상	3
슈퍼픽셀에서 HSI 평균값	색상	3
이미지의 RGB 평균값	색상	3
이미지의 HSI 평균값	색상	3
이미지와 슈퍼픽셀의 RGB 차이	색상	3
이미지와 슈퍼픽셀의 HSI 차이	색상	3
슈퍼픽셀이 포함하는 에지 개수	에지	1
에지의 평균 위치	에지	1
그라디언트 에지에서 최대민값	에지	1

3. DNN을 이용한 실내 공간 인식

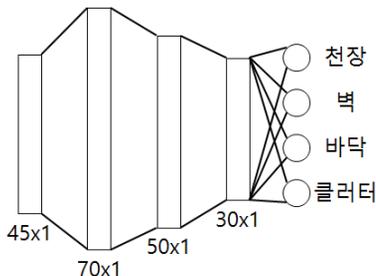
인식을 위한 학습 알고리즘으로는 다양한 방법들이 존재한다. 특히 배경 인식이나 실내 공간 인식을 위해 사용된 방법으로는 Boosted Forest [8] 알고리즘이 기존에 많이 연구되어 왔다. 하지만, 실내 공간을 위한 Boosted Forest는 알고리즘을 개선해도 만족스러운 결과를 얻지 못한다. [8] 이러한 한계점을 해결하기 위해 본 논문에서는 DNN(Deep Neural Network)를 활용하여 인식률을 개선하는 방법에 대해 제안한다.



(그림 4) 하나의 인식을 위한 DNN 구조

(Figure 4) Example of pixel relation measurement

그림 4에서는 본 논문에서 제안하는 DNN의 구성도를 나타낸다. 가장 왼쪽 레이어는 입력 레이어로서 슈퍼픽셀로부터 추출된 특징을 입력으로 사용된다. 가장 오른쪽 레이어는 출력 레이어로 위쪽 노드가 참일 확률을, 아래쪽 노드가 거짓일 확률을 나타낸다. 입력이 천장의 슈퍼픽셀이고, 천장 인식을 위한 DNN일 경우, 출력은 위쪽 노드는 1, 아래쪽 노드는 0으로 할당하여 학습을 수행한다. 학습을 위해 사용된 환경으로 Optimizer는 Adam (Adaptive Moment Estimation)을 사용하였으며, Drop out 값은 0.9로 할당하였다.



(그림 5) 천장, 벽, 바닥 구분을 위한 DNN 구조

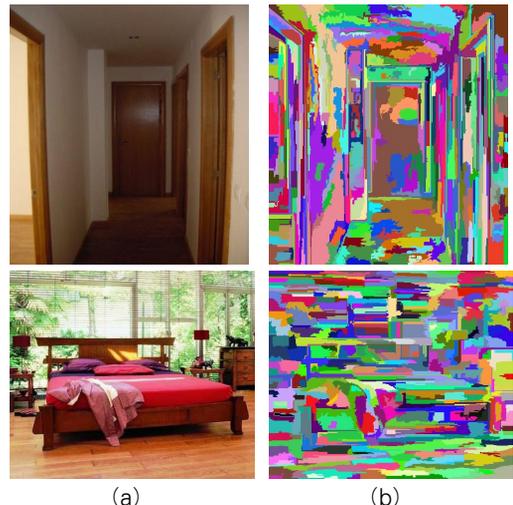
(Figure 5) DNN structure for classify ceiling, wall, floor

그림 5에서는 그림 4에서 추출된 값을 입력값으로 사용한다. 입력으로는 그림 4에서 가장 오른쪽에서 하나 전인 레이어를 입력으로 사용된다. 그림 5는 천장, 바닥, 벽으로 구분되기 때문에, 45개의 노드가 입력 레이어로 사용된다. 출력으로는 4개의 노드로 구분된다. 가장 위쪽으로부터 아래 순서로 천장, 벽, 바닥, 그 외로 구분된다. 그 외라고 할당되는 부분은 천장, 바닥, 벽이 아닌 기타 슈퍼픽셀로 소파, 테이블 등 객체가 이 노드로 표현된다.

4. 실험 결과

본 논문에서 제안한 방법을 개발 및 실험하기 위한 환경은 Intel(R) CoreTM i7-4790 3.60Ghz CPU와 16GByte 메모리를 사용하였으며, 그래픽 카드는 nVidia Geforce GTX 1060 6GB를 사용하였다. 운영체제는 Windows 8 Enterprise K 64bit를 사용하였으며, 개발 툴은 Visual Studio 2013의 MFC, OpenCV 2.4.13, Python 3.5.2, 그리고 Tensorflow 1.4를 사용하였다. 실험 데이터로는 직접 촬영한 것과 인터넷에서 수집한 영상들로 구성되었으며, 학습용 데이터로는 429장의 영상으로 구성되었으며, 실험용 데이터로는 386장의 영상을 준비하였다.

그림 6에서 좌측은 입력영상을 나타내며, 우측은 2장에서 설명한 슈퍼픽셀 추출 알고리즘을 적용 후 세그먼트로 나누어진 영역에 대한 결과를 나타낸다. 슈퍼픽셀에서 최대 임계값은 100으로 사용하였다.



(a) (b)

(그림 6) 슈퍼픽셀 추출 결과

(Figure 6) Result of Super-pixel Extraction

그림 7에서는 원하는 영역의 분할과 슈퍼픽셀로 구분할 경우, 제대로 분류 가능 여부에 대해 비교한 결과이다. 매칭해본 바, 약간의 오차가 존재한다. 이러한 오차는 그림 7서와 같이 클러스터로 인해 가려져서 구분할 수 없는 부분이 생기는 문제가 있다. 하지만, 이러한 부분은 자연스러운 문제로서 벽면을 인식하는데 있어 크게 문제가 되지 않은 것으로 판단된다.



(그림 7) 슈퍼픽셀과 경계선 비교

(Figure 7) Comparison between boundary line and super-pixel

표 3~5는 기존의 Boosted Forest를 통해 벽면을 인식 했을 경우의 인식률과 제안하는 방법을 이용한 인식률을 나타낸다. 인식률은 TPR(True Positive Rate), FPR(False Positive Rate), TNR(True Negative Rate), FNR(False Negative Rate)로 구분하였다. 표 3은 천장을, 표 4는 벽면을, 표 5는 바닥의 인식률을 나타낸다. 기존의 방법에 비해 특히 부정데이터의 인식률이 증대된 것을 확인할 수 있다.

(표 3) 천장 검출 비교

(Table 3) Comparison of Ceiling detection

	Boosted Forest[8]	제안하는 방법
TPR	90.67%	93.26%
FPR	9.33%	6.73%
TNR	72.53%	88.08%
FNR	27.47%	11.92%

(표 4) 벽면 검출 비교

(Table 4) Comparison of Wall detection

	Boosted Forest[8]	제안하는 방법
TPR	89.63%	91.71%
FPR	10.37%	8.29%
TNR	77.72%	90.16%
FNR	22.28%	9.94%

(표 5) 바닥 검출 비교

(Table 5) Comparison of Floor detection

	Boosted Forest[8]	제안하는 방법
TPR	81.34%	86.52%
FPR	18.67%	13.58%
TNR	82.38%	89.64%
FNR	17.62%	10.36%

그림 8은 영상 인식을 했을 때 비교 결과를 나타낸다. 검정색 부분은 바르게 구분한 영역이며, 하얀 부분은 잘못 구분한 영역이다. 오류가 나는 영역은 조금 다르지만, 인식률 측면에서 제안하는 방법이 기존의 방법에 비해 높은 것을 확인 할 수 있다. 해당 영상 뿐만 아니라 전체 테스트 영상을 통해 실험해 본 바, 기존의 방법은 72.10%의 인식률을 보이지만, 제안하는 방법은 81.92%의 인식률을 보였다.



(그림 8) Boosted Forest와 제안하는 방법의 오차 비교
(Figure 8) Comparison difference between Boosted Forest and suggested method

5. 결 론

본 논문에서는 실내 공간 인식을 위해 DNN 학습을 이용하여 벽면 인식하는 알고리즘에 대해 소개하였다. 벽면 인식을 위해 우선 슈퍼픽셀 알고리즘을 통해 영상에서 세그먼트들을 추출하였고, 그 추출된 각 세그먼트마다 특징을 추출하였다. 이러한 특징들을 입력으로 사용한 DNN 모델을 제안하였으며, 이를 통해 각 세그먼트를 인식하여 실내 공간 인식 모델을 제안하였다. 실험 결과 기존의 알고리즘과 비교하여 성능이 향상되었음을 확인할 수 있었다.

본 논문은 세그먼트들을 이용한 벽면 인식 방법으로 다른 실내 공간 인식과 융합 될 것으로 예상된다. 예를 들어 소실점, 코너점 등 레이아웃을 통한 인식 방법과 융합할 경우 더욱 향상된 실내 공간 인식을 할 수 있을 것으로 생각된다.

참고문헌(Reference)

- [1] Techemergence, Everyday Examples of Artificial Intelligence and Machine Learning, 2017.
<https://www.techemergence.com/everyday-examples-of-ai/>
- [2] Structure, 3D Scanner. <https://structure.io/>
- [3] IKEA, IKEA Place.
<http://www.ikea.com/au/en/apps/IKEAPlace.html>
- [4] Microsoft, HoloLens.
<https://www.microsoft.com/en-us/hololens>
- [5] S. Liu, M. Atia, T. Karamat and A. Noureldin, "A LiDAR-Aided Indoor Navigation System for UGVs", The Journal of Navigation, Vol.68, No.2, pp.253-273, 2015.
<https://doi.org/10.1017/S037346331400054X>
- [6] D. Lee, M. Hebert and T. Kanade, "Geometric reasoning for single image structure recovery", Computer Vision and Pattern Recognition, pp.2136-2143, 2009.
<https://doi.org/10.1109/CVPR.2009.5206872>
- [7] A. Gupta, M. Hebert, T. Kanade and D. Blei, "Estimating spatial layout of rooms using volumetric reasoning about objects and surfaces", Advances in neural information processing systems, pp.1288-1296, 2010.
<http://papers.nips.cc/paper/4120-estimating-spatial-layout-of-rooms-using-volumetric-reasoning-about-object-s-and-surfaces>
- [8] K. Kim and H. Choi, "Ensemble of fuzzy decision tree for efficient indoor space recognition", The Korean Society Of Computer And Information, Vol.22, No.4, pp.33-39, 2017.
<https://doi.org/10.9708/jksci.2017.22.04.033>
- [9] X. Yang, X. Gao, D. Tao, X. Li and J. Li, "An Efficient MRF Embedded Level Set Method for Image Segmentation", IEEE Transactions on Image Processing, Vol.24, No.1, pp.9-21, 2015.
<https://doi.org/10.1109/TIP.2014.2372615>
- [10] C. Carson, S. Belongie, H. Greenspan and J. Malik, "Blobworld: image segmentation using expectation-maximization and its application to image querying", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.24, No.8, pp.1026-1038, 2002.
<https://doi.org/10.1109/TPAMI.2002.1023800>
- [11] M. Zheng and Z. Zhou, "A k-nearest neighbor based algorithm for multi-label classification", Granular Computing, IEEE International Conference on, Vol.2, pp.718-721, 2005.
<https://doi.org/10.1109/GRC.2005.1547385>
- [12] D. Hoiem, A. Efros and M. Hebert, "Recovering surface layout from an image", International Journal of Computer Vision, Vol.75, No.1, pp.151-172, 2007.

◎ 저 자 소 개 ◎

**김기상(Kisang Kim)**

2007년 숭실대학교 컴퓨터학부 공학사
 2009년 숭실대학교 대학원 미디어학과 공학석사
 2017년 숭실대학교 대학원 미디어학과 공학박사
 2017년~현재 숭실대학교 산학연구원
 관심분야 : 컴퓨터비전, 머신러닝, 영상처리, 인공지능
 E-mail : kimkisang@ssu.ac.kr

**최형일(Hyung-Il Choi)**

1979년 연세대학교 전자공학과 공학사
 1983년 미시간대학 전기전산학과 공학석사
 1987년 미시간대학 전기전산학과 공학박사
 1989년~1999년 숭실대학교 컴퓨터학부 교수
 2000년~현재 숭실대학교 미디어학과 교수
 관심분야 : 컴퓨터비전, 패턴인식, 증강현실, 영상처리 등
 E-mail : hic@ssu.ac.kr