

스테레오 영상 기반의 객체 탐지 및 객체의 3차원 위치 추정

손행선*, 이선영*, 민경원*, 서성진**

Object Detection and 3D Position Estimation based on Stereo Vision

Haengseon Son*, Seonyoung Lee*, Kyoungwon Min*, Seongjin Seo**

요약 본 항공기에 스테레오 카메라를 장착하여 영상 기반의 비행 객체 탐지 및 탐지된 객체의 3차원 위치를 추정하는 방법을 제시하였다. 구름 사이에 존재할 수 있는 원거리의 작은 객체를 탐지하기 위한 방법으로 PCT 기반의 Saliency Map을 생성하여 이용하였으며, 이렇게 탐지된 객체는 좌우 스테레오 영상에서 매칭을 수행하여 스테레오 시차(Disparity)를 추출하였다. 정확한 Disparity를 추출하기 위하여 비용집적(Cost Aggregation) 영역을 탐지 객체에 맞추어 가변되도록 가변 영역으로 사용하였으며, 본 논문에서는 Saliency Map에서 객체의 존재 영역으로 검출된 결과를 사용하였다. 좀 더 정밀한 Disparity를 추출하기 위하여 Sub-pixel interpolation 기법을 사용하여 Sub-pixel 레벨의 실수형 Disparity를 추출하였다. 또한 이에 카메라 파라미터를 적용하여 실제 탐지된 비행 객체의 3차원 공간 좌표를 생성하여 객체의 공간위치를 추정하는 방법을 제시하였다. 이는 향후 자율비행체의 영상기반 객체 탐지 및 충돌방지 시스템에 활용될 수 있을 것으로 기대된다.

Abstract We introduced a stereo camera on the aircraft to detect flight objects and to estimate the 3D position of them. The Saliency map algorithm based on PCT was proposed to detect a small object between clouds, and then we processed a stereo matching algorithm to find out the disparity between the left and right camera. In order to extract accurate disparity, cost aggregation region was used as a variable region to adapt to detection object. In this paper, we use the detection result as the cost aggregation region. In order to extract more precise disparity, sub-pixel interpolation is used to extract float type-disparity at sub-pixel level. We also proposed a method to estimate the spatial position of an object by using camera parameters. It is expected that it can be applied to image - based object detection and collision avoidance system of autonomous aircraft in the future.

Key Words : Flying Object Detection, Stereo Vision, Saliency Map, PCT, Collision Avoidance

1. 서론

자율비행체란 사람의 조종이나 개입 없이 수송, 이동 등의 임무를 수행하는 비행체를 의미한다. 그중 무인항공기는 처음 군용으로 개발되기 시작하여 최근 그 기술이 고도화되고 지능화 되는 등 비약적인 발전을 해 왔다. 무인 드론 등은 미래의 다양한 분야에서 활용될 것으로 기대되고 있으며, 국

방 및 감시 정찰, 응급환자 구조, 재난 구조 등 다양한 분야에서 무인기의 활용이 추진되고 있다. 현재의 무인비행체처럼 사람의 조종을 받는다면 그 활용부분이 많이 제한적일 것이다. 앞으로 미래는 자율이동 기술이 항공분야에도 적용될 것으로 판단되며 향후 자율비행 기술이 비약적으로 발전하게 되면 향후 사람의 이동 수단이 될 것이라는 전망이 지배적이다. 이러한 자율비행이 가능하게 하려면

This work was supported by the Technology Innovation Program(10059035) funded By the Ministry of Trade, Industry & Energy(MOTIE, Korea)

*Korea Electronics Technology Institute

**Corresponding Author : Hyundai MOBIS Technical R&D center (sjseo@mobis.co.kr)

Received July 21, 2017

Revised July 30, 2017

Accepted August 04, 2017

충돌방지를 위한 센싱 기술이 선행적으로 개발되어야 한다. 다양한 센서들이 서로 융합하여 공간을 센싱하는 기술들이 개발되고 있으며, 카메라를 이용한 영상 기반의 센싱 시스템은 초음파 센서(수십 m) 대비 원거리(수 km)까지 탐지가 가능하며, 특히 물체 식별 기능을 함께 구현할 수 있다. 또한 라이다(Lidar)센서에 비해 가격이 매우 저렴하며, 공간 해상도가 높다는 특징을 가지고 있다.

본 논문에서는 비행공간에서의 다른 비행체와의 충돌을 회피하기 위하여, 스테레오 영상을 기반으로 객체를 탐지하고 객체의 3차원 위치를 추정하는 기술에 대하여 제안하고자 한다. 2장에서는 카메라 영상을 기반으로 한 객체의 검출에 대해서 설명하였으며, 3장에서는 객체의 3차원 위치를 추정하기 위하여 스테레오 영상 처리에 대해서 설명하고자 한다. 4장에서는 항공영상 취득 환경과 실험 결과를 제시하였다.

2. 영상 기반의 객체 검출

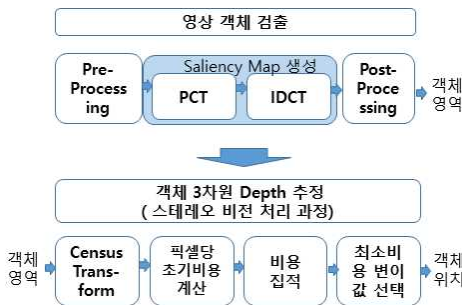


그림 1. 객체 검출 및 위치 추정을 위한 흐름도
Fig. 1. Flow Chart for Object Detection and Position Estimation

2.1 Saliency Map

Saliency Map이란 영상처리 알고리즘을 통하여 주변 영역에 비하여 시각적으로 돌출되어 두드러지게 보이는 영역을 나타내는 것이며, Saliency Map을 이용하여 영상분할, 압축, 객체인식 등 다양한 응용 분야에서 활용될 수 있다. Saliency라는 용어는 Tsotsos[1]와 Olshausen[2], 그리고 Itti[3]

의 연구에서 사용되기 시작하였다. Saliency Map 추출 알고리즘은 크게 생물학적 모델링을 통한 방법과 순수하게 계산적 방법, 조합을 이용한 방법의 3가지로 구분할 수 있다. 3가지 방법론들 모두 색 정보, 명암, 방향성의 데이터를 이용해 Contrast 정보가 주변 영역에 비해 어떠한 특징과 차이를 가지는지 정도를 이용해 Saliency를 추출한다. 본 논문에서는 DCT(Discrete Cosine Transform)를 이용하는 PCT(Pulse Cosine Transform)[4]알고리즘을 활용하여 Saliency Map을 구하고 이를 기반으로 객체를 검출하였다.

2.2 PCT를 이용한 Saliency Map 생성

본 논문에서는 PCT를 이용하여 Saliency Map을 생성한다. 항공 영상 내에서 일반적으로 객체는 구름과 같이 넓은 영역에 단조롭게 분포되어 있는 영역 내에 작은 객체 형태로 존재하며 이를 검출할 필요가 있다.

입력 영상 M에 대하여, M은 항공영상의 특성상 색상에 대한 특징이 크지 않아 Gray 영상만을 고려하였다. 식(1)을 입력 영상 M의 PCT라 정의한다. PCT는 DCT를 수행하여 주파수 영역으로 변환한 후, 이들의 계수 각각에 대하여 Signum 함수를 수행한 것이다. Signum함수에 의해 DCT계수들의 크기는 무시하고 0보다 큰 양수일 경우 1로, 음수일 경우 -1이 되어 부호(1 and -1)만 남게 된다. PCT를 수행한 후, P를 식(2)와 같이 Inverse DCT를 수행하여 시간 영역으로 다시 변환한 후 식(3)와 같이 가우시안 필터를 적용하여 최종 SM(Saliency Map)을 구한다. 식(2)에서 abs(.)는 절대치를 구하는 함수, 식(3)에서 G(.)는 2차원 가우시안 Low-pass 필터를 의미한다.

$$P = \text{sign}(DCT(M)) \quad (1)$$

$$F = \text{abs}(IDCT(P)) \quad (2)$$

$$SM = G * F^2 \quad (3)$$

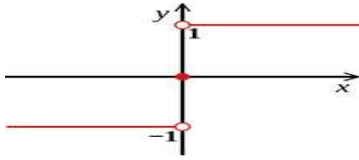


그림 2. 시그널 함수 정의
Fig. 2. Signum Function Definition

2.3 후처리 및 최종 검출

PCT를 이용한 Saliency Map을 기반으로 최종 객체를 잡음환경에서 좀 더 정확히 검출하기 위하여 후처리를 수행한다. Saliency Map 값이 큰 부분에 대하여 이진화를 수행하고, 수행한 이진화 영상에 대하여 Median Filter를 적용하여 노이즈를 제거한다. 이후 모폴로지 필터를 적용한 후 라벨링 알고리즘을 거쳐 최종 객체를 검출한다.



그림 3. PCT를 이용한 Saliency Map 결과
Fig. 3. Saliency Map using PCT

3. 검출 객체의 스테레오 영상 처리

3.1 스테레오 비전

스테레오 비전[5]은 카메라 두 대로 구성되어 있는 스테레오 카메라를 사용하여 거리감을 인식하는 기술을 의미한다. 인간이 양안으로 물체까지의 거리를 인식하는 것과 마찬가지로 스테레오(양안) 카메라를 이용하고 양안카메라에 나타나는 물체의 시차(Disparity) 정보를 추출함으로써 물체까지의 Depth(깊이)를 추정할 수 있다.

양안 스테레오 비전을 이용하여 Depth를 추출하는 원리는 그림 5와 같다. 여기에서 두 개의 카메라는 baseline만큼 떨어진 위치에 서로 나란히 장착되었고 초점 거리가 정확히 같다고 가정한다. 실제적으로 두 카메라의 초점 거리는 다르며, 두

카메라가 정확히 일직선상에 정렬되기 힘들므로, 두 카메라의 체크 보드와 같은 마커를 기반으로 캘리브레이션을 수행하고, 영상을 정렬하는 렉티피케이션(Rectification)을 수행하게 된다.

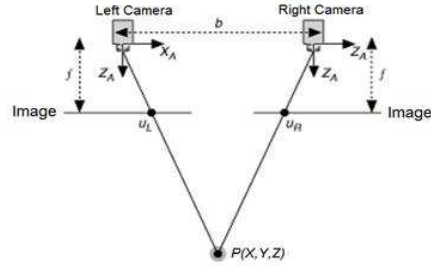


그림 4. 스테레오 비전 시스템
Fig. 4. Stereo Vision System

b(baseline): 두 카메라 사이의 거리

f: 카메라의 초점 거리

P: X, Y, Z로 정의된 실제 물체의 3차원 위치

u_L : 물체 P를 왼쪽 카메라 이미지로 투사점

u_R : 물체 P를 오른쪽 카메라 이미지로 투사점

두 대의 카메라가 "b" 간격만큼 떨어져 있으므로, 두 카메라에 의해 획득된 영상 내에 투사된 P점은 u_L 과 u_R 의 위치에 나타나게 되고 u_L 과 u_R 의 u 좌표는 아래와 같다. 투사된 두 점 사이의 거리를 "disparity"라고 하며 disparity 값을 사용하여 깊이(depth) 정보를 계산할 수 있다. 즉, 실제 점 "P"와 스테레오 비전 시스템 사이의 거리이다.

$$u_L = f * X/Z, \quad u_R = f * (X-b)/Z \quad (4)$$

$$\text{disparity} = u_L - u_R = f * b/Z \quad (5)$$

$$\text{depth} = f * b/\text{disparity} \quad (6)$$

3.2 항공객체의 Depth 추출

3.1에서 설명한 depth를 추출하기 위해서는 좌우 영상 내에서 동일한 물체가 어디에 나타나는지를 찾아내는 픽셀 대 픽셀의 매칭과정을 수행하게 된다. 본 논문에서는 Saliency Map을 이용하여 객

체의 탐지를 선행하여 수행하였고, 최종 검출된 객체에 대하여 스테레오 매칭을 수행하였다. 스테레오 매칭을 위해서는 매칭을 수행하고자 하는 이미지의 좌표를 중심으로 윈도우를 설정하고 특징벡터(Feature Vector)를 추출하여 매칭을 수행하는 것이 일반적인 방법이다.

3.2.1 센서스 변환 (Census Transform)

본 논문에서 사용하는 이미지에 대한 특징벡터는 센서스 변환[6]값을 사용하였다. 센서스 변환은 중심 픽셀과 윈도우 내의 해당 픽셀과의 크기를 비교하여 크기가 크면 1, 그렇지 않으면 0으로 인코딩한다. 이때 윈도우의 크기는 7X5를 사용하였다. 센서스 변환 결과 34bit의 이진데이터 형태로 특징벡터가 추출된다.

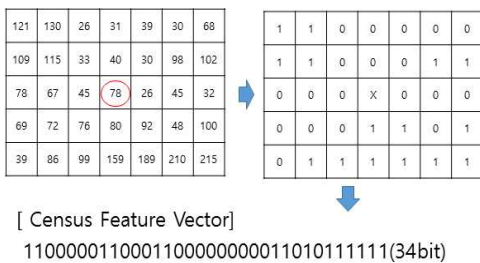


그림 5. 센서스 변환
Fig. 5. Census Transform

3.2.2 비용(Cost) 계산

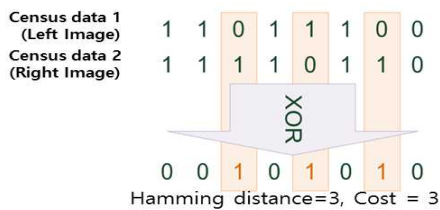


그림 6. 센서스 변환 값의 비용 계산
Fig. 6. Cost Calculation for Census Value

두 개의 센서스 변환 값에 대한 유사도를 측정하는 비용(Cost) 계산은 두 이진 값에 대해 XOR 연산을 수행하고 1의 개수를 세는 Hamming

Distance를 계산한다. 두 개 센서스 값이 같다면 Cost=0 이며, 최대 Cost는 모든 bit가 다른 경우 이므로 34 이다. 영상의 기준 좌표는 스테레오 영상의 좌영상(Left Image)을 기준으로 설정된다. 좌측 영상 픽셀(u, v) 에서 disparity d에 대한 비용은 다음과 같은 식(7)으로 나타낼 수 있다. CL(u, v)는 (u, v)위치에 대한 좌영상의 센서스변환값, CR(u-d, v)는 우영상에 대한 센서스 변환 값을 의미한다. disparity d에 대한 Cost는 이 두 값에 대한 HD(Hamming Distance)로 정의할 수 있다.

$$C(u, v, d) = HD(CL(u, v), CR(u - d, v)) \quad (7)$$

3.2.3 비용 집적 (Cost Aggregation)

좀 더 정확한 스테레오 매칭을 수행하기 위해서 해당 픽셀의 Cost만을 고려하는 것이 아니라 주변 픽셀의 Cost를 누적하는 비용 집적 과정을 거친다. 같은 물체를 이루는 픽셀들은 같은 Disparity 값을 가질 가능성이 크므로, 같은 물체가 존재할 것으로 판단되는 영역 내에서 비용을 집적함으로써 정확한 스테레오 매칭 결과를 얻을 수 있다. 본 논문에서는 이러한 비용 집적 영역을 Saliency Map에서 검출된 객체 영역(A)으로 정의하여 사용한다. 따라서 객체의 검출 결과에 따라 비용 집적 영역의 모양과 크기가 달라진다. 식(7)과 같이 픽셀(u, v) 에 대한 Disparity d에 대한 Cost는 식(8)과 같이 객체검출영역 A내 있는 Cost를 모두 합산하여 Disparity d에 대한 함수로 나타낼 수 있다. 본 논문에서는 Disparity d는 0부터 63까지 모두 64개의 레벨을 고려하였다.

$$C_{object}(d) = \sum_{u, v \in A} C(u, v, d) \quad (8)$$

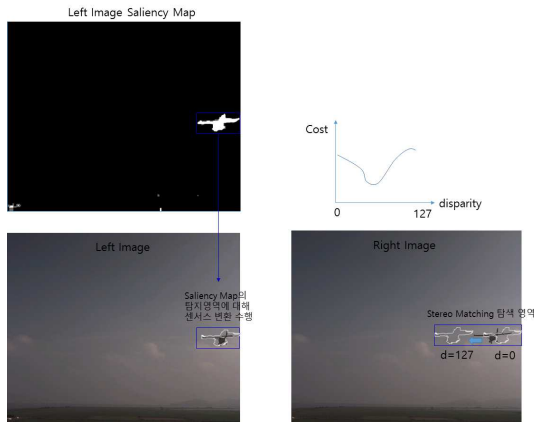


그림 7 Cost Aggregation 과정 및 최종 Disparity
Fig. 7. Cost Aggregation Process and Final Disparity

3.2.4 최적 변위값 및 Sub-Pixel Interpolation

최적의 변위값(Disparity)은 최소의 C_{object} 를 가지는 변위값 d_{min} 으로 결정할 수 있다. 여기에서 변위값의 범위는 0~63까지 정수값이나 좀 더 정밀한 깊이값 추출을 위하여 실수형의 변위값을 Sub-Pixel Interpolation[7] 기법으로 추정하였다.

최적 변위값 $d_0 = d_{min}$, 좌우 변위값 $d_{-1} = d_{min} - 1$ 과 $d_1 = d_{min} + 1$ 의 비용을 각각 C_{-1} , C_0 , C_1 이라 명명하고 이 세 개의 점을 2차원 함수로 Interpolation을 수행한 후, 최소의 Cost를 가지는 지점을 찾고 이때의 변위값 d_f 를 실수형으로 다음과 같이 정밀하게 구한다.

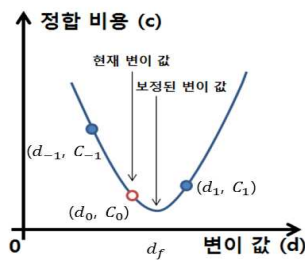


그림 8. 부화소 보간법
Fig. 8. Sub-Pixel Interpolation

$$d_f = d_{min} + \frac{C_{-1} - C_1}{-2(C_1 + C_{-1} - 2C_0)} \quad (9)$$

3.2.5 최종 탐지 객체의 3차원 위치 추정

영상 내에 탐지 객체의 중심좌표 (u, v) 와 변이값 d를 알면, 아래와 같이 객체의 3차원 위치(거리, 방위각, 고각)는 아래와 같이 계산 된다. 이때 c_u, c_v 는 카메라의 주점의 좌표이며, f는 초점거리 b는 스테레오 카메라의 양안의 거리(baseline), d는 변이값(disparity)이다. HFOV는 카메라 센서의 수평화각, VFOV는 카메라 센서의 수직화각이며 R_H 는 카메라의 수평 해상도, R_V 는 수직해상도이다.

$$Z_c(\text{거리}) = f \frac{b}{d}$$

$$\theta_c(\text{방위각}) = (u - c_u) \frac{HFOV}{R_H}$$

$$\phi_c(\text{고각}) = (c_v - v) \frac{VFOV}{R_V} \quad (10)$$

4. 실험결과

비행 객체를 탐지하기 위하여 아래와 같이 스테레오 비전 시스템을 구축하여 항공 영상을 촬영하였다. 스테레오 비전의 카메라 양안의 길이 Baseline은 30cm 로 제작하였다. 자율비행에 적용하기 위해서는 항공기에 스테레오 카메라를 장착하여 영상을 취득하여야 하나, 1차적으로 다양한 표적비행체의 시나리오를 구현하고 이에 대한 영상을 확보하기 위하여 지상에 스테레오 카메라 지그를 설치하여 취득하였다. 표적비행체는 아래와 같은 시나리오로 비행을 하였으며 이에 대한 스테레오 영상 시퀀스를 확보하여 실험을 수행하였다. 그림 10과 같이 취득 영상에 대하여 표적기 탐지를 성공적으로 수행할 수 있었고, 표적기의 3차원 위치 정보를 추출할 수 있었다. 알고리즘 처리 성능은 Intel i7-6700K CPU, 64GB 메모리 환경에서 1280x1024 스테레오 영상을 15 Frame/s 이다. 그림10는 PCT 의 최종 Saliency Map, Left 영상에서의 객체탐지 결과와 Right 영상에서 스테레오 비전 매칭 결과를 보여주고 있다.



그림 9. 스테레오 항공 영상 수집 장치 및 표적기
 Fig. 9. Stereo Image Acquisition for Target object
 : image resolution: 1280x1024, Focal length: 8mm

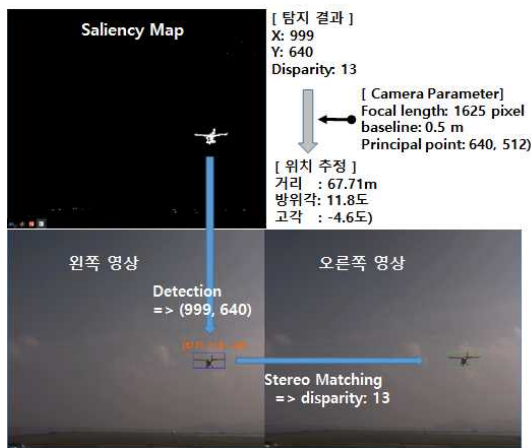


그림 10. 최종 탐지 및 위치 추정 결과
 Fig 10. Final Detection and Position Estimation

5. 결론

본 논문에서는 항공 영상 내에서 침입 객체를 인식하고 이의 위치를 추적하는 알고리즘을 제시하였다. 구름과 같은 단조로운 환경 내에서 작은 침입기를 탐지하기 위하여 PCT기반의 Saliency Map을 사용하여 탐지를 수행하였으며, 탐지된 객체에 대하여 좌우 스테레오 영상에서 Disparity 정보를 추출하고 이를 기반으로 3차원 위치를 추정하였다. 초점거리 8mm 렌즈를 사용하고 스테레오 Baseline 30cm로 구현하였을 때 유효 인식거리는 약 300m 이내였으며, 알고리즘 처리성능은 약 15 FPS를 달성하여, 최적화를 시킨다면 임베디드 환경에서도 실시간 처리 가능성이 높을 것으로 판단된다.

REFERENCES

- [1] Tsotsos J. K., Culhane S. M., Wai W. Y. K, Lai Y, Davis N. and Nuflo F.(1995), "Modeling visual attention via selective tuning," Artificial Intelligence, vol. 78, pp.507-545.
- [2] Olshausen B. A., Anderson C. H. and Van Essen D. C.(1993), "A Neurobiological Model of Visual Attention and Invariant Pattern Recognition Based on Dynamic Routing of Information," The Journal of Neuroscience, vol. 13, no. 11, pp.4700-4719.
- [3] Itti L., Koch C. and Niebur E.(1998), "A Model of Saliency-based Visual Attention for Rapid Scene Analysis," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 20, no. 11, pp.1254-1259.
- [4] Yu, Ying, Bin Wang, and Liming Zhang. "Pulse discrete cosine transform for saliency-based visual attention." Development and Learning, 2009. ICDL 2009. IEEE 8th International Conference on. IEEE, 2009.
- [5] Banz, Christian, et al. "Real-time stereo vision system using semi-global matching disparity estimation: Architecture and FPGA-implementation." Embedded Computer Systems (SAMOS), 2010 International Conference on. IEEE, 2010.
- [6] Froba, Bernhard, and Andreas Ernst. "Face detection with the modified census transform." Automatic Face and Gesture Recognition, 2004. Proceedings. Sixth IEEE International Conference on. IEEE, 2004.
- [7] Kanade, Takeo, et al. "A stereo machine for video-rate dense depth mapping and its new applications." Computer Vision

and Pattern Recognition, 1996. Proceedings CVPR'96, 1996 IEEE Computer Society Conference on. IEEE, 1996.

저자약력

손 행 선(Haengseon Son) [정회원]



- 2000년 2월 : 한양대학교 전자전기공학부 (전자공학 학사)
 - 2002년 2월 : 포항공대 전자전기공학과(전자공학 석사)
 - 2002년 1월~2005년 11월: 국방과학연구소 연구원
 - 2005년 12월 ~ 현재 : 전자부품연구원 선임연구원
- 영상탐지인식, 자율이동, 임베디드 시스템

<관심분야>

이 선 영(Seonyoung Lee) [정회원]



- 2000년 2월 : 한국외국어대학교 전자공학과 (전자공학 석사)
- 2009년 8월 : 한국외국어대학교 전자공학과 (전자공학 박사)
- 2009년 1월 ~ 현재 : 전자부품연구원 책임연구원

<관심분야>

자율주행, 임베디드 시스템

민 경 원(Kyoungwon Min) [정회원]



- 1994년 2월 : 홍익대학교 전자공학과 (전자공학 학사)
- 1996년 8월 : 홍익대학교 전자공학과 (전자공학 석사)
- 1996년 9월 ~ 현재 : 전자부품연구원 수석연구원

<관심분야>

자율주행, 임베디드 시스템

서 성 진(Sungjin Seo) [정회원]



- 2004년 8월 : 광운대학교 대학원 제어계측공학과 (공학석사)
- 2004년 2월 ~ 2006년 3월 : 한단정보통신 연구원
- 2006년 4월 ~ 현재 : 현대모비스 기술연구소 책임연구원/그룹장

<관심분야>

정보통신, 센서 및 자율주행