

다수의 평면을 가지는 모델기반 카메라 추적방법*

이인표, 남보담, 홍현기
 중앙대학교 첨단영상대학원 영상학과
 {inpyo, spaurh}@wmcau.ac.kr, honghk@cau.ac.kr

Camera Tracking Method based on Model with Multiple Planes

Inpyo Lee, Bodam Nam, Hyunki Hong
 Dept. of Imaging Science and Arts, GSAIM, Chung-Ang University

요 약

본 논문에서는 다수의 평면을 가지는 모델기반 카메라 추적 시스템이 제안된다. 상품의 정보를 표기하기 위한 2차원 바코드(barcode)로 널리 사용되는 QR(Quick Response) 코드를 인식하여 해당 물체의 3차원 모델을 임포트한다. 그리고 관련 기하정보를 이용하여 모델의 주요 정점(vertex)을 추출하고 옵티컬 플로우(optical flow)를 이용하여 추적한다. 클리핑 알고리즘으로 다수의 평면을 가지는 물체의 평면 영역을 구별하고 매칭된 특징으로부터 호모그래피를 계산하여 초기 단계의 대략적인 카메라 움직임 파라미터를 추정한다. 이후 카메라의 움직임에 따라 다양한 평면에 존재하는 특징점과 해당 3차원 정보를 선형 방정식으로 구성하고 DLT(Direct Linear Transform) 방법을 적용한다. 최종 단계에서 번들 조정(Bundle Adjustment) 알고리즘을 이용해 카메라의 움직임 파라미터에 포함된 에러를 최소화 한다.

ABSTRACT

This paper presents a novel camera tracking method based on model with multiple planes. The proposed algorithm detects QR code that is one of the most popular types of two-dimensional barcodes. A 3D model is imported from the detected QR code for augmented reality application. Based on the geometric property of the model, the vertices are detected and tracked using optical flow. A clipping algorithm is applied to identify each plane from model surfaces. The proposed method estimates the homography from coplanar feature correspondences, which is used to obtain the initial camera motion parameters. After deriving a linear equation from many feature points on the model and their 3D information, we employ DLT(Direct Linear Transform) to compute camera information. In the final step, the error of camera poses in every frame are minimized with local Bundle Adjustment algorithm in real-time.

Keywords : Augmented Reality, QR code, Camera Tracking, Clipping Algorithm, Direct Linear Transform

접수일자 : 2011년 05월 23일 일차수정 : 2011년 06월 22일 심사완료 : 2011년 07월 04일

교신저자(Corresponding Author) : 홍현기

* 본 논문은 교육과학기술부 재원으로 학술연구재단의 지원을 받아 수행된 2단계 두뇌한국(BK)21 사업 및 기초연구사업(2010-0010398)입니다.

1. 서 론

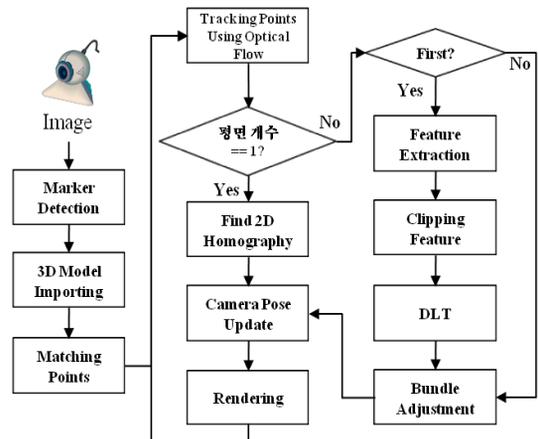
증강현실(Augmented Reality) 분야에서 카메라의 움직임 정보를 실시간에 추정하는 카메라 추적(tracking) 알고리즘은 매우 중요한 연구 주제이다. 기존의 카메라 추적 방법은 크게 마커(marker) 추적 방식과 비마커 추적 방식으로 구분된다.

마커는 카메라로부터 입력되는 실제 공간과 가상의 공간을 연결하기 위한 공간 분석 과정에서 널리 사용되고 있다. 최근에는 바코드 리더와 같은 단말기 사용의 증가에 따라 보다 많은 정보를 포함하면서도 정확하게 인식하기 위한 방법으로 QR(Quick Response) 코드가 표준화되었으며, 이를 이용한 증강현실 연구가 제안되었다[1]. 비마커 추적 방식은 카메라로 관찰되는 장면의 특징점 등을 분석하여 카메라의 움직임을 해석하는 방식으로 다양한 연구가 진행되어 실시간 증강현실 분야에서 좋은 결과를 발표하였다[2,3].

실제 공간에 합성된 가상의 물체는 카메라의 움직임에 따라 동일하게 영향을 받아야 하며, 이 과정에서 카메라의 움직임을 정확하게 추정해야 한다. 이를 위한 공간 분석 알고리즘에는 평면 위에 존재하는 특징을 이용하는 2차원 호모그래피(homography)[4], 3차원 점 클라우드(cloud) 정보를 활용한 선형(linear) 추정 알고리즘[5], 그리고 카메라 움직임을 보다 정확하게 추정하기 위한 비선형(non-linear) 추정 알고리즘[6] 등이 있다.

증강현실을 적용하기 위해 공간을 분석하는 관련 연구는 모델링되어 있지 않은 실내 공간에 대한 3차원 정보를 추출해 장면 맵을 구성하고 임의의 가상 물체를 합성하는 단계까지 진행되었다[3]. 그러나 알려져 있지 않은 공간을 분석하기 위해서는 많은 계산이 필요하며, 넓은 공간을 해석하기에는 많은 어려움이 존재한다. 이러한 계산 복잡도를 극복하기 위해서는 장면에 대한 많은 제약 조건을 고려하거나 장면에 존재하는 물체 등 구성요소의 3차원 모델, 크기, 구조정보를 이용하는 방법 등이 제안되었다[6,7,8,9].

본 논문에서 제안된 방법은 2차원 바코드인 QR 코드를 해석하고, 각 코드에 따라 설정된 3차원 모델의 임포트(import)를 통해 다수의 평면을 가지고 있는 모델의 중요 모서리의 정점(vertex) 등을 참조하여 중요한 특징점을 추출 및 추적한다. 이로부터 카메라의 움직임 정보를 추정하고, 번들 조정(bundle adjustment)을 통한 최적화 과정을 수행한다. 본 논문에서 제안하는 모델은 적어도 한개 이상의 평면을 포함해야 하며, 실험을 위해 직육면체 형태의 모델을 사용하였다.



[그림 1] 제안된 방법의 흐름도

2. 본 론

2.1 인식된 모델 기반 카메라 추적

본 논문에서 제안된 방법의 흐름도를 [그림 1]에 나타내었다. 먼저 입력영상에서 QR 코드를 인식하고 해당 모델의 3차원 정보 등을 임포트한다. 그리고 해당 모델을 구성하는 정점 정보에 해당하는 특징점을 영상으로부터 검출 및 추적하며, 모델의 각 표면 영역을 구분하고 호모그래피를 추출하여 카메라의 움직임 정보를 추정한다. 그리고 중요 정점과 다양한 평면에 존재하는 영상 내 특징점과 해당 3차원 정보를 선형 방정식으로 구성하고

DLT(Direct Linear Transform) 방법을 적용한다. 최종 단계에서 카메라의 움직임 파라미터를 보다 정확하게 추정하기 위해 비선형 추정 알고리즘인 LM(Levenberg-Marquadt) 기반의 다변수 추정 시스템인 번들 조정을 통해 최종적인 카메라의 위치를 계산한다.

2.2 QR 코드 인식을 통한 모델 임포팅

3차원 모델을 임포팅하기 위한 QR 코드의 인식 알고리즘은 기존 연구[1]를 참조하였다. 즉, QR 코드의 위치 정보를 나타내는 세 개의 모서리 영역(좌상, 우상, 좌하)의 사각형 패턴을 인식하여 해당 코드의 네 개의 모서리 점들을 추정한다. 그리고 추출된 모서리 점들을 기준으로 내부 영역의 패턴 정보로부터 코드를 인식한다. 이후 QR 코드 상의 네 개의 모서리 점을 옵티컬 플로우(optical flow)[10]로 추적하면서 마커가 위치한 해당 평면의 호모그래피를 계산하고 증강현실을 위한 카메라 정보를 얻을 수 있다.

```

Input : QR 코드 영상
Output : QR 코드 모서리 점
1: 사각형 찾기
2: 사각형 중심점 찾기
3: 중심점들 사이 거리 계산
4: if 거리 == MAX then
    QR 코드의 중심점 추정
    각 사각형 모서리 점들 사이 거리 계산
end
5: if 모서리 점들 사이 거리 == MAX then
    QR 코드의 모서리 점
end
7: if 모서리 점들 사이 거리 != MIN then
    네 번째 QR 코드 모서리 점 추출
end
8: return QR 코드 모서리 점
    
```

[그림 2] QR 코드 인식 알고리즘

2.3 DLT를 이용한 선형 카메라 추적

입력되는 영상에서 모델 평면이 한 개가 존재하는 경우에는 2차원 호모그래피를 구하고 이로부터 카메라 정보를 계산한다[2]. 즉, 임포트된 모델을 구성하는 평면 정보와 입력 영상에서 해당 면과의 대응관계로부터 호모그래피를 계산하고 이를 분해하여 카메라 움직임 파라미터를 계산한다.

$$P = K[R \ T] = \begin{bmatrix} f_x & s & u \\ 0 & f_y & v \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \end{bmatrix} \quad (\text{식 1})$$

여기서 사영행렬 P는 카메라의 내부(intrinsic) 파라미터 K, 그리고 회전 R, 이동 T의 외부(extrinsic) 파라미터의 곱으로 표현된다. (식 1)로부터 실제 세계의 평면 상의 한 점과 영상 평면 상의 한 점의 관계를 표현하면 (식 2)와 같다.

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = K \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \end{bmatrix} \begin{bmatrix} X \\ Y \\ 0 \\ 1 \end{bmatrix} = K \begin{bmatrix} r_{11} & r_{12} & t_x \\ r_{21} & r_{22} & t_y \\ r_{31} & r_{32} & t_z \end{bmatrix} \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} \quad (\text{식 2})$$

여기서 $(u, v, 1)$ 은 영상 평면상에서의 한 점, $(X, Y, 0, 1)$ 은 이에 대응하는 실제 세계에서의 동일한 점을 각각 의미한다. 각 대응관계는 동차(homogeneous) 좌표계로 표현되며, 실제 세계에 존재하는 한 점의 Z좌표가 0인 이유는 z값이 0인 평면상에 존재하기 때문이다. 따라서 P는 (식 4)와 같은 호모그래피 행렬로 표현된다.

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} \quad (\text{식 3})$$

(식 3)을 이용해 카메라의 움직임 파라미터를 추출하기 위해 다음의 (식 4)를 이용한다.

$$r_1 = \lambda K^{-1}h_1, r_2 = \lambda K^{-1}h_2, r_3 = r_1 \times r_2, t = \lambda K^{-1}h_3 \quad (\text{식 4})$$

여기서 K 는 정칙행렬로서 역행렬이 존재한다. r_3 는 회전행렬의 직교 성질을 이용해 r_1, r_2 의 외적(cross product)으로 계산될 수 있으며, λ 는 행렬의 크기 인자(scale factor)로 아래의 (식 5)의 방법으로 구한다.

$$\lambda = 1/\|K^{-1}h_1\| = 1/\|K^{-1}h_2\| \quad (\text{식 5})$$

제안된 방법은 초기 단계에서 하나의 평면에 대한 호모그래피를 계산하여 카메라의 움직임 파라미터를 추정할 수 있다. 영상내에 여러 개의 평면이 입력되면, 추적되는 특징점과 3차원 정보를 선형 방정식으로 구성하고 DLT 방법을 적용한다. 이 과정에서 얻어진 카메라 움직임 정보는 이후 보다 안정적으로 카메라 움직임을 추정하기 위한 번들 조정의 초기값으로 사용된다[5,11].

$$Am = 0$$

$$A = \begin{bmatrix} X_1 & Y_1 & Z_1 & 1 & 0 & 0 & 0 & 0 & -x_1X_1 & -x_1Y_1 & -x_1Z_1 & -x_1 \\ 0 & 0 & 0 & 0 & X_1 & Y_1 & Z_1 & 1 & -y_1X_1 & -y_1Y_1 & -y_1Z_1 & -y_1 \\ X_2 & Y_2 & Z_2 & 1 & 0 & 0 & 0 & 0 & -x_2X_2 & -x_2Y_2 & -x_2Z_2 & -x_2 \\ 0 & 0 & 0 & 0 & X_2 & Y_2 & Z_2 & 1 & -y_2X_2 & -y_2Y_2 & -y_2Z_2 & -y_2 \\ \vdots & \vdots \\ X_N & Y_N & Z_N & 1 & 0 & 0 & 0 & 0 & -x_NX_N & -x_NY_N & -x_NZ_N & -x_N \\ 0 & 0 & 0 & 0 & X_N & Y_N & Z_N & 1 & -y_NX_N & -y_NY_N & -y_NZ_N & -y_N \end{bmatrix}$$

$$m = [m_{11}, m_{12}, \dots, m_{33}, m_{34}]^T \quad (\text{식 6})$$

DLT는 영상에서 검출된 특징점과 실제 세계에 존재하는 3차원 점들과의 대응 관계를 통해 카메라의 사영 행렬을 계산하는 방법으로 최소 6개 이상의 대응점이 필요하다. 이 관계를 (식 6)에 나타내었으며, (X_N, Y_N, Z_N) 과 (x_N, y_N) 은 각각 실제 세계에 존재하는 점들과 이에 대응되는 영상평면의 점들을 의미한다. 여기서 m 은 사영행렬로서 행렬 A 에 대하여 SVD(Singular Value Decomposition)를 수행하고, 분해된 고유(eigen) 벡터 행렬 중 최

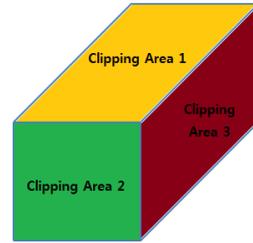
소 고유값에 대응되는 벡터로서 사영 행렬을 추정할 수 있다.

2.4 모델 표면의 클리핑

```

Input : 3차원 모델 정보
Output : 분류된 특징점
1: 특징점 추출
2: 영역 설정
3: if 특징점이 영역 안에 존재 then
    특징점 저장
    if 모델의 특징 면에 존재 then
        2차원 좌표를 3차원으로 변환
        3차원 좌표 저장
    end
end
4: return 대응점 집합
    
```

[그림 3] 3차원 모델에서 클리핑 알고리즘



[그림 4] 3차원 모델의 클리핑 영역

입력된 3차원 모델을 구성하는 중요 모서리 정보는 사전에 강건한 검출 및 추적을 위해 일반적으로 다해상도 정보로 저장된다. 그러나 각 면에 해당하는 표면 텍스처 정보까지 저장하면 모델기반 방법에서 고려해야 하는 인덱싱(indexing) 정보가 너무 많아지며, 사전에 저장된 영상에 비해 입력 조건에서 조명 변화 등의 환경적 영향에 대응관계 설정이 어렵다.

본 논문에서는 모서리에 해당하는 정점을 기준으로 모델 평면을 구분하기 위해 [그림 3]의 클리핑 알고리즘을 적용한다. 즉, 중요 정점으로부터 각 평면상에 존재하는 특징들을 분류하여 호모그래피를 계산한다. 3차원 직육면체의 구분된 각 평면

영역을 [그림 4]에 나타내었다. 각 면의 3차원 구조를 기반으로 해당 특징점들을 3차원 좌표로 변환한 다음, DLT를 적용하여 카메라 정보를 구한다.

2.5 번들 조정

카메라의 움직임은 비선형적 특성을 가지므로 이에 적합한 해석 모델을 적용하는 것이 바람직하다. 본 논문에서는 LM(Levenberg-Marquadt) 알고리즘을 이용하여 카메라 움직임에 대한 비선형 방정식의 해를 구한다. 번들 조정은 움직임 정보를 여러 프레임 동안 누적하여 조정하는 방식으로 3차원 점 X 와 영상내 투영 좌표 x 를 카메라의 투영 행렬 P 를 이용한 함수 $f(P, X)$ 로 구성한다. 그리고 실제로 관찰된 특징점 x 와의 차이를 오차로 놓고 이를 최소화 하는 방식으로 (식 7)과 같다.

$$\min_{(P, X)} \sum_{i=1}^M \sum_{j=1}^N \|x_{ij} - f(P, X_j)\| \quad (\text{식 } 7)$$

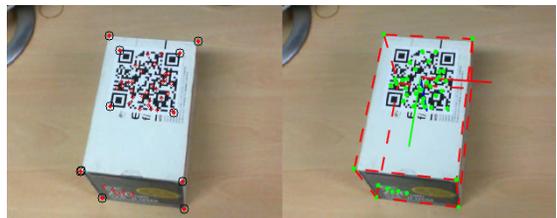
번들 조정은 카메라 정보를 누적하므로 프레임 수와 3차원 좌표점 개수가 많아질수록 복잡도가 증가한다. 예를 들어 듬성한(sparse) 행렬 알고리즘을 사용하는 경우, 계산량은 $\Theta(mn)$ (n : 특징점 수, m : 카메라 프레임 수)이 된다. 본 논문에서는 번들 조정을 실시간 환경에 적용하기 위해서 누적 조건을 검사하여 키 프레임을 해석하는 지역(local) 번들 조정 방식을 사용한다.

3. 실험 및 결과 분석

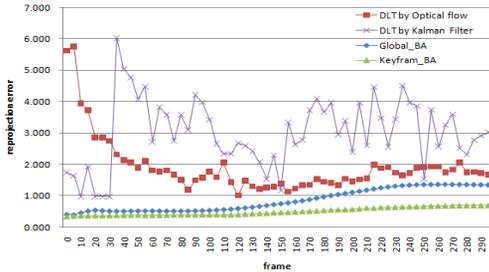
영상 시퀀스를 대상으로 제안된 알고리즘에 의한 모델 인식과 카메라 추적 결과를 [그림 5]에 나타내었다. [그림 5]의 왼쪽 영상에서 ●점들은 검출 및 추적되고 있는 모델의 각 모서리 정점들과 QR 코드의 각 꼭지점들을, ○점들은 클리핑에 의한 모델의 각 평면상의 특징점을 나타낸다. 오른쪽

영상은 제안된 카메라 추적방법을 통해 얻어진 정보로부터 3차원 모델 정보를 화면에 표시한 결과로 임포트된 3차원 모델의 와이어 프레임을 점선으로 표시하였고, 모델의 각 면위에 표시된 점들은 각 평면 영역에 존재하는 특징점이 호모그래피 계산에 따라 변환된 3차원 점들을 의미한다.

DLT, 칼만 필터, 전역 번들 조정, 로컬 키프레임 번들 조정에 따른 재사영 오차 결과를 [그림 6]에 나타내었다. 상대적으로 칼만 필터가 DLT보다 오차가 더 큰 것을 확인하였으며, DLT만 수행하였을 때보다 번들 조정을 함께 사용하는 경우 보다 적은 오차가 발생함을 확인하였다. 지역 키프레임 번들 조정과정에서 카메라가 최소 20mm 이상 이동한 프레임만을 사용했기 때문에 내부적으로 계산되는 자코비안(Jacobian) 행렬 값이 커지고 고유값도 증가하므로 상대적으로 작은 고유값을 가지는 전역(global) 번들 조정보다 정확한 결과를 얻었다. 또한 전역 번들 조정보다 변수의 수가 적어 수행 속도가 더 빠른 장점이 있다. [표 1]의 각 단계별 계산시간에서 가장 많은 시간이 소요되는 QR 코드 인식은 초기에 한번 수행하는 과정이므로 제안된 증강현실 방법이 실시간 응용에 충분히 사용될 수 있음을 확인하였고, 실제 외부 환경에서 촬영된 시퀀스에 적용된 결과 장면을 [그림 8]에 각각 나타내었다.



[그림 5] 제안된 알고리즘의 실험 결과



[그림 6] 제안된 알고리즘의 재사영 에러

[표 1] 각 단계별 수행 시간

| | QR 코드 인식 | 호모 그래피 계산 | 클리핑 | DLT | 번들과정 | |
|------------|----------|-----------|------|------|-------|------|
| | | | | | 키프레임 | 전역 |
| 수행 시간 (ms) | 42.4 | 0.16 | 0.02 | 0.15 | 13.26 | 41.2 |



[그림 8] 영상시퀀스 적용결과-간판(좌), 자판기(우)

4. 결 론

본 논문에서는 제품 정보를 전달하기 위해 널리 사용되는 QR 코드를 인식하여 해당 3차원 모델을 임포팅하고 모델 기반의 실시간 카메라 추적 알고리즘이 제안된다. 다양한 카메라 움직임을 고려하기 위해 2차원 평면 간의 호모그래피, 3차원 대응 관계를 이용하는 DLT, 다수의 프레임을 이용한 비선형 번들 조정 방법(평균 오차 0.499 픽셀) 등을 이용하여 카메라의 움직임 정보를 칼만 필터(평균 오차 2.975 픽셀) 보다 비교적 정확하게 추정하였다. 또한 키프레임 기반의 지역 번들 조정을 적용하여 칼만 필터만 사용하였을 때 보다 강건하고, 전역 번들 조정보다 빠른 수행 성능을 얻었다. 제

안된 알고리즘은 이후 증강현실 게임이나 다양한 모바일 어플리케이션 등에 활용될 예정이다.

참고문헌

- [1] T.W. Kan, C.H. Teng and W.S. Chou, "Applying QR code in augmented reality applications," in Proc. of Int. Conf. on Virtual Reality Continuum and its Applications in Industry, 2009.
- [2] G. Simon, A. W. Fitzgibbon and A. Zisserman, "Markerless tracking using planar structures in the scene," in Proc. Int. Symposium on Augmented Reality, 2000.
- [3] G. Klein and D. Murray, "Parallel tracking and mapping for small AR workspaces," in Proc. of the Conf. on Computer Vision and Augmented Reality, 2007.
- [4] Z. Zhang, "A flexible new technique for camera calibration," IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 1330-1334, 2000.
- [5] E. Trucco and A. Verri, "Introductory Techniques for 3-D Computer Vision". Prentice Hall, 1998.
- [6] K. Kim, V. Lepetit and W. Woo, "Keyframe-based modeling and tracking of multiple 3D objects," Int. Symposium of Mixed and Augmented Reality, 2010.
- [7] L. Vacchetti, V. Lepetit, and P. Fua, "Stable real-time 3d tracking using online and offline information," IEEE Trans. on PAMI, vol. 26, pp. 1385-1391, 2004.
- [8] R. Freeman and A, "Stead, interactive modelling and tracking for mixed and augmented reality," in Proc. of the ACM symposium on Virtual Reality Software and Technology, 2006.
- [9] V. Lepetit, P. Laguerre and P. Fua, "Randomized trees for real-time keypoint recognition," IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, 2005.
- [10] J. Y. Bouguet, "Pyramidal implementation of the lucas-kanade feature tracker description

of the algorithm,” Intel Corporation Microprocessor Research Labs, 2000.

- [11] F. Dornaika and C. Garcia, “Robust camera calibration using 2D to 3D feature correspondences,” Optical Science Engineering and Instrumentation, Videometrics V, Volume 3174, 1999.
- [12] 김기백, 이석한, 최중수, “히스토그램 명세화와 칼만 필터를 이용한 급격한 밝기 변화에 강건한 증강현실 시스템,” 전자공학회논문지, 제 48권 제 2호, 1-148쪽, 2011년 3월



이 인 표 (Lee, Inpyo)

2010년 한국산업기술대학교 게임공학과 졸업
2010년-현재 중앙대학교 첨단영상대학원 영상학과 석사과정

관심분야 : 증강현실, 컴퓨터비전, 컴퓨터그래픽스 등



남 보 담 (Nam, Bodam)

2010년 중앙대학교 컴퓨터공학과 졸업
2010년-현재 중앙대학교 첨단영상대학원 영상학과 석사과정

관심분야 : 컴퓨터비전, 컴퓨터그래픽스 등



홍 현 기 (Hong, Hyunki)

1998년 중앙대학교 전자공학과 박사
1998년-1999년 서울대학교 자동제어 특화연구센터 연구원
2002년-2003년 Post-doc., Univ. of Colorado at Denver
2000년-현재 중앙대학교 첨단영상대학원 영상학과 교수 재직 중.

관심분야 : 증강현실, 컴퓨터비전, 컴퓨터그래픽스 등
