

텍스처가 없는 환경에서 증강현실을 구현하기 위한 색상 히스토그램 지역 서술자

김강수 박정식 서병국 박종일

한양대학교

{vistavision, nangsik, bkseo}@mr.hanyang.ac.kr, jipark@hanyang.ac.kr

Color-Histogram Descriptor for Augmented Reality on Non-Textured Objects

Kim, Kangsoo Park, Jungsik Seo, Byung-Kuk Park, Jong-Il

Hanyang University

요약

물체 인식 및 추적 기술은 기계가 영상 정보를 기반으로 주변을 인지하고 정황을 파악하는 컴퓨터 비전 분야의 매우 중요한 연구 영역 중 하나이다. 현재까지 이러한 물체 인식/추적에 대한 다양한 연구들이 있어 왔고, 최근 증강현실에 대한 높은 관심을 바탕으로 증강현실을 위한 영상 정보 기반의 정확하고 정교한 추적 기술에 대한 관심 또한 매우 높아졌다. 본 논문에서는 텍스처가 없는 단색의 블록에 대해 증강현실을 실현하기 위한 물체 추적 방식을 제안한다. 제안하는 방식은 다수의 블록들을 조합하여 구성하고, 이 조합으로부터 추출한 특징점에 색상 정보 기반의 지역 서술자를 정의함으로써 사전에 정의된 서술자와의 비교를 통해 물체를 추적하는 방식이다. 제안된 추적 방식은 사전에 기준이 되는 지역 서술자를 정의함에 있어서 기준 영상에 다양한 어파인 변환을 적용함으로써 카메라와 대상물과의 각도가 큰 입력 영상에 대해서도 추적이 실패하지 않는다. 실험을 통해 제안된 방식을 집 모양으로 구성된 블록 조합에 적용하여 3차원 가상 콘텐츠를 증강시켜 봄으로써 제안된 방식의 유용성을 확인하였다. 제안된 방식은 텍스처가 없는 환경에서 사용자의 상호작용으로 텍스처를 구성하고 이를 추적하는 방식으로 향후 아이들을 위한 교육 프로그램, 모바일 기기에서의 응용 프로그램 등으로 적용 가능하다.

1. 서론

컴퓨터 비전 분야에서 주어진 영상에서 물체를 인식하고 추적하는 기술은 증강현실, 로봇 비전 등에 활용성이 높은 매우 중요한 연구 주제 중 하나이다. 특히 3차원 가상 콘텐츠를 실사 영상에 투영하여 보여주는 증강현실(Augmented Reality)에 있어서 정확하고 정교한 물체 추적 기술은 필수적인 기술이다. 이러한 물체 추적 기술은 카메라와 대상 물체와의 위치 및 방향 관계(카메라 포즈)를 정의하는 과정으로 생각할 수 있으며, 그 관계는 위치/방향 센서와 같은 기기를 이용하거나 카메라를 통해 획득한 영상으로부터 추정할 수 있다. 위치/방향 센서를 이용하는 경우 카메라와 대상 물체와의 관계를 쉽게 얻을 수 있는 반면 정교함이 떨어지는 문제가 있다. 그러므로 원만한 증강현실 구현은 영상 기반의 정보 해석을 통해 카메라와 대상 물체와의 관계를 정교하게 추정해내는 고차원적인 기술을 요구한다.

현재까지 정교한 증강현실 구현을 위한 영상 기반의 물체 추적 기술은 다방면에서 연구되어 왔으며, 이러한 추적 방식은 마커 기반 추적 방식과 비마커 자연영상 기반 추적 방식으로 분류할 수 있다. 마커 기반 추적 방식의 경우 영상에서의 가독성이 높은 특정 마커를 제작하여 이용하는 방식으로 ARToolkit[1] 등의 증강현실 구현을 위한 라이브러리에서 이용되고 있다. ARToolkit 에서는 추적이 용이한 사각형태의 테두리와 내부에 식별이 가능한 그림으로 구성된 마커를 제작하고, 영상에서 이를 추적함으로써 카메라와 대상 마커와의 위치/방향 관계를 얻어 낸다. 비마커 기반 추적 방식은 자연 영상에서 두드러진 특징을 보이는 특징점을 찾아내고 이를 묘사할 수 있는 지역 서술자(Local Descriptor)를 정의함으로써 물체를 추적해내는 방식과 영상의 특징점

에 대한 확률 모델을 이용한 패턴 분류 방식으로 구분할 수 있다. 전자의 예로는 SIFT(Scale Invariant Feature Transform)[2] 나 SURF(Speeded Up Robust Features)[3] 지역 서술자가 있으며, 후자의 경우 Ferns[4] 와 같은 방식을 예로 들 수 있다. SIFT는 특징점 주변 지역의 영상에 대하여 방향성을 찾아내고 이를 이용하여 128차원의 서술자를 정의하였고, SURF의 경우 특징점 주변에서의 Haar-wavelet 응답 분포를 기반으로 64차원의 서술자를 제한함으로써 SIFT의 속도의 한계를 개선하였다. Ferns는 확률 모델을 이용한 사전 학습에 의해 특징점 인식의 문제를 단순한 이분법적 분류과정으로 정의하고, 앞선 방법들과 비교하여 정확성을 잃지 않으면서도 속도를 높이는 성과를 얻어냈다.

그러나 마커 기반의 경우 육안 상 자연스럽지 못한 마커를 사용하기 때문에 환경 친화적인 형태를 갖출 수 없고, 비마커 자연영상 기반의 경우 영상에서 추출되는 특징점이 많아야 하고 이를 서술하기 위해 지역 영상의 변화 또한 많아야 유리하기 때문에 일상에서 일반적으로 발생하는 텍스처가 부족한 환경, 즉 특징점 추출이 어렵고 영상 변화가 많지 않은 경우에 적용하기 어렵다. 본 논문에서 이용한 어린이를 위한 놀이기구인 단색 나무 블록(그림 1)은 텍스처가 없는 7가지 단색 표면을 갖고 있어, 앞서 언급한 자연영상 기반의 추적 방식을 이용하여 특징점을 서술하기에 그 환경이 용이하지 못하다(그림 2).

본 논문에서는 텍스처가 없는 단색 나무 블록을 추적하여 증강현실을 구현하기 위한 색상 히스토그램 기반의 지역 서술자를 제안하고, 이를 이용하여 증강현실을 구현한다. 제안된 방식을 적용함에 있어서 추적 대상물은 단순하고 균일한 색상의 물체로 가정한다. 제안된 방식은 텍스처가 없는 환경에서 블록들의 조합을 이용하여 인위적으로 텍



그림 1. 단색 나무 블록(7가지 다른 색의 블록으로 구성).

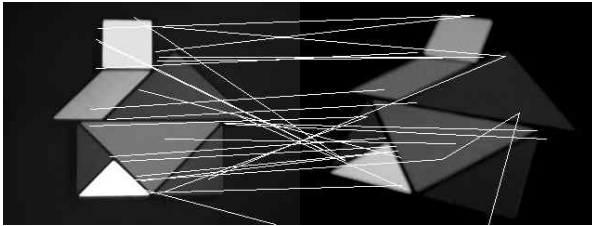


그림 2. SURF 지역 서술자를 이용한 추적 결과. 기준 영상(좌), 카메라 입력 영상(우).

스처를 구성하고 블록 간의 경계에서 얻은 경계점을 색상 히스토그램 서술자를 이용하여 정의함으로써 전체 블록의 조합을 추적할 수 있는 방법이다. 본 논문의 구성은 색상 히스토그램 기반 지역 서술자를 규정하는 방법과 이를 이용하여 블록들의 조합 위에 증강현실을 구현하는 실험 결과, 결론 및 향후 계획으로 이루어진다.

2. 색상 히스토그램 지역 서술자를 이용한 추적

가. 전체 흐름도

제안된 추적 방식의 전체 흐름(그림 3)은 크게 특징점 추출, 지역 서술자 정의, 카메라 포즈 추정, 3차원 가상 콘텐츠 증강으로 구분할 수 있다. 사전 정의된 기준 영상에서 추출된 특징점과 카메라로부터 입력되는 영상 프레임에서 추출된 특징점을 각각 지역 서술자로 정의하고, 이를 이용하여 서로 대응되는 특징점을 찾아냄으로써 카메라의 포즈를 추정할 수 있다. 기준 영상과 어파인 변환(Affine Transformation)된 기준 영상에 대해서 각 특징점에 지역 서술자를 정의함으로써 한 특징점에 대해 카메라의 다양한 위치에 따른 여러 개의 지역 서술자가 정의된다. 이는 카메라 입력 영상과의 대응점을 찾는 과정에서 카메라의 위치에 따라 획득된 영상이 기준 영상과 매우 달라 대응되는 특징점을 찾지 못하는 문제를 개선한다. 대응되는 특징점을 찾아내면 그 대응 관계를 이용하여 카메라 포즈를 추정하고 추정된 포즈를 3차원 가상 콘텐츠에 적용하여 증강된 영상을 얻을 수 있다.

나. 특징점 추출

사전에 정의된 기준 영상에서 Shi-Tomasi 코너[5]를 특징점으로 검출한다(그림 4). Shi-Tomasi 코너는 Harris 코너[6]의 추출방식을 기반으로 한 것으로 어파인 변환을 갖는 영상에 대해 보다 강건한 특징점 추출 성능을 보인다. 앞서 언급한 것과 같이 어파인 변환에 강건한 특징점을 정의하기 위해 원본 기준 영상에 임의의 어파인 변환을 적용하여 여러 개의 변환된 기준 영상을 얻어 낸 후 변환된 영상에서

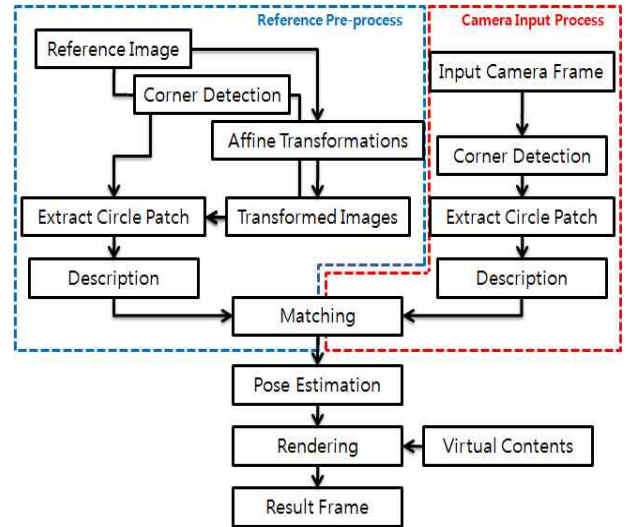


그림 3. 전체 흐름도.

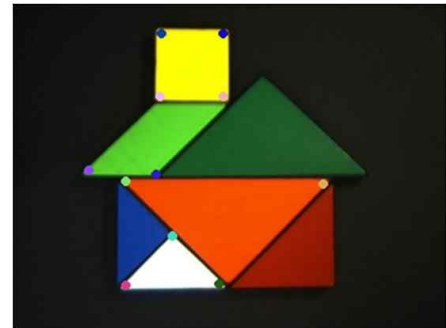


그림 4. 원본 기준 영상과 추출된 특징점.

도 같은 특징점을 검출해 낸다. 변환된 영상에서의 특징점의 위치는 원본 기준 영상과 변환된 영상간의 변환 관계로부터 계산해 낼 수 있다. 어파인 변환을 적용하기 위해 다음의 식이 이용된다[7].

$$\begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} = \begin{bmatrix} a_{11} & a_{12} & t_x \\ a_{21} & a_{22} & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \quad (1)$$

$$x' = H_A X = \begin{bmatrix} A & t \\ 0^T & 1 \end{bmatrix} X \quad (2)$$

$$R(\theta)R(-\phi)diag(\lambda_1, \lambda_2)R(\phi) \quad (3)$$

식 (1)은 식 (2)와 같이 표현할 수 있고, 여기서 x 와 x' 은 각각 변환 전의 좌표와 변환 후의 좌표를 의미한다. 식 (2)에서의 회전변환에 대한 부분인 A 는 식 (3)과 같이 분해될 수 있다. 여기서 $diag(\lambda_1, \lambda_2)$ 는 크기변환을 의미하며, $R(z)$ 는 각도 z 만큼의 회전변환을 의미한다. 제안된 방식에서는 λ_1 와 λ_2 를 모두 1로 하여 크기 변화를 고려하지 않고, θ 와 ϕ 는 극단적인 각도를 배제하고 그 조합의 수가 30개가 되도록 한다(그림 5).

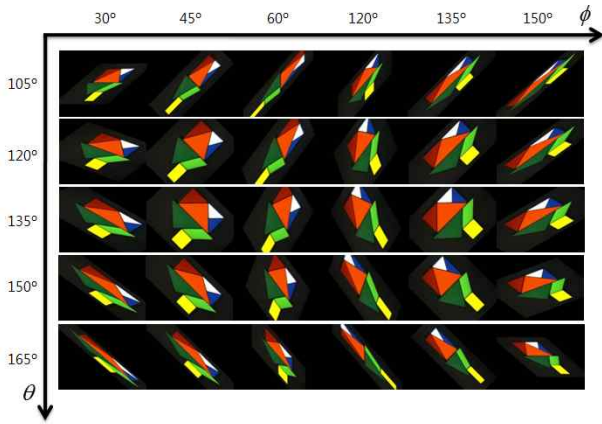


그림 5. 어파인 변환된 기준 영상.

다. 색상 히스토그램 기반 지역 서술자

지역 서술자 정의는 특징점 주변의 패치 추출, 색상 구별을 통한 분할 및 색상 별 픽셀 수 획득, 히스토그램 작성의 과정을 거치게 된다(그림 6). 지역 서술자를 정의하기 위해 원본 기준 영상과 어파인 변환된 영상에서 검출된 특징점을 중심으로 일정 크기의 원형 패치를 추출한다(그림 7). 제안하는 방식에서는 추출하는 원형 패치의 크기를 실험적으로 반지름 20픽셀로 규정하고 이 원형 패치는 7가지 나무 블록의 색상 중 몇 가지를 포함하게 된다. 여기서 패치를 구성하는 각 픽셀의 색상을 분별하여 패치 내에 포함된 색상별 픽셀 수를 얻을 수 있고, 이를 이용하여 히스토그램을 작성한다(그림 8). 제안하는 방식에서는 색상 분별을 위해서 사전에 YCrCb, HSV, RGB와 같은 색상 모델로부터 각 색상의 범위를 실험적으로 얻어 낸다. 원형 패치에서의 색상 히스토그램을 사용하기 때문에 회전변환에 강건한 지역 서술자를 얻을 수 있다. 작성된 히스토그램은 정규화 과정을 거쳐 지역 서술자로서 정의된다.

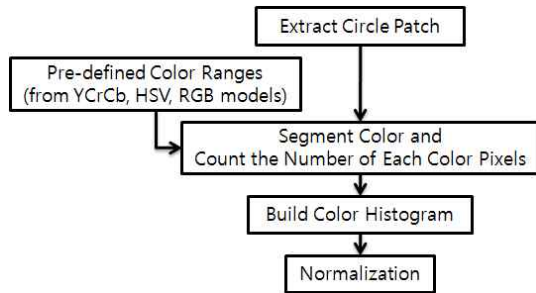


그림 6. 지역 서술자 정의 과정.

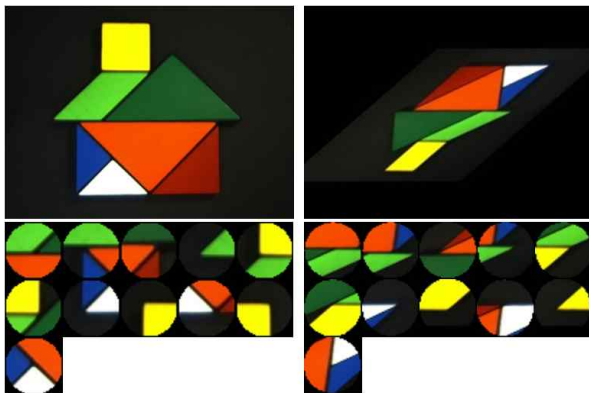


그림 7. 기준 영상에서 추출된 원형 패치. 원본 기준 영상과 원형 패치(좌), 어파인 변환된 기준 영상과 원형 패치(우).

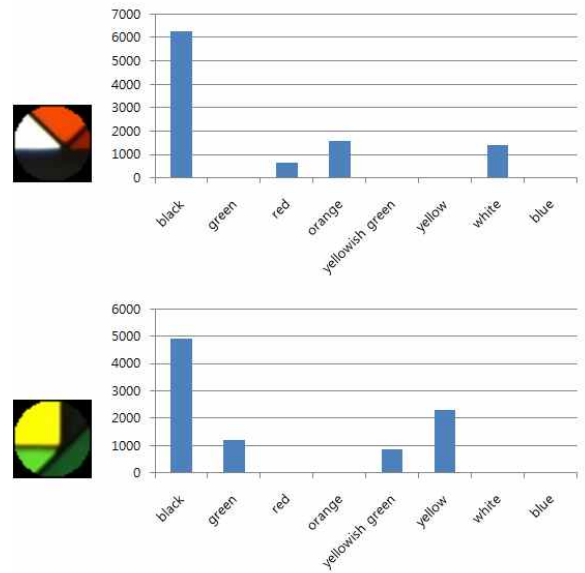


그림 8. 원형 패치에 대한 색상 히스토그램의 예. 원형 패치(좌), 색상 히스토그램(우).

라. 대응점 검출 및 카메라 포즈 추정

색상 히스토그램으로 정의된 지역 서술자를 이용하여 기준 영상과 카메라로부터 입력받은 영상 사이의 특징점 간 대응 관계를 찾아낸다. 대응점은 기준 영상과 카메라 입력 영상에서 추출된 지역 서술자의 각 성분 간 차를 제공하여 그 합이 최소가 되는 점으로 규정한다. 이는 다음의 식으로 표현 가능하다.

$$\hat{j} = \underset{j}{\operatorname{argmin}} \sum_k (h_{j,k}^{\text{ref}} - h_{i,k}^{\text{cam}})^2 \quad (4)$$

여기서 i 는 카메라 입력 영상에서 추출된 특징점, j 는 기준 영상들에서 추출된 특징점을 의미하며, h 는 서술자 히스토그램 원소의 값, k 는 히스토그램 원소의 가지수(단색 블록의 색상 수)를 나타낸다. 최종적으로 카메라 입력 영상에서의 특징점 i 와 대응하는 기준 영상에서의 특징점 j 를 검색하게 된다. 모든 대응점들을 찾고 난 후, Direct Linear Transformation(DLT) 방법[7]을 이용하여 기준 영상과 카메라 입력 영상간의 기하관계인 호모그래피(Homography)를 계산해 낸다(그림 9). 찾아낸 대응관계 중 올바르지 못한 것들의 영향을 배제하기 위해 RANSAC[7]을 적용하여 호모그래피를 획득한다. 카메라의 내부 특성(Intrinsic matrix)과 두 영상 간의 호모그래피를 알고 있을 때, 이들을 이용하여 카메라의 포즈를 추정할 수 있다.

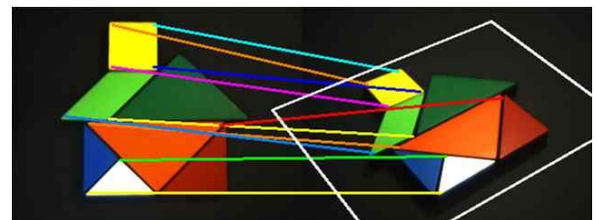


그림 9. 기준 영상(좌)와 카메라 입력 영상(우)간의 대응점 표시. 계산된 호모그래피를 이용하여 기준 영상의 네 모서리를 입력 영상에 투사하여 흰 색 사각형으로 표시.

3. 실험 및 결과

사전에 집 모양으로 맞춰 놓은 블록 조합을 촬영한 기준 영상과 같은 형태로 맞춰 놓은 블록 조합을 촬영한 카메라 입력 영상 간 추정된 카메라 포즈를 이용하여 3차원 가상 콘텐츠를 증강시켜 봄으로써 텍스처가 없는 물체에 대한 제안된 색상 히스토그램 지역 서술자의 유용성을 확인하였다(그림 10).

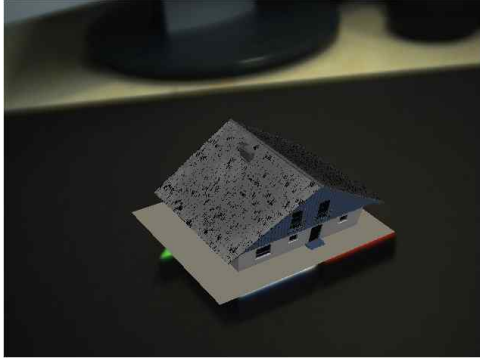


그림 10. 3차원 가상 콘텐츠를 증강한 결과 영상.

제안하는 추적 방식은 촬영 영상의 크기 및 회전 변화에 대해서 대상물을 강건하게 추적하는 결과를 보였으며(그림 11), 앞서 언급한 것과 같이 어파인 변환을 사전에 고려해 줌으로써 어파인 변환이 발생하는 영상에 대해서도 추적에 실패하지 않을 수 있었다(그림 12). 결과 영상의 흰 사각형은 추적된 물체를 표시한다.

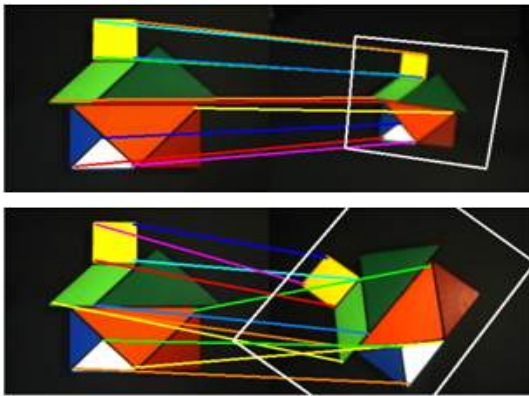


그림 11. 크기 변화(상) 및 회전 변화(하)에 대한 추적 결과.

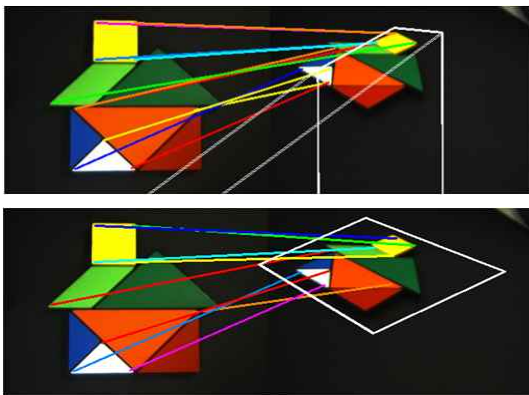


그림 12. 어파인 변환을 고려하지 않은 경우(상), 고려한 경우(하)의 결과.

4. 결론 및 향후 계획

본 논문에서는 7가지 단색의 나무 블록 조합 위에 증강현실을 구현하기 위해 색상 히스토그램 기반의 지역 서술자를 이용한 물체 추적 방식을 제안하고, 실제 블록 조합에 3차원 가상 콘텐츠를 증강시켜 봄으로써 제안된 방식의 유용성을 확인하였다. 이러한 방식은 기존의 마커를 이용하는 방식에서 문제시 되는 육안 상의 불편함으로부터 벗어날 수 있으며 텍스처가 부족한 환경에서 이용이 가능하다. 또한 기준 영상에 다양한 어파인 변환을 사전 적용함으로써 카메라의 촬영 각도 변화에 대해 비교적 강인한 추적 결과를 얻을 수 있었다. 그러나 제안된 방식은 색상 정보를 이용하기 때문에 조명의 변화가 적고 색상 종류가 적은 환경에서 적용이 용이하며, 배경의 색상이 복잡하고 동적인 경우 추적의 정확도가 낮아질 수 있다. 또한 사람이 인위적으로 블록의 위치를 조작하여 조합을 만드는 과정에서 얻게 되는 블록 조합은 엄밀하게는 실제 기준 영상과 차이가 있기 때문에 대상 블록 조합을 추적함에 있어서 어려움이 존재한다.

현재 제안된 방식은 극단적인 조명의 변화나 사용되는 카메라에 따른 카메라 특성 변화와 같은 환경 변화에 대해 추적의 강건함을 유지하기 위해 연구 중에 있으며, 기준 영상과 사후 구성된 블록 조합과의 차이에 대해서도 안정적인 추적을 가능하게 하는 개별 블록 추적 방식에 대해 지속적인 연구를 진행하고 있다. 또한 향후 아이들을 위한 상호작용형 증강현실 교육 프로그램, 모바일 기기 등을 활용한 응용 프로그램 등으로의 발전을 구상 중에 있다.

감사의 글

본 연구는 문화체육관광부 및 한국콘텐츠진흥원의 2009년도 문화콘텐츠산업기술지원사업의 연구결과로 수행되었음(2차년도).

참고문헌

- [1] H. Kato and M. Billinghurst, "Marker tracking and HMD calibration for a video-based augmented reality conferencing system," IEEE and ACM International Workshop on Augmented Reality (IWAR), pp. 85-94, 1999.
- [2] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," International Journal of Computer Vision, vol. 60, pp. 91-110, November 2004.
- [3] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," Computer Vision and Image Understanding, vol. 110, pp. 346-359, June 2008.
- [4] M. Ozuysal, M. Calonder, V. Lepetit, and P. Fua, "Fast keypoint recognition using random ferns," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, pp. 448-461, March 2010.
- [5] J. Shi and C. Tomasi, "Good features to track," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 593-600, June 1994.
- [6] C. Harris and M. Stephens, "A combined corner and edge detector," Alvey Vision Conference, pp. 147-152, 1988.
- [7] R. Hartley and A. Zisserman, "Multiple view geometry in computer vision," 2nd Edition, Cambridge University Press, March 2004.