

구면 파노라마 영상으로부터 사람의 자세 추정

임예슬, 박종승
 인천대학교 컴퓨터공학과
 imthree@inu.ac.kr, jong@inu.ac.kr

Human Pose Estimation from Spherical
Panorama Image

Ye-Seul Im, Jong-Seung Park
 Dept. of Computer Engineering, Incheon National University

요 약

사람의 자세는 구면 파노라마에서 다양한 형태로 왜곡되어 나타날 수 있다. 따라서 구면 파노라마에서의 자세 추정은 평면 이미지에서의 경우보다 정확도가 떨어진다. 본 논문에서는 인식률이 높은 얼굴 인식 기법을 도입하여 구면 파노라마 영상에서 안정적으로 사람의 자세를 추정하는 방법을 제시한다. 먼저 구면 파노라마에서 얼굴을 인식한 후에 이에 기반하여 사람의 전신 영역을 추정하고 전신 영역을 포함하는 평면 영상을 획득한다. 획득된 평면 영상에서 자세를 추정하여 스켈레톤을 얻고 이를 캐릭터 모델에 적용한다. 제안 방법을 실영상에 적용하여 실험한 결과 평면 이미지에서와 동일한 수준의 정확도를 보임을 확인하였다.

1. 서론

메타버스는 3차원에서 현실과 비현실이 공존하는 확장 가상 세계로 최근 AR(Augmented Reality)과 VR(Virtual Reality)의 발전과 함께 가상 세계를 넘어 현실 세계의 활동과 연결되고 있다. 최근 코로나 19 사태로 인한 비대면 시대에서 메타버스의 원격 회의 또는 원격 교육은 오프라인에서 생기는 많은 제약에서 벗어나게 한다[1]. 3차원 가상 세계에 대한 콘텐츠와 연구가 증가하면서 카메라를 통해 사람의 움직임을 추적하고 이를 가상 세계에 있는 캐릭터에 적용하는 연구의 중요성이 대두되고 있다.

사람을 움직임을 추적하는 연구들은 대부분 평면 카메라를 통해 진행되어 왔다. 평면 카메라를 이용한 영상은 카메라 시야 밖으로 사람이 이동하게 되면 더 이상 사람을 추적할 수 없다는 단점이 존재한다. 이를 해결하기 위해 구면 파노라마 영상을 이용하여 사람의 움직임을 추적하는 연구들이 진행되고 있다. 구면 파노라마 영상을 이용하는 기존 해당 분야의 연구는 여러 사람이 있는 영상에서 두 개의 시간적으로 정렬된 360° 이미지 시퀀스에서 사람을 하나씩 추적한 후 두 개의 카메라 위치에 상대적인 3D 스켈레톤을 생성하는 Shere 등의 연구, 구면 파

노라마 이미지에서 사람의 위치와 자세를 추정하기 위해 EPnP 알고리즘을 수정한 Fei 등의 연구 등 다양하게 진행되어 왔다[2][3].

본 논문은 평면 카메라의 단점을 해결하기 위해 360° 구면 카메라를 사용하여 얻은 단일 구면 파노라마 영상을 사람 중심의 평면 이미지를 얻어내고 이를 통해 사람의 자세를 추정한다. 구면 파노라마 이미지는 수평 360° 수직 180°의 시야각을 가지기 때문에 2:1의 비율로 생성되며 기존 평면 카메라의 고정된 시야의 제약에서 벗어나 많은 정보를 포함한다[4]. 구면 파노라마 영상으로부터 사람의 자세를 추정하기 위해서는 기존의 평면 영상으로 3차원 키포인트를 추적하는 기술을 사용하기 위해 구면 파노라마 영상을 평면 영상으로 변환하는 과정이 필요하다.

본 논문은 구면 파노라마 영상으로부터 사람의 자세를 추정하는 방법을 제안한다. 2장에서는 구면 파노라마 영상에서 사람을 찾아내기 위하여 얼굴을 검출하고 구면 파노라마가 사용하는 구면 좌표계를 평면 좌표계로 변환해 구면 파노라마 영상이 가지는 왜곡을 제거한 사람 중심 평면 영역을 확보한다. 평면 영역을 OpenPose를 사용하여 사람이 아닌 부분이 검출된 평면 영역을 삭제하고 사람의 자세를 추

정하는 방법을 설명한다. 3장에서는 원본 비디오와 비교하여 구면 파노라마 이미지로부터 사람 중심 평면 이미지를 검출하고 사람의 자세를 추정하였다는 것을 보여준다. 마지막으로 4장에서 결론을 맺는다.

2. 구면 파노라마로부터 사람의 자세 추정

구면 파노라마 영상으로부터 사람의 자세를 추정하기 위해서는 영상 내 사람의 위치를 찾아야 한다. 찾아낸 사람의 위치 정보를 통해 구면 파노라마 영상에서 사람 중심의 유클리드 좌표로 변경한다. 사람 중심의 평면을 확보하기 위해 유클리드 좌표계를 다시 구면 좌표계로 변경하고 사람 중심 평면 이미지 화소에 대응하는 구면 파노라마 이미지의 화소 좌표를 계산한다. 생성된 사람 중심 평면 이미지를 OpenPose에 적용하여 얼굴 검출에서 사람이 아닌 다른 부분을 검출한 이미지를 제외한다. 최종적으로 생성된 사람 중심 평면 이미지를 딥러닝을 사용하는 기술을 이용해 사람의 자세를 추정한다.

2.1. 얼굴 검출

캐릭터 애니메이션 생성을 위해서는 구면 파노라마 영상 내의 정보 중 사람의 정보가 필요하다. 구면 카메라가 인식할 수 있는 범위 내에 있는 모든 사람을 검출하기 위해 얼굴 검출을 이용한다. 구면 파노라마 영상에 OpenCV가 제공하는 Haar feature 기반 cascade classifier를 이용한 face detection을 사용해 (그림 1)과 같이 영상 속 사람을 모두 검출한다[5].



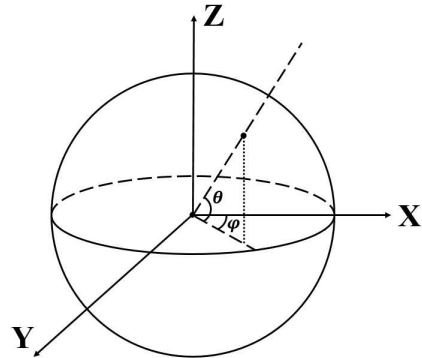
(그림 1) 구면 파노라마에서 얼굴 검출

2.2. 사람을 중심으로 한 평면 영역 확보

구면 파노라마 이미지는 구면 좌표계를 기반으로 하기 때문에 평면 이미지와는 다르게 공간이 왜곡되어 표현된다. (그림 2)에 나타난 것과 같이 구면 파노라마 이미지의 화소는 수평각 θ 와 수직각 ϕ 을 가지며 반지름이 1인 구로 나타낼 수 있다[6]. 구면 파

노라마 영상의 정 중양을 원점이라 가정하고 반지름이 1인 구형 투영면을 가질 때 모든 화소의 좌표를 (x, y) 에서 (θ, ϕ) 로 나타낼 수 있다. 식 (1)은 구면 파노라마 영상의 좌표 (x, y) 를 (θ, ϕ) 로 변환하는 수식이다. w 는 구면 파노라마 영상에서의 수평 길이, h 는 수직 길이를 나타낸다.

$$\begin{aligned} \phi &= 2x^*\pi/w - \pi \\ \theta &= y^*\pi/h - \pi \end{aligned} \tag{1}$$



(그림 2) 공간상의 구면 좌표계

구면 파노라마 영상에서 검출한 얼굴의 위치 (x, y) 를 통해 사람의 중심이 구면 파노라마의 어느 위도 경도에 위치하는지 식 (1)을 통해 계산할 수 있다. (그림 3)은 구면 파노라마 영상에서 큐브맵 하나의 평면 크기에 맞게 정 중양을 원점으로 하는 uv 좌표계로 변환한 다음 앞에서 얻어낸 사람 중심의 구면 좌표계 (θ, ϕ) 를 중심으로 유클리드 좌표계 (x_u, y_u, z_u) 로 변환하는 식이다.

```

Algorithm 1 Calculate human body location in Euclidean space
1: Input: Human body location in spherical space  $\{\theta, \phi\}$ ,
2:   Spherical image height, width  $h, w$ ,
3:   angle of view  $\omega$ .
4: Output: Human body coordinate set in Euclidean space  $p_u$ .
5: begin
6:   initialize  $p_u$ ;
7:    $size \leftarrow w/4; i \leftarrow 0; j \leftarrow 0$ ;
8:   while ( $i < size$ ) {
9:     while ( $j < size$ ) {
10:       $u \leftarrow (2 * j/size - 1) * \omega$ ;
11:       $v \leftarrow (2 * i/size - 1) * \omega$ ;
12:       $x_u \leftarrow \cos \phi \cos \theta - u \sin \phi + v \cos \phi \sin \theta$ ;
13:       $y_u \leftarrow \sin \phi \cos \theta + u \cos \phi + v \sin \phi \sin \theta$ ;
14:       $z_u \leftarrow -v \sin \theta \cos \theta$ ;
15:      add  $\{x_u, y_u, z_u\}$  to  $p_u$ ;
16:       $j \leftarrow j + 1$ ;
17:    }
18:     $i \leftarrow i + 1$ ;
19:  }
20: return  $p_u$ ;
21: end
    
```

(그림 3) 유클리드 좌표계로 나타낸 사람 중심 좌표

Grimm 등이 연구한 구면 파노라마 영상과 큐브맵 영상과의 관계에 대해 설명한 것과 같이 구면 파노라마를 여섯 개의 평면 이미지로 생성할 수 있다[7]. 큐브맵에서 하나의 평면 이미지는 구면 파노라마 이미지에서 수평 90° 수직 90°의 정보를 담고 있다. 구면 카메라에 가까이 있는 사람의 경우 큐브맵으로 생성한 하나의 평면 이미지에 사람이 전부 들어오기 어렵다. 식 (2)와 같이 검출된 얼굴의 크기 ℓ 을 사용하여 평면 이미지가 가질 수 있는 화각의 크기 ω 를 계산하고 사람의 정보를 하나의 이미지에 들어올 수 있도록 한다. (그림 4)는 화각을 조정하지 않았을 때의 평면 이미지와 화각을 조정하였을 때의 이미지의 비교를 나타낸다.

$$\begin{aligned} size &= w/4 \\ \omega &= \ell * 10 / size \end{aligned} \quad (2)$$



(그림 4) 화각을 조정한 이미지

식 (3)은 구면 파노라마 영상에서 사람 중심으로 한 평면 영역을 확보하기 위해 유클리드 좌표계 (x_u, y_u, z_u) 를 다시 구면 좌표계 (r, θ, ϕ) 로 변환하는 수식이다. 식 (4)는 사람 중심 평면 이미지 화소 좌표 (x, y) 에 대응되는 구면 파노라마 이미지 화소 좌표 (x_s, y_s) 를 계산하는 수식이다.

$$\begin{aligned} r &= \sqrt{x_u^2 + y_u^2 + z_u^2} \\ \theta &= \sin^{-1}(z_u/r) \\ \phi &= \sin^{-1}(y_u/\sqrt{x_u^2 + y_u^2}) \quad (x_u \geq 0) \\ \phi &= \cos^{-1}(x_u/\sqrt{x_u^2 + y_u^2}) \end{aligned} \quad (3)$$

$$\begin{aligned} x_s &= (\phi/\pi+1)*(w/2) \quad (x \geq 0, y \geq 0) \\ x_s &= (-\phi/\pi+1)*(w/2) \\ y_s &= (2\theta/\pi+1)*(h/2) \end{aligned} \quad (4)$$

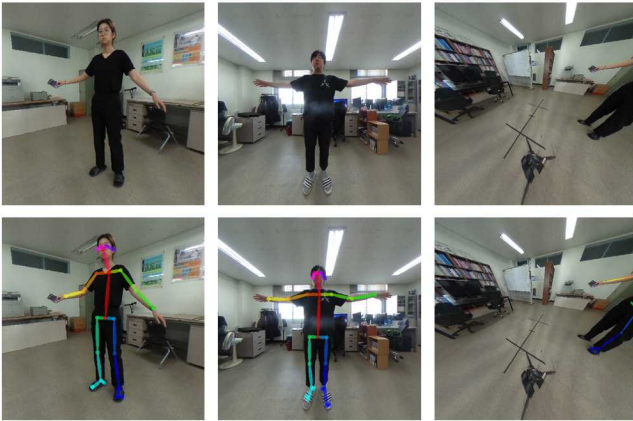
사람 중심의 평면 이미지를 생성한 후 OpenPose를 통해 구면 파노라마 영상에서 검출된 얼굴이 진짜 사람인지 재검증한다. 평면 이미지를 OpenPose를 이용하여 사람의 keypoints를 추적해 스켈레톤 조인트의 2차원 위치 좌표를 얻는다[8]. OpenPose의 결과가 없는 이미지의 경우 사람이 아닌 것으로 판단하여 제거하고 최종 사람 중심 평면 이미지를 획득한다.

평면 이미지로부터 사람의 포즈를 추정하기 위해서 AMASS를 활용하여 temporal pose 및 shape regression network에 의해 생성된 동작을 구별하는 학습 프레임워크를 통해 운동학적으로 그럴듯한 모션 시퀀스를 생성하는 Kocabas 등이 제안한 VIBE를 사용하여 캐릭터에 적용한다[8].

3. 실험 결과

본 논문에서 제안하는 방법은 구면 파노라마 영상으로부터 사람의 자세 추정이다. 본 논문의 실험은 Intel Core i5 CPU, NVIDIA Geforce GTX 970, 32GB RAM, Windows 10 64bit OS의 시스템에서 수행하였다. 실험 코드는 Unreal Engine과 C++언어, OpenCV 라이브러리, OpenPose 라이브러리, AMASS 데이터 세트를 사용하여 Human Body Pose와 Shape를 추정하는 연구의 오픈 소스 라이브러리를 이용하여 구현하였다.

본 논문의 실험은 구면 파노라마 영상에서 하나의 프레임을 사용하여 수행하였다. (그림 5)에서 위의 영상은 구면 파노라마에서 OpenCV를 이용해 얼굴을 검출하고 이를 이용하여 얻어낸 사람 중심 평면 이미지이다. 아래 영상은 평면 이미지로부터 OpenPose를 적용하여 사람의 스켈레톤 조인트 keypoints를 추적한 영상이다. 생성된 사람 중심 평면 이미지 중 keypoints 정보가 존재하는 두 개의 이미지만을 사용하여 사람의 자세를 추정한다. (그림 6)은 VIBE를 사용하여 캐릭터에 적용하고 사람 중심 평면 이미지를 통해 그럴듯한 모션 시퀀스를 생성하였음을 보여주는 영상이다.



(그림 5) 사람 중심 평면 이미지를 통한 키포인트 추적



(그림 6) 사람의 자세 추정

4. 결론

본 논문에서는 구면 파노라마 영상으로부터 사람의 자세를 추정하는 방법을 제안하였다. 제안하는 방법은 실험을 통해 구면 파노라마에서 사람의 얼굴을 추적하고 얻어낸 정보를 바탕으로 사람 중심의 평면 이미지를 얻어내 사람의 자세를 추정할 수 있음을 확인하였다. 제안 방법을 통하여 평면 영상보다 더 많은 정보를 담고 있는 구면 파노라마 영상으로부터 여러 사람의 자세를 추정할 수 있다는 점에서 기여점이 있다.

향후 연구 과제로는 모든 프레임마다 사람의 위치를 추적하고 평면 이미지를 얻어내야 한다는 점에서 불필요한 작업을 줄일 수 있도록 다음 프레임에 사람이 어느 위치에 존재하는지 찾을 수 있는 알고리즘을 추가할 수 있는 후속 연구가 필요하다. 또한 얼굴을 검출할 때 구면 파노라마에서 왜곡이 심한 위치에 얼굴이 존재할 경우 검출하기 매우 어렵다는 한계가 존재하기 때문에 이를 해결하기 위한 연구가 필요하다.

Acknowledgments

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No. NRF-2019R1F1A1060828).

참고문헌

- [1] 김광집, “메타버스 사례를 통해 알아보는 현실과 가상 세계의 진화”, 방송과 미디어, Vol. 26, No. 3, pp. 10-19, 2021.
- [2] M. Shere, H. Kim, A. Hilton, “3D Human Pose Estimation From Multi Person Stereo 360° Scenes”, Conference on Computer Vision and Pattern Recognition, pp. 1-8, 2019.
- [3] D. Fei, W. Yousi, H. Yulei, C. Hongxia, “Position and Pose Estimation of Spherical Panoramic Image with Improved EPnP Algorithm”, Acta Geodaetica et Cartographica Sinica, Vol. 45, No. 6, pp. 677-684, 2016.
- [4] G. Fangi, C. Nardinocchi, “Photogrammetric processing of spherical Panoramas”, The Photogrammetric Record, Vol. 28(143), pp. 293-331, 2013.
- [5] X. Fan, F. Z hang, H. Wang, X. Lu, “The System of Face Detection Based on OpenCV”, 2012 24th Chinese Control and Decision Conference(CCDC), pp. 648-651, 2012.
- [6] G. Tong, J. Gu, “Locating Objects in Spherical Panoramic Images”, IEEE International Conference on Robotics and Biomimetics, pp. 818-823, 2011.
- [7] C.M. Grimm, B. Niebruegge, “Continuous cube mapping”, Journal of Graphics Tools, pp. 25-34, 2007.
- [8] M. Kocabas, N. Athanasiou, M. J. Black, “VIBE: Video Inference for Human Body Pose and Shape Estimation”, CVPR, pp. 5253-5263, 2020.