

다수의 건설인력 위치 추적을 위한 스테레오 비전의 활용

Simultaneous Tracking of Multiple Construction Workers Using Stereo-Vision

이용주¹⁾, 박만우²⁾

Lee, Yong-Ju¹⁾ · Park, Man-Woo²⁾

Received March 6, 2017; Received March 20, 2017 / Accepted March 21, 2017

ABSTRACT: Continuous research efforts have been made on acquiring location data on construction sites. As a result, GPS and RFID are increasingly employed on the site to track the location of equipment and materials. However, these systems are based on radio frequency technologies which require attaching tags on every target entity. Implementing the systems incurs time and costs for attaching/detaching/managing the tags or sensors. For this reason, efforts are currently being made to track construction entities using only cameras. Vision-based 3D tracking has been presented in a previous research work in which the location of construction manpower, vehicle, and materials were successfully tracked. However, the proposed system is still in its infancy and yet to be implemented on practical applications for two reasons. First, it does not involve entity matching across two views, and thus cannot be used for tracking multiple entities, simultaneously. Second, the use of a checker board in the camera calibration process entails a focus-related problem when the baseline is long and the target entities are located far from the cameras. This paper proposes a vision-based method to track multiple workers simultaneously. An entity matching procedure is added to acquire the matching pairs of the same entities across two views which is necessary for tracking multiple entities. Also, the proposed method simplified the calibration process by avoiding the use of a checkerboard, making it more adequate to the realistic deployment on construction sites.

KEYWORDS: Construction worker, Construction automation, Computer vision, Image processing, Tracking

키워드: 건설인력, 건설 자동화, 컴퓨터 비전, 영상처리, 위치추적

1. 서론

1.1 연구의 배경

오늘날의 건설 구조물들은 복잡한 구조로 인해 건설 자재, 장비, 인력 등의 다양한 건설 자원들이 건설 현장에 복잡하게 산재되어 있다. 현장 지휘자는 해당 자원들을 적시 적소에 사용할 수 있도록 현장을 지휘해야 하지만 복잡한 건설 환경으로 인해 효과적인 현장 지휘가 점점 어려워지고 있다. 이러한 이유로 각종 센서들을 활용하여 건설 자원들의 데이터를 확보하고자 하는 노력이 진행되고 있다. 건설 장비에 각종 센서를 부착하여 장비의 가동률을 파악하거나 현재 어떤 작업에 투입되어 있는지를 파악할 수 있으며 GPS나 RFID 등을 이용하여 건설 자재나 건설 인력의 위치와 상태 정보를 획득할 수도 있다. 이러한 데이터들은 공정 관리, 안전 관리 등의 목적으로 사용될 수 있다 (Carbonari et al., 2011; Turkan et al., 2012). 건설 현장에서

수집되는 다양한 데이터 중에서 위치 데이터는 특히 중요하다. 특정 센서들을 이용하면 복잡한 건설 현장에서 지정된 자원들의 위치를 비교적 정확하게 파악할 수 있으며 이를 실시간으로 업데이트할 수 있다. 이를 이용하여 공정 진행 상황을 실시간으로 시각화할 수 있으며 건설 장비 간의 간섭 여부 등을 미리 파악하여 사전에 경고함으로써 안전 관리에도 사용될 수 있다(Chan et al., 2016). 건설 자원의 위치 데이터를 확보하기 위해 GPS (Global Positioning System), RFID(Radio Frequency IDentification), UWB(Ultra Wide Band) 등의 다양한 위치 추적 기술들이 사용되고 있거나 연구가 진행되고 있다(Domdouzis et al., 2007; Cheng et al., 2011; Vasenev et al., 2014). 하지만 위에서 언급한 위치 추적 기술들은 무선 주파수 기술을 기반으로 하고 있기 때문에 데이터를 생성하기 위해서 태그(Tag)와 리더(Reader)를 이용하여 무선 신호를 송수신해야 한다. 따라서 추적하고자 하는 개체에 반드시 태그를 부착해야만 한다. 이는 태그의 관리에

¹⁾학생회원, 명지대학교 토목환경공학과 석사과정 (leetoday@mju.ac.kr)

²⁾정회원, 명지대학교 토목환경공학과 조교수 (mwpark@mju.ac.kr) (교신저자)

다른 시간과 비용적 측면에서의 노력을 필요로 한다. 또한, 추적하고자 하는 개체 수가 많아지게 되면 필요한 관리 노력 또한 증가하게 된다. 영상 기반 위치 추적 방법은 카메라와 영상을 처리할 프로세서만을 이용하여 대상을 추적하기 때문에 별도의 태그를 필요로 하지 않는다. 기존의 방법들과 달리 추적하고자 하는 개체의 수가 증가해도 추가적으로 소요되는 관리 노력이 매우 적다는 장점이 있다. 건설 현장에서 영상 기반 위치 추적 방법을 적용하기 위해 다양한 연구들이 진행되어 왔다. 하지만 대부분의 기존 연구들은 2D 영상에서 특정 개체를 검출해내는 것에 그치고 있으며 영상들을 이용하여 특정 개체의 3D 위치 좌표를 획득하는 방법에 대한 연구는 매우 드물다(Seo et al., 2015). 3D 위치 데이터 획득을 위해 연구된 기존의 방법 또한 몇 가지 한계를 가지고 있다(Park et al., 2012). 해당 연구에서 제안하는 방법은 카메라만을 이용하여 차량, 인력 등의 건설 자원의 3D 위치를 성공적으로 추적해낼 수 있으나 개체 별도의 자동화된 개체 매칭(Entity matching) 과정이 없기 때문에 다수의 개체를 동시에 추적하기에는 어려움이 있다. 스테레오 영상을 이용하는 영상 기반 위치 추적 방법의 특성상 촬영된 영상에서 개체를 검출하고 추적하는 것은 물론, 두 영상에서 검출된 개체 중 동일 개체를 짝 지어주는 과정이 반드시 필요하다. 이 과정이 자동화되지 않는다면 영상을 이용하여 다수 개체의 위치를 추적하는 것은 사실상 불가능하다.

본 연구에서는 등극선 기하학(Epipolar geometry)에 근거한 개체 매칭 과정을 추가함으로써 기존 연구들이 가지는 한계를 보완하여 다수 개체를 추적할 수 있는 영상 기반 위치 추적 방법을 제시하고자 한다. 개체 매칭 과정에서는 기초 행렬(Fundamental matrix)로부터 계산 가능한 에피폴라 라인(Epipolar line)을 개체 매칭 정보로 이용한다.

1.2 연구의 범위 및 목적

영상 기반 위치 추적의 최종적인 목적은 특정 개체의 3D 위치 데이터 확보에 있으나 기존 연구들은 대부분 건설 현장 안전 관리 목적으로 진행되었으며 대부분의 연구들은 2D 영상에 국한되어 있다(Seo et al., 2015). 또한 3D 위치 좌표 획득을 목적으로 진행된 연구(Park et al., 2012)의 경우에도 건설 현장에 적용하기에 제한사항으로 작용할 수 있는 몇 가지 한계점들을 가지고 있다.

카메라 캘리브레이션은 카메라의 초점 거리, 위치 및 자세 등의 특성 정보를 계산하는 과정으로, 스테레오 카메라를 이용한 영상 기반 위치 추적을 위해서는 반드시 필요한 과정이다. 이 과정에서 일반적으로 체커보드가 사용된다. 체커보드는 일정한 크기의 체크박스로 이루어져 있으며 각 체크박스의 변의 길이는 모두 동일하다. 일반적인 카메라 캘리브레이션은 카메라로

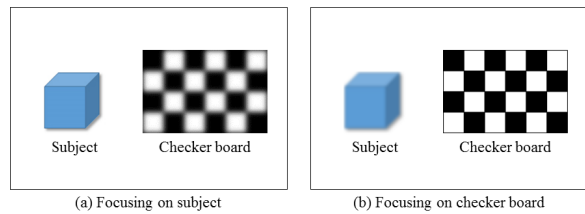


Figure 1. Correlation of camera calibration and focal length

촬영된 체커보드로부터 체크박스의 꼭지점을 검출하여 각 꼭지점의 위치와 거리를 이용해 카메라의 특성 값을 계산하는 과정이다. 기존 연구는 스테레오 카메라의 캘리브레이션 과정에 체커보드를 사용했다. 하지만 건설 현장과 같은 야외 환경에서 체커보드를 이용한 카메라 캘리브레이션은 몇 가지 문제점을 가지고 있다. 일반적으로 스테레오 영상에서 깊이를 구분할 수 있는 한계 거리는 카메라 간의 거리(기저선, Baseline)와 비례한다. 다시 말해, 넓은 영역을 다루기 위해서는 기저선의 길이가 그만큼 길어져야 한다. 하지만 기저선의 길이가 길어지면 카메라 캘리브레이션 과정에서 문제가 발생할 수 있다. 기저선이 너무 긴 경우, 하나의 체커보드를 사용한 캘리브레이션이 불가능할 수 있다. 스테레오 카메라의 캘리브레이션 과정은 두 카메라가 동시에 같은 체커보드를 촬영하여 진행되는데, 기저선이 길어지면 두 카메라에 체커보드가 완전히 비춰지지 않는다. 따라서 카메라로부터 더 먼 거리에서 체커보드를 촬영해야 하며 체커보드와 카메라 간의 거리가 멀어지는 만큼 정확한 캘리브레이션을 위해서는 더 큰 체커보드가 필요하게 된다. 두 카메라를 따로 캘리브레이션을 하는 방법도 있지만 이 경우에도 초점거리로 인해 문제가 발생할 여지가 있다.

카메라의 자동 초점 기능을 이용하면 초점 거리 등의 특성 값이 수시로 변화할 수 있기 때문에 정확한 캘리브레이션을 할 수 없게 된다. 따라서 영상 기반 위치 추적을 위한 카메라는 자동 초점 기능을 사용하지 않는다. 카메라와 피사체의 거리가 멀다면 피사체를 온전히 촬영하기 위해서 초점이 원거리에 고정되어야하므로 근거리에서 체커보드를 사용할 경우, 초점이 맞지 않아 상이 흐릿하게 맺히기 때문에 정확한 카메라 캘리브레이션이 불가능할 수 있다(Fig. 1 (a)). 반대로 카메라 캘리브레이션을 위해 근거리에 초점을 고정하게 되면 카메라 추적하고자 하는 개체의 상이 흐릿하게 맺히기 때문에 정확한 데이터를 확보하기 어렵다(Fig. 1 (b)).

기존 연구가 가진 또 다른 한계점은 좌/우 영상에서 검출된 개체의 매칭이 불가능하다는 것이다. 3D 위치 좌표를 계산하기 위해선 스테레오 영상에서의 대응점이 필요하다. 따라서 좌/우 영상에서 독립적으로 검출된 개체 중 동일한 개체끼리 매칭시켜 줄 필요가 있다. 즉 Figure 7과 같이 좌측 영상에서 검출된 하나

의 개체가 오직 영상에서 검출된 세 개의 개체 중 어떤 개체와 동일한 개체인지를 판단할 수 있어야 한다. 이 과정을 본 연구에서는 개체 매칭이라 한다. 특히 다수의 인물을 추적하기 위해서는 영상으로부터 검출된 다수의 인물 중 동일한 인물끼리의 매칭쌍을 찾을 수 있어야하므로 이 개체 매칭 과정이 반드시 필요하다. 하지만 기존 연구에서는 하나의 개체를 추적하는데 중점을 두고 있어, 개체의 매칭 과정을 수동적으로 진행했기 때문에 다수의 개체를 추적하기에는 적합하지 않다. 개체 매칭 과정을 자동화하기 위해서는 검출된 개체들을 식별할 수 있는 별도의 정보가 필요하다.

본 연구에서는 캘리브레이션 과정에서 체커보드를 사용함으로써 인해 발생할 수 있는 문제들을 해결하기 위해 체커보드 대신 GCP(Ground Control Point)를 사용한다. GCP는 실측된 위치 좌표로서 이를 이용해서 카메라의 특성 값을 계산할 수 있다. 본 연구에서는 보다 정확한 결과를 위해 다수의 GCP를 이용하여 카메라 특성 값을 계산한다. 추적 대상이 되는 인물들의 주변 환경으로부터 GCP를 획득하기 때문에 체커보드를 사용할 때 발생할 수 있는 초점거리, 기저선의 문제를 해결할 수 있다. 또한, 다수의 건설 인력을 추적하기 위해서 별도의 개체 매칭 과정을 추가한다. 개체 매칭 과정에서는 기초행렬로부터 계산할 수 있는 에피폴라 라인(Epipolar line)을 이용하여 동일한 개체를 찾아낸다(Lee et al., 2016). 최종적으로 두 과정의 결과를 이용하여 건설 인력의 3D 위치 좌표를 계산하고, 정확도를 확인하여 본 연구에서 제안하는 다수 건설 인력의 영상 기반 위치 추적 방법의 적용 가능성을 검토해보고자 한다.

2. 연구 방법

2.1 다수 인력의 위치 추적

Figure 2는 본 연구에서 제안하는 영상 기반 다수 건설 인력 위치 추적 방법의 전체 진행 과정을 도식화한 것이다. 본 연구에서 제안하는 다수 건설 인력의 위치 추적 방법은 기존의 연구 성과들에서 카메라 캘리브레이션 과정과 개체 매칭 과정을 보완하는 것을 목표로 한다. 본 연구에서는 다수 개체의 위치를 추적하기 위해 별도의 개체 매칭 과정을 추가했으며 체커보드 대신 GCP를 사용하여 카메라 캘리브레이션을 수행한다. 이를 위해 각 영상의 첫 번째 프레임으로부터 기초행렬의 계산이 선행된다. 계산된 기초행렬은 카메라 캘리브레이션 과정뿐만 아니라 개체 매칭 과정에서도 사용된다. 각 영상에서는 독립적으로 개체를 검출 및 추적한다. 추적된 개체들은 계산된 기초 행렬을 이용하여 개체 매칭 과정을 수행하고, 이 결과들을 이용하여 최종적으로 3D 위치 좌표를 계산한다.

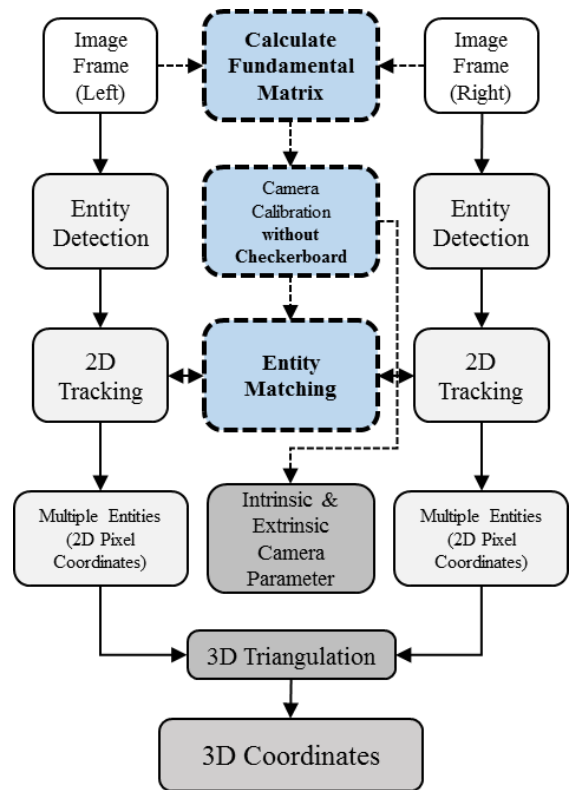


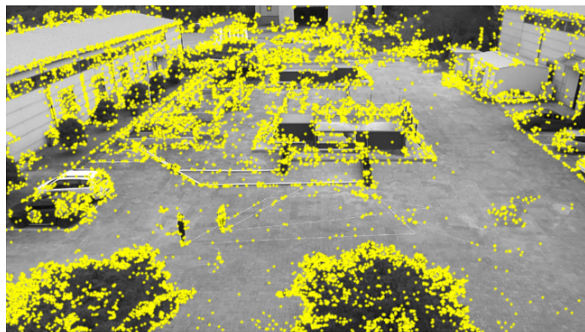
Figure 2. Workflow of vision based 3D tracking

2.2 개체 매칭

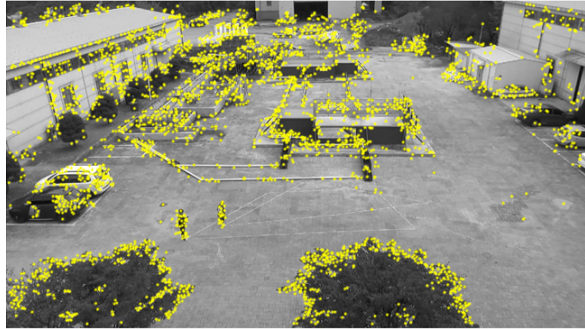
본 연구에서 사용된 개체 매칭 방법은 기존의 연구 방법을 채용하였으나(Lee et al., 2016) 향상된 결과를 얻기 위해 몇 가지 제한점을 두었다. 기존 방법은 개체 매칭을 위해 등극선 기하학을 적용했다. 개체 추적 과정에서 얻어진 개체 중심점으로부터 기초행렬을 이용해 에피폴라 라인을 계산하고, 계산된 에피폴라 라인과 추적 개체의 중심점과의 거리를 비교하여 개체 매칭 여부를 판단한다. 기초행렬 계산을 위해 각 영상의 첫 번째 프레임을 이용하여 특징점을 추출한다. 다양한 특징점 추출 알고리즘이 있지만 본 연구에서는 가장 보편적으로 사용되는 SIFT(Scale Invariant Feature Transform)(Lowe, 2004) 알고리즘과 SURF(Speeded Up Robust Features)(Bay et al., 2008) 알고리즘을 비교하여 더 많은 특징점을 반환하는 SIFT 알고리즘을 사용했다(Fig. 3). 좌/우측 영상에서 추출된 특징점들을 비교하여 동일한 대응점만을 기초 행렬 계산에 사용했다. 이 과정에서 대응점 판단을 위해 MSAC(M-estimate SAmple Consensus)(Torr & Murray, 1997)이 사용됐다.

2.2.1 Entity distance threshold

본 연구에서는 개체 중심점과 에피폴라 라인 사이의 수직 거리를 Entity distance d_e 라 정의했으며 d_e 의 크기에 따라 매칭



(a) SIFT Feature extraction results



(b) SURF Feature extraction results

Figure 3. Comparison of SIFT and SURF

여부를 판단한다. d_E 가 정해진 $threshold(T_E)$ 보다 작은 경우에만 매칭시키며 반대의 경우, 에피플라 라인이 개체 중심점으로부터 너무 멀리 떨어져 있으므로 매칭 여부를 정확히 결정할 수 없다고 판단하여 매칭을 보류하게 된다.

2.2.2 Entity distance ratio threshold

에피플라 라인과 개체 중심점 사이의 수직거리에 따라 매칭 여부를 판단하게 되면서 하나의 에피플라 라인 주변에 다수의 개체 중심점이 근접할 때 그 중 어떤 개체와 매칭을 이뤄야 하는지에 대한 문제가 발생할 수 있다. Figure 4와 같이 p 의 에피플라 라인 l 주변에 p_1, p_2 두 개의 개체 중심점이 근접할 때, Entity distance threshold에 의해 한 개체가 탈락했다면 문제가 되지 않지만 두 개체 모두 통과한 경우에는 p 가 p_1, p_2 중 어떤 개체와 매칭을 이뤄야 할지 결정하기가 모호하다.

이 경우, 기본적으로 에피플라 라인에 더 가깝게 위치한 개체를 더 적합한 개체로 판단하여 매칭하게 된다. p_1 과 l 사이의 수직거리를 d_1 , p_2 과 l 사이의 수직거리를 d_2 라 할 때, 만약 $d_1 < d_2$ 라면 p 는 p_1 과 매칭된다. 반대로 $d_1 > d_2$ 라면 p 는 p_2 과 매칭된다. 하지만 각종 오차들로 인해 그 결과가 항상 정확하지는 않다. 픽셀 좌표의 차이, 개체 검출 및 추적 과정에서의 오차, 기초행렬 계산의 오차 등이 그 원인이다. 본 연구에서는 이 오류를 최대한 방지하기 위해서 Entity distance ratio threshold T_R

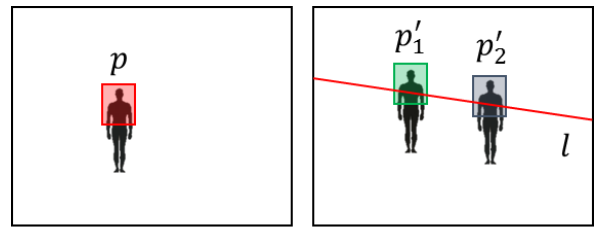


Figure 4. Entity distance ratio threshold

(이하 Ratio threshold)를 적용했다. 하나의 에피플라 라인에 근접한 다수의 개체 중에서 가장 가까운 거리에 위치한 개체를 선정하되, 다른 후보 개체와의 격차가 명확하게 큰 경우에만 해당 개체로 확정짓는 방법이다. 이를 비교하기 위해 Entity distance ratio R_E 를 정의했으며 다음 식과 같다.

$$R_E = \frac{d_E(2nd\ closest)}{d_E(1st\ closest)}$$

2.3 3D 위치 좌표 계산

영상으로부터 3D 위치 좌표를 계산하기 위해서는 두 영상에서의 대응점($x_i \leftrightarrow x'_i$)이 필요하다. Figure 5와 같이 실제 공간상의 점 X_i 와 그것의 영상 평면에서의 대응점($x_i \leftrightarrow x'_i$)이 위치해 있을 때, 대응점 x_i, x'_i 이 위치한 영상 평면과 X_i 가 위치한 공간간의 기하학적 상관관계를 통해 대응점으로부터 X_i 의 좌표를 계산할 수 있다. 이 과정에서 카메라 행렬 P 와 P' 이 필요하며 대응점과 카메라 행렬의 관계는 다음 식을 만족한다.

$$x_i = PX_i, \quad x'_i = P'X_i \quad (\text{for all } i)$$

카메라 행렬은 기초행렬로부터 계산할 수 있지만 정확한 계산을 위해서는 충분히 많은 수의 대응점을 필요로 한다. 본 연구에 적용된 방법을 통해 영상의 대응점들로부터 3D 위치 좌표를 계산하는 과정은 다음과 같다.

- (1) 특징점의 매칭쌍들로부터 기초행렬을 계산한다.

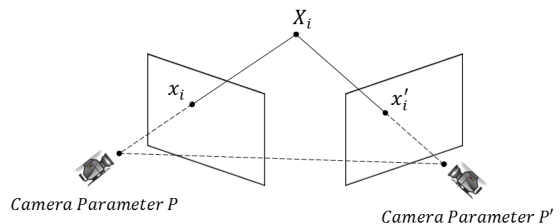


Figure 5. Relation of 3D and 2D pixel location

- (2) 계산된 기초행렬로부터 카메라 행렬을 가정한다.
- (3) 가정된 카메라 행렬을 이용해 GCP의 위치 좌표를 계산한다.
- (4) GCP의 실측 좌표와 계산된 좌표를 비교하여 호모그래피 행렬을 계산한다.
- (5) 호모그래피 행렬을 이용해 가정된 카메라 행렬을 보정한다.
- (6) 보정된 카메라 행렬을 이용해 추적 개체 대응점의 위치 좌표를 계산한다.

본 연구에서는 개체 매칭 과정에 사용된 기초행렬을 3D 위치 좌표 계산 과정에도 동일하게 사용하므로 카메라 행렬 가정을 위한 추가적인 기초행렬 계산은 필요로 하지 않는다.

2.3.1 카메라 행렬 계산 및 보정

본 연구에서는 카메라 행렬 계산을 위해 기초행렬이 사용된다. 기초행렬로부터 카메라 행렬을 가정하고, GCP를 이용하여 이를 보정한다. 카메라 행렬과 기초행렬의 관계는 다음 식과 같이 나타낼 수 있다.

$$P = [I \mid 0]$$

$$P' = [[e']_x F \mid e']$$

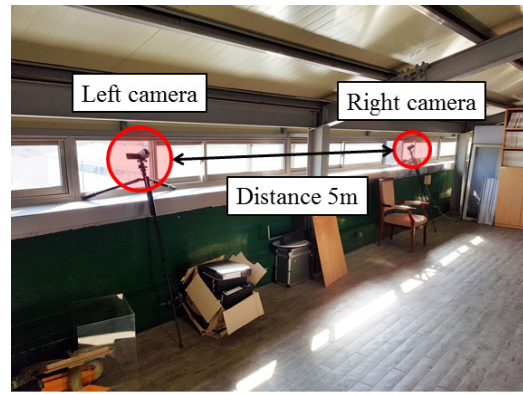
이 관계로부터 최초 카메라 행렬 P와 P'을 가정하고, 가정된 카메라 행렬을 이용해 재구성(Reconstruction)된 장면의 위치 좌표와 실측 위치 좌표(GCP) 간의 기하학적 관계를 이용해 카메라 행렬을 다시 보정한다. 스테레오 영상에서 여러 개의 대응점을 이용하여 유일한 기초행렬이 정해졌다면 기초행렬과 해당 대응점을 이용하여 카메라와 각 대응점의 3D 위치를 재구성할 수 있다(Hartley & Zisserman, 2003). 오차를 최대한 제거한 형태로 재구성하기 위해서는 카메라 행렬의 보정이 필요하며 이 보정을 위해 호모그래피 행렬(Homography matrix)을 사용한다. 호모그래피 행렬은 실측된 3D 위치 좌표를 이용해 계산할 수 있으며 본 연구에서는 GCP를 활용하여 이를 계산한다.

2.3.2 3D 위치 좌표 계산

보정된 카메라 행렬과 3D 위치 좌표 계산을 위한 영상의 대응점들이 주어졌다면 삼각화(Triangulation)를 통해 3D 위치 좌표를 계산할 수 있다. 3D 위치 좌표와 계산에 필요한 값들의 관계는 다음 식과 같다.

$$X = f(x, x', P, P')$$

본 연구에서는 3D 위치 좌표 계산을 위해 선형 삼각화(Linear triangulation) 방법을 사용한다. 삼각화 방법은 실제 3D 공간상의 한 점 X_i 가 영상에 투영되어 생긴 대응점 x_i, x'_i 로부터 X_i 의 위치 좌표를 계산하는 방법이다.



(a)



(b)

Figure 6. Video data acquisition

3. 실험 결과

3.1 실험 구성

본 연구에서는 에피폴라 라인을 이용하여 다수 개체의 매칭을 가능하게 했으며 이를 이용해 3D 위치를 추적할 수 있는 방법에 대해 제안했다. 제안된 방법을 통해 개체 매칭을 실시하고, 개체 매칭의 결과로 3D 위치 좌표를 계산하여 계산된 좌표와 실측 위치 좌표의 비교를 통해 적용 가능성을 확인해보고자 한다.

실험을 위해 목표 추적 개체는 인물로 한정했으며 실험 영상은 건설 현장과 같이 피사체와 카메라 간의 거리가 비교적 멀리 위치해 있는 실외 환경에서 두 대의 카메라를 이용해 촬영했다. Figure 6 (a)와 같이 좌/우측 시점의 카메라를 5m 간격을 두고 설치했으며 두 카메라는 Figure 6 (b)와 같이 촬영 대상 영역을 비추고 있다.

영상에는 첫 화면부터 마지막 화면까지 두 명의 인물이 등장한다. 개체 검출 및 추적 과정에서 별도의 모델을 추가하지 않았기 때문에 직립 보행 상태만을 검출할 수 있다. 따라서 영상에 등장하는 인물들은 지속적으로 보행을 실시하며 보행 외의 다른 자세를 취하지 않도록 했다. 계산된 위치 좌표와 실측 위치 좌표



Figure 7. GCP and designated path

의 비교 검증을 위해 두 인물은 정해진 경로를 따라 이동한다. 이동 중인 개체의 실시간 위치 좌표를 획득할 수 없기 때문에 영상 촬영 이전에 각 경로의 좌표를 획득해두었으며 실험 결과는 이 좌표와 비교할 수 있도록 했다. 경로의 위치 좌표는 토탈 스테이션을 이용해 획득했다. 토탈 스테이션은 경로의 위치 좌표 외에도 호모그래피 행렬 계산을 위한 GCP 위치 좌표를 획득할 때에도 사용됐다. GCP는 총 25개로 구성했다. 이동 경로의 좌표와 GCP의 구성은 Figure 7과 같으며 그 좌표는 모두 1번 GCP를 원점으로 하여 변환한 상대 좌표이다. 계산 결과 또한 1번 GCP를 원점으로 한다. 실험 결과는 개체 매칭 결과와 이를 이용한 3D 위치 좌표 계산 결과를 나타냈다. 개체 매칭 과정을 위해서 먼저 기초행렬을 산정했으며 산정된 값을 개체 매칭과 카메라 파라미터 계산에 사용했다. 개체 매칭 과정에서는 Threshold를 변경하며 실험 데이터에 적합한 값을 확인하여 적용했다. 개체 매칭 결과는 각 Threshold 별로 모두 계산했지만 3D 위치 좌표 계산에는 최적 Threshold 값을 사용한 하나의 개체 매칭 결과만을 사용했다.

개체 검출 및 추적 과정은 Microsoft 社의 Visual Studio 2015를 이용해 C#으로 코딩했으며 컴퓨터 비전 라이브러리인 EmguCV Library가 사용됐다. 개체 매칭과 3D 위치 좌표 계산은 Mathworks 社의 MATLAB 2016a를 사용해 코딩했으며 기본 제공되는 빌트인(Built-in) 함수 외에도 Image processing toolbox, Computer vision toolbox가 사용됐다. 영상 촬영에는 Sony 社의 HandyCam HDR-CX450 두 대가 사용됐다. 자동 초점(AF) 기능을 지원하는 캠코더이지만 자동 초점 기능은 사용하지 않았으며 촬영 영역에 초점을 고정된 상태로 촬영했다. 촬영된 영상의 화면 비율은 16:9이며 해상도와 촬영 프레임은 1080×1920와 29.97fps이다. 영상은 별도의 프로그램을 이용하여 싱크를 맞췄으며 촬영 전의 실험 준비 과정 등을 편집하여 제외하고 1분 20초, 2495프레임이 사용됐다. 원활한 연산을 위해 이 중에서 10fps만을 사용하여 최종적으로는 1분 20초, 832프레임이 결과 계산에 사용됐다. 연산 시간은 21분가량 소요됐다.

3.1 기초행렬 계산

각 영상에서 추출된 특징점의 개수와 MSAC을 통해 획득한 Inlier의 개수는 Table 10에 표기했다. 추출된 전체 특징점 중 약 10%정도인 1,123개의 특징점들이 Inlier로 결정되었으며 이 Inlier들을 이용하여 기초행렬을 계산했다.

Table 1. The number of extracted feature points and inliers

	Left view image (#1 frame)	Right view image (#1 frame)
Extracted points	9,685	9,425
Inlier matched points	1,123	

계산된 기초행렬 F는 다음과 같다.

$$F = \begin{bmatrix} -2.484 \times 10^{-7} & -5.088 \times 10^{-5} & 0.012 \\ 7.582 \times 10^{-6} & 6.734 \times 10^{-6} & 0.377 \\ -0.003 & -0.335 & 0.873 \end{bmatrix}$$

3.2 파라미터 결정 및 개체 매칭 결과

3.2.1 파라미터 결정

Entity distance threshold(T_E)는 추적되는 개체의 영역의 크기를 고려하여 결정했다. 본 연구에서는 T_E 를 1부터 20까지 1씩 증가시키면서 정확도와 재현율을 계산하여 가장 좋은 결과를 얻을 수 있는 값을 사용하기로 결정했다. 그 결과를 Table 2에 나타냈다. 정확도는 평균적으로 98% 이상으로 나타나고 있으며 T_E 가 3을 넘게 되면 큰 차이를 보이지 않는 것을 확인할 수 있다. 하지만 재현율은 T_E 를 낮게 설정할 경우 50% 미만의 결과를 나타낼 수도 있으며 10 이상의 값에서부터 완만한 상승을 보이고 있다. 16부터는 정확도와 재현율 모두에서 일정한 결과를 유지하는 것을 확인할 수 있다. 본 연구에서는 정확도와 재현율 모두에서 최댓값을 갖는 16으로 T_E 를 설정했다.

Table 2. Precision/Recall of entity distance thresholding

Threshold	Precision (%)	Recall (%)	Threshold	Precision (%)	Recall (%)
1	97.39	16.89	11	99.02	91.61
2	98.54	38.32	12	99.02	91.84
3	99.00	55.90	13	99.03	92.52
4	99.17	67.57	14	99.03	92.74
5	99.26	76.19	15	99.03	92.86
6	98.93	83.56	16	99.03	93.08
7	98.97	87.41	17	99.03	93.08
8	98.99	88.89	18	99.03	93.08
9	99.01	90.36	19	99.03	93.08
10	99.01	90.93	20	99.03	93.08

Table 3. Precision/Recall of entity distance ratio thresholding

Threshold	Precision (%)	Recall (%)	Threshold	Precision (%)	Recall (%)
1,0	99,01	90,93	5,0	99,57	78
1,5	99,25	90,02	5,5	99,55	75,62
2,0	99,37	89,12	6,0	99,54	73,36
2,5	99,36	87,98	6,5	99,52	70,98
3,0	99,35	86,28	7,0	99,51	69,16
3,5	99,6	84,13	7,5	99,49	66,89
4,0	99,59	82,09	8,0	99,48	65,65
4,5	99,58	80,27			

다음은 결정된 $T_E=16$ 을 사용해 적절한 Entity distance ratio threshold(R_E)를 결정해야 한다. R_E 를 1부터 8까지 0.5씩 변화시켜 결과를 계산했다. 결과는 Table 3과 같다. 정확도는 $R_E=3.5$ 이후 큰 변화가 없으며 재현율은 Threshold가 높아짐에 따라 서서히 하락하는 것을 확인할 수 있다. 가장 높은 정확도를 갖는 값은 3.5이며 해당 값에서 재현율이 85%정도로 낮아졌지만 충분히 납득할만한 수준으로 판단했으며 가장 높은 정확도를 갖는 Threshold를 사용하는 것이 3차원 위치 좌표 계산에서 더 낮은 오차를 확인할 수 있을 것으로 예상하여 최종적으로 $T_E=16$, $R_E=3.5$ 로 결정했다.

3.2.2 개체 매칭 결과

결정된 T_E 와 R_E 를 이용하여 개체 매칭을 실시했으며 개체 매칭에는 1분 20초, 2495 프레임의 29.97fps로 촬영된 영상이 사용되었고, 이 중에서 10fps만을 결과 계산에 사용했다. 따라서 총 832개의 계산 결과가 생성된다. 832개의 프레임 중 실제 매칭될 수 있는 매칭 쌍은 총 882개이다. 추적 대상 개체는 Figure 8과 같이 각각 E1과 E2로 표기했다. Table 4는 개체 매칭 결과를

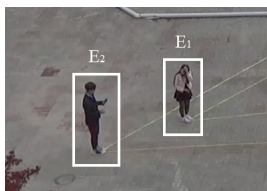


Figure 8. Target entities

Table 4. Entity matching results

Expected matching (TP+FN)	882
Counted matching (TP+FP)	745
True positive (TP)	741
False positive (FP)	1
False negative (FN)	3
Precision (%)	84.13
Recall (%)	99.60

정리한 것이다. 총 745개의 매칭이 이루어졌으며 그 중에서 매칭 오류로 판단할 수 있는 것은 4개로 확인됐다. 이에 따라 정확도와 재현율은 각각 99.60%와 84.13%로 계산됐다.

3.3 3D 위치 좌표 계산 결과

3D 위치 좌표 계산 결과는 최적 Threshold로 결정된 $T_E=16$, $R_E=3.5$ 로 설정한 상태에서 진행됐다. 개체 추적 결과 반환되는 개체 중심점은 추적 대상 인물의 가슴 높이에 위치한다(Lee et al., 2016). 따라서 검증을 위해 수집된 경로의 실측 위치 좌표와 추적 개체의 계산된 위치 좌표의 정확한 비교를 위해서는 대상 인물들의 가슴 높이를 고려해야 한다. E1의 키는 H_1 , E2의 키는 H_2 로 정의했으며 $H_1=1,580m$ $H_2=1,780m$ 이다. 개체 추적 과정에서 개체의 상반신을 추적하도록 설정되었고, 그 중심 좌표를 개체 중심점으로 반환하기 때문에 계산을 통해 생성되는 개체 중심점의 좌표는 지면에서 연직 방향으로 인물의 키의 3/4 지점에 위치하게 된다. 3D 위치 좌표 계산 결과는 다음 Table 5, Figure 9와 같다.

3D 거리에서 최대 오차는 2,213m로 나타났으며 최소 오차는 0m이다. 개체의 가슴 높이를 추적한 결과라는 점을 감안하더라도 3m 내의 최대 오차를 갖는 것은 기존 GPS 태그와 비교해볼 만한 가치가 있다고 사료된다. 또한, 최대 오차가 발생하는 지점은 개체 매칭 과정에서 발생한 오류 데이터가 생성된 지점으로, 만약 개체 추적 결과가 100%의 정확도를 갖는다면 이보다 적은

Table 5. 3D tracking results

Entity index	E1 Trajectory			E2 Trajectory		
	3D	2D (XY)	Vertical (Z)	3D	2D (XY)	Vertical (Z)
Counted matching	371			374		
Maximum distance (m)	1,992	1,622	1,430	2,213	1,112	1,913
Minimum distance (m)	0,941	0,000	0,828	0,996	0,007	0,976
Average (m)	1,107	0,331	1,040	1,270	0,257	1,230
Standard deviation (m)	0,136	0,208	0,108	0,108	0,183	0,111

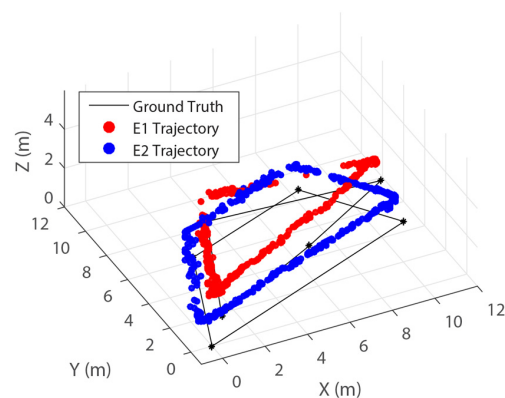


Figure 9. 3D tracking results

최대 오차를 가질 수 있을 것으로 생각된다. 수직 거리의 평균 오차는 E_1 의 경우 1.04m, E_2 의 경우 1.23m로 계산됐다. 본 연구에서 개체의 가슴 높이는 0.75H로 정의했으며 $H_1=1.580m$, $H_2=1.780m$ 일 때, $0.75H_1=1.185m$, $0.75H_2=1.335m$ 이다. 이와 비교했을 때 수직 거리의 평균 오차는 납득할만한 수준이라고 생각된다.

4. 결론 및 향후 연구 진행 방향

영상을 이용한 건설 현장 모니터링 방법에 대한 연구가 활발하게 진행 중이지만 대부분 2D에 국한되어 있으며 건설 현장 개체의 3D 위치 좌표를 획득하기 위한 연구는 그 필요성에 비해 연구 결과가 다소 부족했으며 다수의 개체를 추적하지 못하는 한계가 있었다. 본 연구에서는 이를 개선하기 위해 등극선 기하학을 적용함으로써 체커보드를 사용해야 했던 기존의 카메라 캘리브레이션 과정을 개선하고, 개체 매칭 과정을 추가하여 다수의 건설 인력을 추적할 수 있는 방법을 제안했으며 그 적용 가능성을 확인해보고자 인물을 대상으로 촬영된 스테레오 영상을 이용하여 개체 매칭과 3D 위치 좌표 계산 결과의 정확도를 확인했다.

스테레오 영상 픽셀 좌표 간의 기하학적 상관관계를 나타내는 기초행렬은 카메라 행렬 계산을 위해 카메라 캘리브레이션 과정에서 사용됐으며 개체 매칭 과정에서는 에피플라 라인 계산을 위해 사용됐다. 개체 매칭 과정에서 영상의 픽셀 좌표를 이용함으로써 발생할 수 있는 문제들을 해결하기 위해서 Entity distance와 Entity distance ratio에 대해 Threshold(T_E , R_E)를 지정했으며 그 값의 변화에 따른 개체 매칭 결과를 통해 최적 값을 결정했다. 본 연구의 실험에 사용된 영상 데이터에 대해서는 $T_E=16$, $R_E=3.5$ 가 가장 적합한 것으로 확인됐다. 이 Threshold는 사용되는 영상 데이터에 따라 달라질 수 있기 때문에 그 값을 결정하기 위한 과정이 반드시 필요할 것으로 사료된다. 최적 Threshold를 이용한 개체 매칭 과정의 정확도는 99% 이상으로 확인됐다. 제안한 개체 매칭 방법의 3D 위치 좌표 계산에 대한 적용 가능성을 판단하고자 개체 매칭을 통해 계산된 각 개체의 대응점을 3D 위치 좌표 계산에 사용했다. 계산된 3D 위치 좌표는 평균적으로 0.35m 정도의 오차를 보이는 것으로 확인됐다.

영상으로부터 3D 위치 좌표를 계산하기 위해서는 동일한 점에 대한 스테레오 영상에서의 대응점을 필요로 한다. 기존 연구 사례에서는 이 과정이 수동적으로 시행되었기 때문에 영상을 이용하여 다수의 개체를 추적하기에는 한계가 있었다. 본 연구에서 제안한 개체 매칭 방법은 연산 처리장치를 통해 자동으로 이루어지며 정확도가 매우 높아 다수 개체를 추적하기 위한 개체 매칭 방법으로 적합하다고 사료된다. 3D 위치 추적 결과 또한

시중에서 판매되는 GPS 태그의 정확도와 비교했을 때 상당히 정확하다고 판단할 수 있다. 또한, 위성 신호가 닿지 않는 터널이나 실내 등의 환경에서도 사용할 수 있다.

하지만 개체 추적, 개체 매칭 과정에서 정확한 대응점을 반환하지 않을 경우에는 정확한 3D 위치 좌표를 계산할 수 없다. 본 연구에 사용된 개체 추적 방법은 추적 대상 개체의 색상 정보를 이용하기 때문에 두 개체가 일부 겹쳐지거나 비슷한 색상 히스토그램을 갖는 개체가 영상에서 검출될 경우 잘못된 영상 좌표를 반환할 수 있다. 개체 매칭 과정에서는 하나의 에피플라 라인에 다수의 개체가 근접할 때 이런 문제가 발생할 수 있다. 개체 매칭 과정에서는 이 문제를 해결하기 위해 두 가지 Threshold를 적용했고, 실험 데이터에 대해서는 납득할만한 결과를 보여주고 있다. 하지만 실험에 사용된 영상 데이터는 두 명의 인물을 대상으로 촬영되었기 때문에 개체가 겹쳐지거나 에피플라 라인에 둘 이상의 개체가 위치하는 경우는 비교적 적게 발생하게 된다. 실제 건설 현장에서는 이보다 더 많은 개체를 추적할 수 있어야 한다. 따라서 더 복잡한 환경에서 촬영된 영상 데이터를 이용하여 확인이 필요할 것으로 사료된다. 이외에도 다른 영상을 이용하게 될 때 최적 Threshold를 결정해야 하는 문제가 있으며 3D 위치 좌표 계산을 위해서는 정확한 GCP와 그에 대한 영상 대응점을 사전에 획득해야만 하는 한계가 있다. 또한, 카메라를 이용하기 때문에 시야각에 따라 촬영 가능한 영역이 제한되며 영상의 해상도와 화질에 따라 촬영 가능한 거리가 제한될 수 있어 매우 넓은 환경에서 사용하기 위해서는 추가적인 카메라의 설치가 필요하다.

영상 기반의 위치추적 시스템은 GPS에 비해 비용 및 설치시간 측면에서 이점을 가질 수 있지만 개체가 시야에서 사라진 동안에는 작동이 불가하다는 점이 GPS 대비 큰 단점이 될 수 있다. 하지만, 건설현장의 한정된 일정 공간 내에서의 위치 추적 및 이동 패턴 분석을 위한 용도로 활용될 수 있는 충분한 잠재적 가치가 있으며, 이는 특정 작업에 투입된 인력, 장비들의 생산성 분석이나 안전 점검을 목적으로 현장에 적용해 볼 수 있을 것이다. 현장조건에 따라 최적의 카메라 배치에 대한 연구가 함께 이뤄져야 할 것으로 사료되며, 전체 시스템을 구성하는 각각의 요소별로 새로운 알고리즘들을 적용해보면서 전체 시스템의 정확도 및 안정성을 향상시키는 노력이 지속적으로 이뤄질 필요가 있다.

References

Bay, H., Ess, A., Tuytelaars, T., Van Gool, L. (2008). Speeded-Up Robust Features (SURF), Computer Vision and Image Understanding, 110(3), pp. 346-359.

- Carbonari, A., Giretti, A., Naticchia, B. (2011). A proactive system for real-time safety management in construction sites, *Automation in Construction*, 20(6), pp. 686-698.
- Cheng, T., Venugopal, M., Teizer, J., Vela, P. A. (2011). Performance evaluation of ultra wideband technology for construction resource location tracking in harsh environments, *Automation in Construction*, 20(8), pp. 1173-1184.
- Domdouzis, K., Kumar, B., Anumba, C. (2007). Radio-Frequency Identification (RFID) applications: A brief introduction, *Advanced Engineering Informatics*, 21(4), pp. 350-355.
- Hartley, R., Zisserman, A. (2003). Multiple View Geometry in Computer Vision.
- Lee, Y., Park, M., Brilakis, I. (2016). Entity Matching across Stereo Cameras for Tracking Construction Workers, *Proceedings of the International Symposium on Automation and Robotics in Construction 2016 (ISARC 2016)*, 33, pp. 1-8.
- Li, H., Chan, G., Wong, J. K. W., Skitmore, M. (2016). Real-time locating systems applications in construction, *Automation in Construction*, 63, pp. 37-47.
- Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints, *International Journal of Computer Vision*, 60(2), pp. 91-110.
- Park, M. W., Brilakis, I. (2012). Construction worker detection in video frames for initializing vision trackers, *Automation in Construction*, 28, pp. 15-25.
- Seo, J., Han, S., Lee, S., Kim, H. (2015). Computer vision techniques for construction safety and health monitoring, *Advanced Engineering Informatics*, 29(2), pp. 239-251.
- Torr, P. H. S., Murray, D. W. (1997). The Development and Comparison of Robust Methods for Estimating the Fundamental Matrix, *International Journal of Computer Vision*, 24(3), pp. 271-300.
- Turkan, Y., Bosche, F., Haas, C. T., Haas, R. (2012). Automated progress tracking using 4D schedule and 3D sensing technologies, *Automation in Construction*, 22, pp. 414-421.
- Vasenev, A., Pradhananga, N., Bijleveld, F. R., Ionita, D., Hartmann, T., Teizer, J., Dorée, A. G. (2014). An information fusion approach for filtering GNSS data sets collected during construction operations, *Advanced Engineering Informatics*, 28(4), pp. 297-310.