AA033

Correlating multiple 2D vision trackers for 3D object tracking on construction sites

Man-Woo Park
(Ph.D. Candidate, School of Civil & Environmental Engineering, Georgia Institute of Technology, Atlanta, USA)

Christian Koch
(Postdoctoral Research Fellow, School of Civil & Environmental Engineering, Georgia Institute of Technology, Atlanta, USA)

Ioannis Brilakis
(Assistant Professor, School of Civil & Environmental Engineering, Georgia Institute of Technology, Atlanta, USA)

Abstract

Tracking methods have the potential to retrieve the spatial location of project related entities such as personnel and equipment at construction sites, which can facilitate several construction management tasks. Existing tracking methods are mainly based on Radio Frequency (RF) technologies and thus require manual deployment of tags. On construction sites with numerous entities, tags installation, maintenance and decommissioning become an issue since it increases the cost and time needed to implement these tracking methods. To address these limitations, this paper proposes an alternate 3D tracking method based on vision. It operates by tracking the designated object in 2D video frames and correlating the tracking results from multiple pre-calibrated views using epipolar geometry. The methodology presented in this paper has been implemented and tested on videos taken in controlled experimental conditions. Results are compared with the actual 3D positions to validate its performance.

Keywords

Vision tracking, epipolar geometry, 3D positioning

1. Introduction

3D object tracking on construction sites has a wide variety of applications. Being able to effectively identify and track personnel, equipment and materials can effectively support progress monitoring, activity sequence analysis, productivity measurements, asset management as well as enhancing site safety. In addition, real time tracking instantly enables the identification of critical activities and problems, which allows for in-situ project control and decision making capabilities. Available tracking solutions are mainly based on Radio Frequency technologies including Global Positioning System (GPS), Radio Frequency Identification (RFID), Wireless Networks (Wi-Fi) and Ultra Wideband (UWB) technologies. They all work under the same principle of having a sensor attached on each entity to be tracked. In outdoor and congested construction sites (e.g. highway construction), due to the large amount of items involved, tracking technologies that require attaching

sensors to each single item have additional time and cost burdens associated with performing sensor installations.
The research, part of which is presented in this paper, proposes an alternate tracking methodology based on vision. Under this method, video-streams are collected from constructions sites and then used to determine the spatial location of project related entities across time. In order to achieve this, the first step is to compare the performance of existing 2D tracking methods. It has turned out that kernel-based methods are most suitable for construction related applications (Makhmalbaf et al., 2010). Using a selected kernel-based 2D tracking algorithm (Ross et al., 2008), the 2D locations of an entity are determined and recorded for corresponding video frames. Since 2D locations are not sufficient to be used for construction management tasks, the next step is to correlate corresponding views in order to calculate depth values. For this purpose, epipolar geometry is used to provide the 3D aspect, which allows the user to retrieve 3D location information in real time. Preliminary results were obtained by testing this method on a scaled model of a highway construction site at the Construction Information Technology Laboratory. The results are compared with the actual 3D position in order to validate that the method proposed can effectively provide accurate localization of construction site entities.

2. Object Tracking

2.1 Current Practice in 3D Tracking

Common tracking methods are based on Radio Frequency and include several types of technologies like Global Positioning Systems (GPS), Radio Frequency Identification (RFID), Bluetooth and Wireless Fidelity (Wi-Fi, Ultra-Wideband, etc). They have been successfully used when tracking prefabricated materials (Song et al. 2006), equipment, inventory (Caldas et al. 2004) and personnel (Teizer 2007). RFID technology does not require line-of-sight (as opposed to vision-based methods), and also it is durable to harsh environments and can be embedded in concrete. Reading range depends on the frequency at which the tag operates, and it varies from several inches up to about 10 feet. Unlike barcodes, reading range is not a fixed distance, allowing tags to be read at any distance within the range. In addition, RFID enables efficient automatic data collection since readers can be mounted to any structure to detect and each reader can scan multiple tags at a given time.

However, all RFID technologies face several limitations, which restrict their applicability in outdoor and congested construction sites. The main disadvantage of frequency based systems is the need for installing a sensor in each entity before any type of tracking information can be acquired. Due to the large number of entities that need to be tracked in congested construction sites, in most cases it is infeasible to attach a sensor on each entity since the cost associated with this work and equipment becomes considerable. Moreover, RFID technology, unless combined with other tools (Ergen et al, 2007), can only report the radius inside which the tracked entity exists, and the near-sighted effect (the location is not reported until the reader reaches the radius of the reading range) prohibits its use in real-time tracking applications. Another important aspect is privacy. Since RFID tags have to be attached to persons, who are to be tracked, this method is not favored by Construction Unions due to the lack of privacy they provide workers on the site (Juels 2006).

2.2 Vision based 2D Tracking

Vision based 2D tracking can be a proper alternative to RFID methods because it removes the need for installing sensors and ID tags of any kind on the tracked entity. For this reason, this technology is (a) highly applicable in dynamic, busy construction sites, where large numbers of equipment, personnel and materials are involved, and (b) more desirable from personnel who wish to avoid being "tagged" with sensors. In Gruen (1997) it is highly regarded for its capability to measure a large number of particles with a high level of accuracy. Generally, vision tracking fuses video cameras and computer vision algorithms to perform several measurement tasks and has been considered as an inference problem (Forsyth and Ponce, 2002). The moving entity has a certain internal state which is calculated in each frame. The tracking algorithms then need to combine these calculations to estimate the state of the entity. Information such as the speed, acceleration, flow and periodic motion of the entity can then be inferred.

Vision tracking methods can be categorized in kernel-based, contour-based, and point-based methods, depending on the way of representing objects. In kernel-based methods, an object is represented by the color or texture in the region of interest, and its position in next frame is estimated based on the region's color or texture information. In contour-based methods, an object is represented by silhouettes or contours that determine the boundary of the object. In point-based methods, an object is represented by a set of feature points extracted from the region that contains the object. Out of the three categories, kernel-based methods are most suitable for construction related applications regarding the construction sites' characteristics such as illumination condition and object types on construction sites (Makhmalbaf et al., 2010). Compared to the kernel-based methods, point-based methods (Mathes et al., 2006) are prone to lose the object whenever it gets difficult to extract sufficient feature points (e.g. when it is very dark or bright). Also, contour-based methods (Vaswani et al., 2010) do not provide a consistent centroid of the object.

3. Vision based 3D Tracking

2D vision tracking results, which just present 2D object positions in an image, are not sufficient to be used for construction management tasks due to the complexity and large scale of the construction sites. To obtain meaningful 3D location information, an additional process of integrating 2D vision tracking results from multiple camera perspectives is required. The objective of this work is to accurately determine the 3D location of distinct construction related objects, such as equipment, personnel and materials of standard shape and size across time.

3.1 Methodology

The tracking methodology proposed in this paper consists of a computer receiving video feeds collected from on site video cameras. The method can be used with several cameras connected simultaneously. These cameras must have at least partially overlapping views of the site object that is to be tracked in order to calculate the 3D location of the entity. After receiving the video feed, the cameras' views are projected in the user's window simultaneously. The user can select an entity of interest in both of the views. Selecting an area corresponding to the object (e.g. safety vest or hardhat) is enough to initialize tracking.

The method then calculates the selected area's centroid, as a representation of the object's 2D location on the image frame. Subsequently, a kernel-based tracking algorithm (Ross et al., 2008) is used to track the areas identified in each of the views. The object tracking is performed simultaneously and independently in each camera. This feature allows the user to locate and track the object in real time across the entire site simultaneously. The result of the 2D tracking algorithm is the 2D position of the object's centroid in each window. Subsequently, epipolar geometry is used to calculate the depth of the centroid that determines the 3D location of the object in each frame. The system records the tracking information and displays the 3D coordinates of the entity in the user's window. The methodology for this algorithm is illustrated in Figure 1.

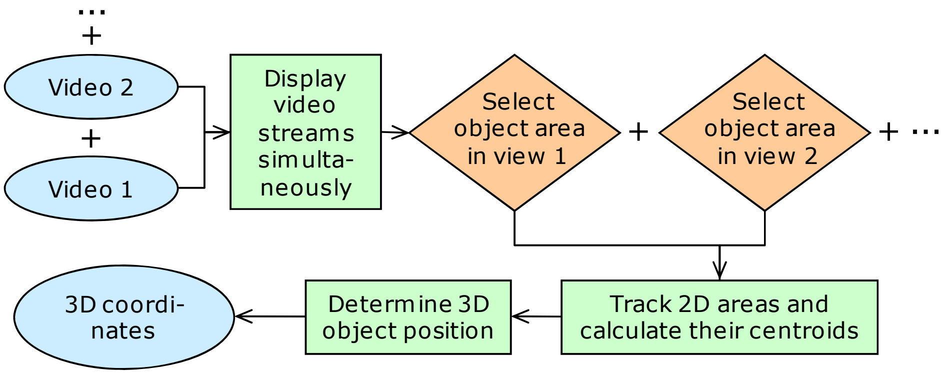
...
+
Video 2
+
Video 1
Display video streams simulta-neously
Select object area in view 1
+
Select object area in view 2
+ ...
Track 2D areas and calculate their centroids
Determine 3D object position
3D coordi-nates

Fig 1: Methodology overview

3.2 Object Selection and 2D Tracking

The identification of the object area of interest is performed interactively by the user. The user selects the same object area in both of the views presented to him by marking it on the screen. The created rectangular areas are used to determine the object location in the image planes by simply calculating the centroids (centers of gravity) of the marked regions. Using the 2D tracking algorithm described above, the object location is recalculated in each subsequent frame and updated as the marked object moves. The result of the 2D object tracking is the real-time object location in each camera's view.

3.3 3D Location Calculation

The 2D tracking results are used to determine the real-time 3D object location based on epipolar geometry. In order to perform epipolar geometry calculations, at least two cameras must be present and their views must overlap in the region that contains the object of interest. Based on the overlapping regions, the 8-point-algorithm is used to determine the fundamental matrix, which geometrically relates corresponding points (P_L, P_R) in stereo images (Fig. 2). The fundamental matrix enables full reconstruction of the epipolar geometry, since it encodes information about both intrinsic and extrinsic parameters of the cameras views. The centroid's coordinates of the object are used to calculate the epipolar lines in each of the camera's views. These epipolar lines are necessary to find the 3D location of the object. Since the centroid of the object is known on both camera views, the projection of that point across the epipolar plane in both of the views leads to an intersecting point P in 3D space (Fig. 2). This intersection point contains the depth coordinate of the object's centroid. This procedure provides 3D perspective reconstruction and allows for the calculation of the 3D object location (P) in every frame. For a more detailed description of epipolar geometry, the 8-point-algorithm, and the fundamental matrix, the interested reader is referred to Hartley and Zisserman (2003).

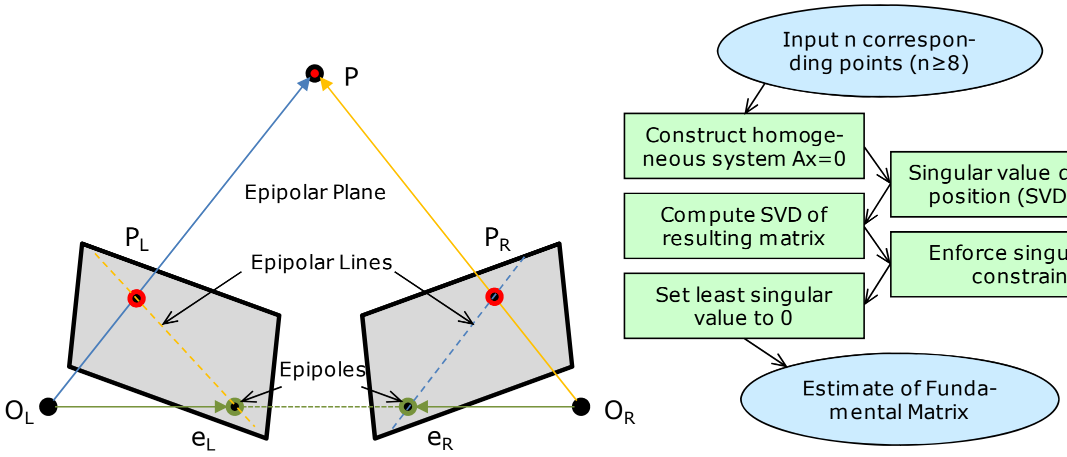
P
Epipolar Plane
P_L
P_R
Epipolar Lines
Epipoles
O_L
O_R
e_L
e_R
Input n correspon-ding points (n≥8)
Construct homoge-neous system Ax=0
Singular value decom-position (SVD) of A
Compute SVD of resulting matrix
Enforce singularity constraint
Set least singular value to 0
Estimate of Funda-mental Matrix

Fig 2: Epipolar Geometry for two cameras O_L and O_R (left) and 8-point-algorithm for Fundamental Matrix calculation (right)

4. Results

The method presented in this paper has been implemented and tested on videos taken in controlled experimental conditions. Results are compared with the actual 3D positions to validate its performance.

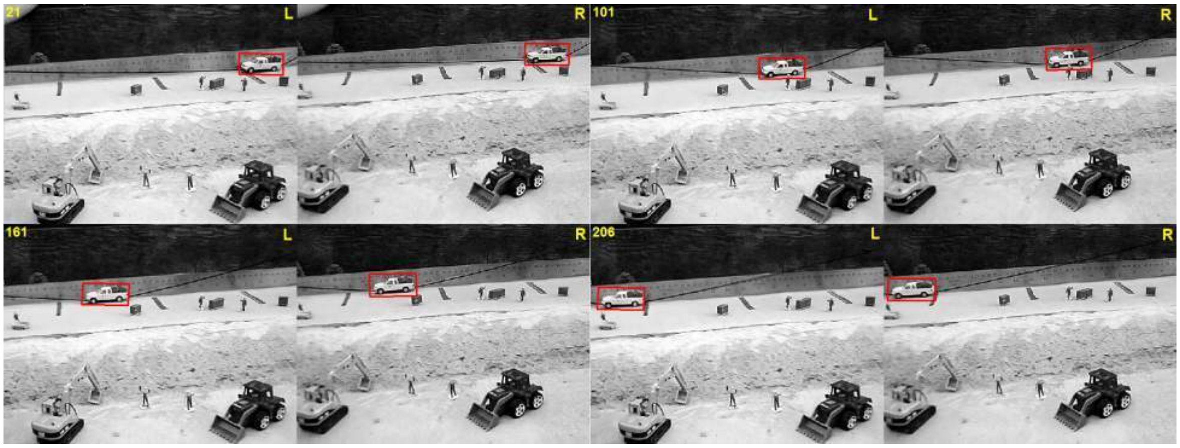

Fig 3: 2D tracking sequence for left (L) and right (R) camera view (frame no. 21, 101, 161 and 206)

Although the method can be implemented with several cameras simultaneously, the basic example (2 cameras) was used for experimental purposes and to reduce the computational load of the system. In order to minimize the error of manually synchronizing a pair of frames according to time, it was decided to use a stereo USB webcam. This webcam by default records synchronized frame sequences. The stereo videos taken contain 191 frames (~ 13 sec) each at a resolution of 640x480

pixels and a frame rate of 15 fps. The highway construction site model used for testing is of scale 1:87. The moving object to be tracked was a small truck at a size of 6.5 mm length, 2.2 mm width and 2.2 mm height (Fig. 3).
The actual positions of the truck were determined by a start point Q_1 and an end point Q_2 that were measured manually. For simplicity reasons it was intended and assumed that the truck moved straight along the line L defined by Q_1 and Q_2. Figure 3 presents example video frames showing the model environment and the tracked truck object.
Based on the corresponding 2D centroid positions P_L and P_R, the 3D location P of the truck was determined for every video frame. The performance results are quantified by an absolute error E_{ABS} that is defined as the distance between the determined 3D position P and the line $L = Q_1 + t\,(Q_2 - Q_1)$.

$$E_{ABS} := |(Q_2 - Q_1) \times (P - Q_1)| \;/\; |(Q_2 - Q_1)| \qquad (1)$$

Figure 4 presents the overall performance results by plotting the absolute errors $E_{ABS}(i)$ against the frame numbers i = 21..211. The graph shows the maximum error E_{MAX}= 105.4 mm at frame #127 and the average error E_{AVG} = 47.8 mm.

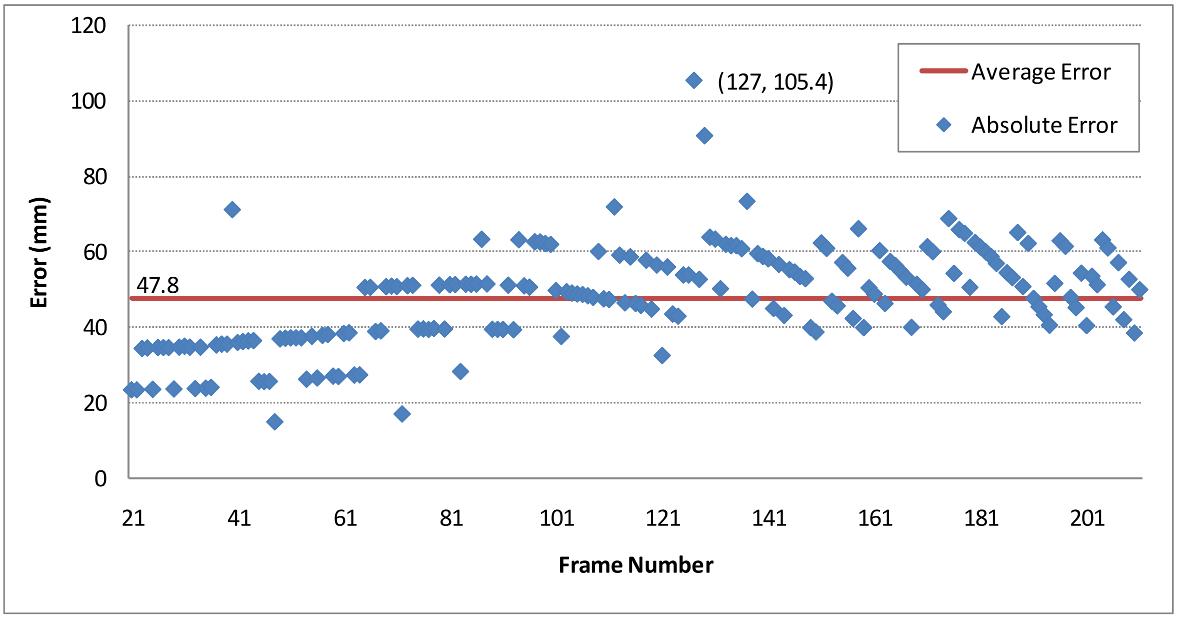

Fig 4: 2D tracking sequence for left (L) and right (R) camera view (frame no. 21, 101, 161 and 206)

Due to the model scale of the test environment the determined errors have to be scaled by factor 87. This yields a maximum error $\bar{E}_{MAX}$ = 9.17 m that obviously can compete stand-alone GPS with an accuracy of around 10 m (Caldas et al. 2004). Moreover, scaling the average error yields $\bar{E}_{AVG}$ = 4.16 m, which is less than the actual length of the tracked object. Based on these error magnitudes it can be stated that the vision based 3D tracking approach presented was successfully tested and verified.
However, it is believed that the first results presented can be improved. Considering frame #127, which is related to the maximum error E_{MAX}, it was recognized that the determined centroid positions P_L and P_R differ to a significant extent. Figure 5 presents the left and right view of frame #127 highlighting the rectangular areas used for 2D tracking and their centroids. The noticed difference is about 3 to 4 pixels according to a fixed position (e.g. bottom right corner of the side window). Misarranged coordinates determined in the triangulation process

result in an incorrect depth value for the 3D location of P. It is concluded, that the position of the selected rectangular area has a significant impact on 3D tracking results. Based on this finding, it is proposed to specify the area by selecting its centroid, rather than by selecting the area's exterior shape and determining the centroid afterwards. Furthermore, the 2D tracking results can be enhanced by utilizing high resolution cameras, which directly affects the accuracy of the 3D tracking result.

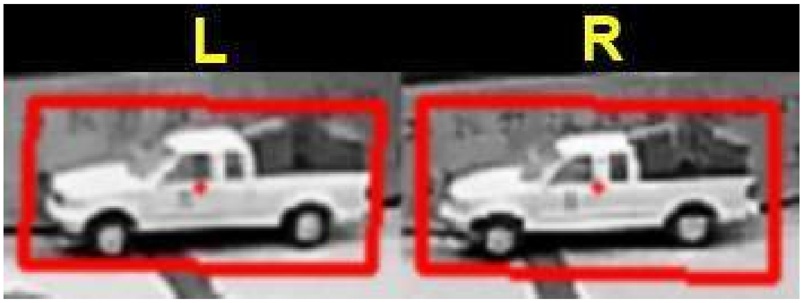

Fig 5: Differing centroid position P_L and P_R after 2D tracking in video frame #127

5. Conclusions

3D object tracking on construction sites has a wide variety of applications. Being able to effectively identify and track personnel, equipment and materials can support progress monitoring, activity sequence analysis, productivity measurements, asset management as well as enhancing site safety. Available tracking solutions based on Radio Frequency technologies have been successfully used when tracking prefabricated materials, equipment, inventory and personnel. However, in open construction sites (e.g. highway construction), due to the large amount of items involved, tracking technologies that require attaching sensors to each single item have additional the time and cost burdens associated with performing sensor installations.

This paper presented an alternate tracking methodology based on vision tracking, which removes the need for object tagging. Under this method, video-streams are collected from constructions sites and then used to determine the spatial location of project related entities across time. Using a selected kernel-based 2D tracking algorithm, the 2D locations of an entity are determined and recorded for corresponding video frames. Epipolar geometry is then used to correlate multiple views and to calculate the depth value, providing the 3D aspect and allowing the user to retrieve 3D location information in real time. Preliminary results were obtained by testing the method on a scaled model of a highway construction site at the Construction Information Technology Laboratory. The maximum error determined is comparable with the standard accuracy of stand-alone GPS, and the average error is even less than the maximum dimension of the tracked object. It was found that the errors are reasonably attributed to incorrect 2D tracking in terms of an inaccurate determination of object centroids. However, the results presented validate that the vision based 3D tracking approach proposed can effectively provide accurate localization of construction site entities. The next step is to investigate how visual pattern recognition methods can be used to automatically recognize and match entities removing the need for manual entity selection.

6. Acknowledgements

The presented work is based upon a research funded by the National Science Foundation Grant # 0933931. NSF's support is gratefully acknowledged.

7. References

Caldas, C.H., Torrent, D.G., and Haas, C.T. (2004), "Integration of automated data collection technologies for real-time field materials management", 21st International Symposium on Automation and Robotics in Construction, Jeju, Korea

Ergen, E., Akinci, B., and Sacks, R. (2007), "Tracking and locating components in a precast storage yard utilizing radio frequency identification technology and GPS", Automation in Construction, 16(3), 354-367

Forsyth, D.A., Ponce, J. (2002), "Computer Vision: A Modern Approach", Prentice Hall

Gruen, A. (1997), "Fundamentals of videogrammetry – A review", Human Movement Science Journal, 16, 155-187.

Hartley, R. and Zisserman, A. (2003). "Multiple View Geometry in computer vision", Cambridge University Press

Juels, A. (2006), "RFID Security and Privacy: A Research Survey", IEEE Journal on Selected Areas in Communications, 24(2), 381-394

Makhmalbaf, A., Park, M.-W., Yang, J., Brilakis, I., and Vela, P. A. (2010), "2D Vision Tracking Methods' Performance Comparison for 3D Tracking of Construction Resources", Construction Research Congress 2010, Banff, Canada

Mathes, T. and Piater, J.H. (2006), "Robust Non-Rigid object Tracking Using Point Distribution Manifolds", 28th DAGM (Deutsche Arbeitsgemeinschaft für Mustererkennung), Berlin, Germany

Ross, D., Lim, J., Lin, R.-S., and Yang, M.-H. (2008), "Incremental Learning for Robust Visual Tracking", Int. J. Comput. Vis., 77, 125-141.

Song, J., Haas, C., Caldas, C., Ergen, E., and Akinci, B. (2006), "Automating Pipe Spool Tracking in the Supply Chain", Automation in Construction, 15/2, 166-177

Teizer, J., Lao, D., and Sofer, M. (2007), "Rapid Automated Monitoring of Construction Site Activities Using Ultra-Wideband", 24th International Symposium on Automation and Robotics in Construction (ISARC 2007). Construction Automation Group, I.I.T. Madras

Vaswani, N., Rathi, T., and Yezzi, A. (2010), "Deform PF-MT: Particle Filter With Mode Tracker for Tracking Nonaffine Contour Deformations", IEEE Trans. Image Processing, 19(4) 841-857.