Rapid Image-based Localization using Clustered 3D Point Cloud Models with Geo-Location Data for AEC/FM Mobile Augmented Reality Applications

Hyojoon Bae[1], Mani Golparvar-Fard[2] and Jules White[3]

[1]Bradley department of Electrical and Computer Engineering, Virginia Tech, Blacksburg, VA, 24060; email: hjbae@vt.edu
[2]Departments of Civil and Environmental Eng. and Computer Science, University of Illinois at Urbana-Champaign, Urbana, IL, 61801; email: mgolpar@illinois.edu
[3]Department of Electrical Engineering and Computer Science, Vanderbilt University, Nashville, TN, 37235; email: jules.white@vanderbilt.edu

ABSTRACT

In this paper, we present a new method for supporting onsite construction and facility management tasks by allowing field personnel to automatically have access to the latest project information in form of Augmented Reality (AR) overlays, visually document onsite issues/progress, and communicate information with other personnel on or off site. Our near real-time and marker-less mobile augmented reality solution builds on top of a new image-based localization method for 3D point clouds that have been reconstructed using a Structure-from-Motion (SfM) pipeline and are clustered based on already available geo-location data. By using images captured from commodity smartphones/tablets, our method computes a precise 6-DOF pose for the camera and delivers relevant project information in form of AR overlays. Our main contributions lie in efficient clustering of 3D point clouds and rapid computation of camera pose by detecting an appropriate cluster of 3D points. Compared to our previous work for AEC/FM mobile augmented reality applications, the experimental results demonstrate that the proposed clustering approach accelerates image-based localization using 3D point clouds, taking 1-2 seconds for a single localization.

INTRODUCTION

Onsite information management is indispensable to successful operations of construction and facility management field activities. Inexpensive and prompt access to reports of inspection, management, and/or specifications facilitates the identification, processing, and communication of quality control issues (Golparvar-Fard et al. 2012; Chen and Kamara 2011). It further enables engineers to proactively decide on corrective actions and minimize the excessive cost and schedule delays in managing the construction or operation of the facilities (Bae et al. 2012; Kim et al. 2013). Until recently, the commonly accepted practice for onsite information management involved manual/monotonous data collections, non-systematic analysis, and visually complex reporting (Golparvar-Fard et al. 2011; Navon and Sacks 2007).

Over the past decade, the advent of smart mobile devices, such as smartphones or tablets, has provided a great opportunity to improve existing practices. A recent survey conducted by McGraw Hill reveals that 93% of a representative sample of general contractors and subcontractors are now using mobile devices on their jobsites to document workflows (ENR 2012). Kim et al. (2013) have summarized that to improve productivity of onsite operations, any onsite information management system should: 1) enable project monitoring capabilities (Golparvar-Fard et al. 2011; Retik et al. 2002), 2) provide easy access to relevant information so that onsite resources could be managed more effectively (Bae et al. 2013a,b; Son et al. 2012), and 3) function in near real-time to share information and facilitate interactions among project participants (Bowden et al. 2006; Aziz et al. 2005).

To achieve these functionalities in form of Augmented Reality (AR), we have proposed a vision-based mobile augmented reality system which identifies location and orientation of mobile devices solely based on a visual features extracted from site photographs (Bae et al. 2013a,b; Bae et al. 2012). This marker-less and infrastructure-independent system, called HD^{4}AR (Hybrid 4-Dimensional Augmented Reality), provides high-precision information retrieval in near real-time without requiring external sensors (e.g., geomagnetic or inertial sensors) or environmental constraints (e.g., GPS satellites or wireless access points), as shown in Figure 1. In our recent work (Bae et al. 2013a,b), we also presented a method for 3D annotation from a single 2D image taken by commodity smart devices to easily and quickly associate project information with actual 3D physical elements in the scene, and thus enable real-time site monitoring and convenient association of project information with real-world 3D building and civil infrastructure components on all commodity smartphones and tablets. Although the latest HD^{4}AR system achieves millimeter-level accuracy and near real-time localization/augmentation using 3D point cloud reconstructed from site photographs, the localization performance of HD^{4}AR still depends on the scale of 3D point cloud, i.e. the number of 3D points in the point cloud. As a consequence, the larger scale of 3D point clouds can cause longer localization time.

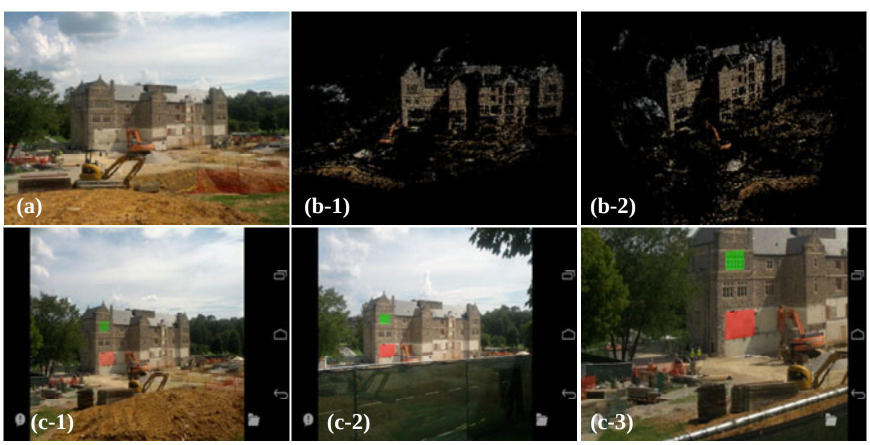

Figure 1. An example of HD^{4}AR application; (a) initial base images, (b) 3D point cloud, and (c) localization/augmentation results – the system precisely renders overlays with significant changes in viewpoint (adopted from Bae et al. 2013b).

This paper builds on our previous work on the HD^{4}AR and proposes a new 3D point cloud clustering scheme to further accelerate localization speed of HD^{4}AR. The approach segments the large-scale 3D point cloud into several smaller point clouds using geo-location data, i.e. 3D site coordinates values from GPS, which can be easily obtained with modern commodity mobile devices. Then, the image-based and model-based localization of HD^{4}AR utilizes these smaller point clouds to reduce the overall localization time by performing direct 2D-to-3D matching against small size of 3D point cloud. After reviewing the overall structure and workflow of the HD^{4}AR system for image-based 3D reconstruction and localization, the paper presents the 3D point cloud clustering approach in detail. The robustness of the proposed system has been validated with real static building on the campus under varying degrees of viewpoint. The potential of the HD^{4}AR, as the basis of smart device based field reporting and operation and maintenance solution for improving the productivity of onsite construction and facility management activities, is also discussed.

HD^{4}AR: HYBRID 4-DIMENSIONAL AUGMENTED REALITY

HD^{4}AR is a mobile augmented reality system that allows site personnel to use existing and already available camera-equipped mobile devices, such as smartphones or tablets, to take pictures and accurately retrieve project information related to users' surrounding context. In order to augment photographs, the HD^{4}AR first requires an initial 3D point cloud of target scene that roles as a reference model for entire localization/augmentation process. Preparing the 3D point cloud, i.e. bootstrapping process, requires initial overlapping photographs and the Structure-from-Motion (SfM) algorithms that estimates 3D position of 2D image feature points. By introducing a new parallelized SfM framework, which accelerates the computational time of an existing 3D reconstruction pipeline by a factor of 30 times, Bae et al. (2013ab;2012) make model-based localization feasible in mobile augmented reality and provide much shorter point cloud preparation time compared to existing work. To speed up the 3D reconstruction, the HD^{4}AR has used four approaches: 1) the combination of several state-of-the-art feature detectors and feature descriptors including binary descriptors, 2) new filtering procedure on the track creation and SfM stages to reduce the noise of a final 3D point cloud, 3) extracting representative 3D descriptors to optimize the memory consumption and enable direct 2D-to-3D feature matching for localization of new images with respect to the point cloud model, and 4) a scheme for use of multi-core CPU and GPU.

Along with the proposed 3D reconstruction framework, the HD^{4}AR also provides a new plane transformation (Homography) based 3D cyber-physical content authoring approach, which purely creates 3D cyber-information using user inputs from a single 2D image and automatically associates user-driven cyber-information with corresponding physical objects in 3D geometry. As described by Bae et al. (2013a,b), user-driven elements on 2D images (e.g., onsite daily construction, QA/QC, punch lists, and facility inspection reports) can be accurately triangulated and associated with components of the building or civil infrastructure in target 3D point cloud. In addition, the generated 3D cyber-information by users can be precisely overlaid on other photographs taken at completely different locations.

Once the point cloud is generated and the cyber-information is aligned, the HD^{4}AR server can augment photos sent from the client running on user's mobile devices. From a high-level perspective, this process operates as shown in Figure 2. Step 1, the field personnel, upon finding a section of the worksite he/she wishes to query, takes a picture of the area using a mobile device. Step 2, the device uploads the captured image to the HD^{4}AR server. Step 3, the server runs feature detection, feature matching, and camera calibration algorithms to identify the relative location and orientation of the camera against the 3D point cloud. Step 4, using the relative rotation and translation information of the image as input, the server determines what cyber-information are within the image's field of view, and where they appear. Step 5, the cyber-objects are sent back to the user's device with positional information and semantic information. Step 6, the user's device renders the captured image overlaid with the returned cyber-objects in the correct position. After a field engineer has an augmented photograph, he or she is able to use a multi-touch interface to select physical components in camera's field of view to retrieve more information.

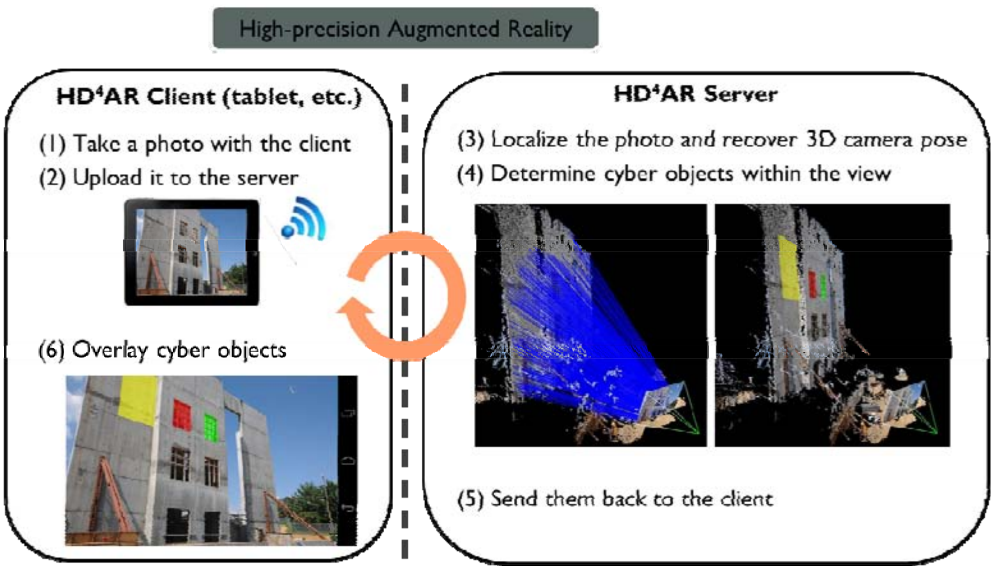

Figure 2. The localization/augmentation process of the HD^{4}AR method.

3D POINT CLOUD CLUSTERING

The localization speed of the HD^{4}AR depends on the number of 3D points in the point cloud. If users build a dense and large-scale 3D point cloud, such as entire construction site or buildings, the localization and augmentation time is getting longer since the system tries to perform direct 2D-to-3D matching with a huge 3D point cloud. For staging the solution on construction sites, in our previous approach, we proposed onsite personnel to create multiple point clouds associated with different locations/areas on the jobsite. For example, separate point cloud models were created for different spaces within the same building floor such as corridor and rooms. This strategy requires the user to choose the location/area from a list on the client's device and enable image-based localization with respect to the corresponding point cloud. To alleviate the requirement for generating separate point cloud models and selecting an appropriate model during onsite information retrieval, in this paper, we propose to segment a large-scale 3D point cloud into several clusters automatically and use each cluster to localize and augment new photograph sent from the client device. To cluster the 3D point cloud, we use GPS latitude and longitude values measured by mobile device and recorded in the image in form of EXIF (Exchangeable Image File Format) tag. There is no need for accurate GPS values as we only use this information for clustering purposes. The overall steps for 3D point cloud clustering are:

1) *Partitioning the base images*: All base images participated in 3D reconstruction are divided into several clusters using latitude and longitude values of each base image. In order to find the proper number of clusters, hierarchical clustering analysis is first used to estimate starting values for the K-means algorithm (Norušis 2011). Based on the resulting number of clusters, K-means is performed to partition base images to each cluster with the nearest mean of GPS values.
2) *Clustering the point cloud*: Once the base images are successfully partitioned, we segment the target 3D point cloud by selecting 3D points that are observed by base images in each cluster. As a consequence, each clustered point cloud contains less 3D points compared to initial 3D point cloud, resulting smaller scale.

The localization process is slightly modified to handle clustered 3D point clouds. In our new method, upon receiving the new photograph from the client device, the HD^{4}AR server first finds the nearest cluster by comparing GPS values recorded in the new photograph to mean value of each cluster. After finding the nearest cluster, the server performs existing localization method, i.e. camera calibration with direct 2D-to-3D matching, to compute a complete pose of the camera. If the new photograph does not include GPS tag, the server attempts to localize the image with all clustered point clouds in parallel. As we will discuss in next section, the proposed clustering approach results faster localization compared to our previous work although it requires mobile devices to enable GPS sensors during the AR cycle.

EXPERIMENTAL RESULTS

This section describes experiments we conducted to assess the performance of localization with the proposed clustering approach. The server side of the HD^{4}AR was running on a desktop computer with 8 gigabytes of 667 MHz DDR3 RAM, and a 4-core Intel i7 CPU 870 (@2.93 GHz) processor running Windows 7. The NVIDIA Geforce GTX 560 Ti graphic card was used for GPU computations. The base image set used to create a 3D point cloud came from an existing building on campus of Virginia Tech. The localization test images were taken at random locations and tested on-site for localization robustness. All the photographs were taken using Samsung Galaxy Nexus smartphone with Android version 4.2.

3D Reconstruction and Point Cloud Clustering. First, a 3D reconstruction procedure with the HD^{4}AR was performed on base images. To validate our approach, we enabled the GPS sensor installed in smartphones and recorded its values in form of EXIF tag during the photo collection. In addition, Fast REtinA Keypoint (FREAK) (Alahi et al. 2012) descriptor is used to minimize feature extraction time and memory consumption during the 3D reconstruction. The resulting point cloud is then partitioned into three clusters using GPS values of each base image. The final results of 3D reconstruction and clustering are summarized in Table 1 and the results show that the initial 3D point cloud is successfully reconstructed and well partitioned into three clusters. Figure 3 shows the final 3D point cloud and its corresponding clusters.

Table 1. 3D reconstruction and clustering results

	Initial Point Cloud	Cluster #1	Cluster #2	Cluster #3
Number of base images	66	15	21	30
Number of 3D points	70,906	24,178	23,098	27,528
Mean re-projection error	0.523 pixels	0.511 pixels	0.553 pixels	0.608 pixels
Centroid (latitude, longitude)	(37.2290, -80.4225)	(37.2293, -80.4227)	(37.2289, -80.4227)	(37.2290, -80.4222)

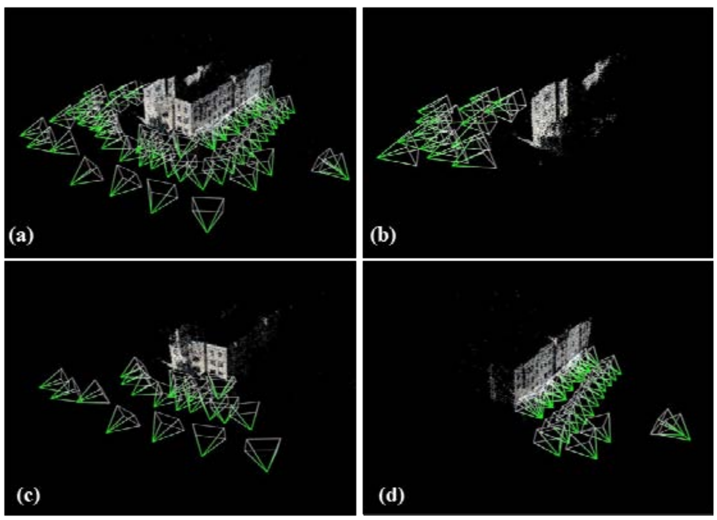

Figure 3. Resulting 3D point clouds with the HD^{4}AR and the proposed clustering method; (a) Initial 3D point cloud, (b) cluster #1, (c) cluster #2, and (d) cluster #3.

Localization with Clustered 3D Point Clouds. The localization success-ratio, mean re-projection error, and the elapsed time using clustered 3D point clouds were measured and compared to results using non-clustered single point cloud. In this experiment, we only measured the localization performance with the sequential requests from a single device although the HD^{4}AR can handle multiple requests of localization from several client devices simultaneously, which leads to increased system capacity. As observed in Table 2, the experimental results show that the clustering approach indeed accelerates the overall localization speed up to 1.535 times with the tested data set, without significantly reducing success-ratio and mean re-projection error. By using geo-location data, all tested images were matched against correct clusters, and thus resulted in 100% success-ratio of localization. In addition, the mean re-projection error of localized photographs with each cluster presents 1-pixel error in all cases.

To further demonstrate the acceleration factor of the proposed approach, we also measured elapsed times for each step in localization, i.e. the file i/o time (loading corresponding 3D point cloud onto memory), cluster selection time, feature extraction time, and the matching/calibration time. As shown in Table 3, the matching and calibration takes the longer time when the scale of 3D point cloud (i.e. number of 3D points) is larger, while the feature extraction time remains constant. Therefore, we can conclude that the proposed clustering approach, which segments large-scale point cloud into smaller point clouds, reduces overall localization speed by reducing the scale of 3D point cloud and direct 2D-to-3D matching time. If we only consider the direct 2D-to-3D matching procedure, the matching/calibration time is up to 1.924 times faster than our previous work.

Table 2. Localization success-ratio and average localization time

	Initial Point Cloud	Cluster #1	Cluster #2	Cluster #3
Localization success-ratio	100% (75/75)	100% (25/25)	100% (25/25)	100% (25/25)
Mean re-projection error	0.958 pixels	0.937 pixels	0.960 pixels	1.037 pixels
Avg. localization time	2.735 sec	1.897 sec	1.782 sec	1.934 sec
Performance gain in localization time	1×	1.442×	1.535×	1.414×

Table 3. Details of localization time

	Initial Point Cloud	Cluster #1	Cluster #2	Cluster #3
Number of 3D points	70,906	24,178	23,098	27,528
File i/o time[a]	1.124 sec	0.327 sec	0.316 sec	0.391 sec
Cluster selection time	0 sec	3.5×10^{-7} sec	3.5×10^{-7} sec	3.5×10^{-7} sec
Feature extraction time	0.759 sec	0.775 sec	0.755 sec	0.760 sec
Matching/calibration time	1.976 sec	1.122 sec	1.027 sec	1.174 sec
Performance gain in matching/calibration time	1×	1.761×	1.924×	1.683×

[a] Due to the server-client architecture, loading 3D point cloud onto memory takes place only once during the AR cycles. Thus, it is excluded from calculating the overall localization time.

CONCLUSION

The HD^4AR was designed with the intent of bringing high-precision mobile augmented reality to field personnel without requiring external sensors or infrastructures. The HD^4AR allows using existing mobile devices to take pictures for accurate localization and visualize project information, punch list, and/or inspection reports on top of the associated building elements in the photographs. The performance of the HD^4AR, with a localization success-ratio of 100% and mean re-projection error less than 1 image pixel, implies that the system can be applied to construction progress monitoring, QA/QC reporting, and/or facility management operations. Despite the accuracy and near real-time performance of the HD^4AR, however, the localization speed needed to be further accelerated to provide better user experience. To address this issue, in this paper, we proposed a clustering method that partitioned 3D point cloud into smaller scale of point clouds using geo-location data.

The clustered approach reduces the scale of 3D point cloud to be searched and accelerates overall localization time compared to our previous work. With the smaller, clustered point clouds, the HD^{4}AR now takes less than 2 seconds for single localization. In future work, we plan to develop a cache based kd-tree approach, which caches 3D points by analyzing users' localization request pattern and uses a small cached point cloud to further accelerate localization speed.

ACKNOWLEDGEMENTS

This material is based upon work supported by the National Science Foundation under CMMI-1360562 award. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

Alahi, A., Ortiz, R., Vandergheynst, P. (2012). "FREAK: Fast Retina Keypoint." *Proc. 2012 IEEE* CVPR, Providence, Rhode Island, 510-517.

Aziz, Z., Anumba, C. J., Ruikar, D., Carrillo, P. M., and Bouchlaghem, D. N. (2005). "Context aware information delivery for on-site construction operations." *Proc., the 22nd CIB-W78 Conference*, Dresden, Germany, 304, 321-332.

Bae, H., Golparvar-Fard, M., and White, J. (2013a). "High-precision vision-based mobile augmented reality system for context-aware architectural, engineering, construction and facility management (AEC/FM) applications." *Visualization in Engineering*, 1(3). doi:10.1186/2213-7459-1-3.

Bae, H., Golparvar-Fard, M., and White, J. (2013b). "High-precision and Infrastructure-independent Mobile Augmented Reality System for Context-Aware Construction and Facility Management Applications." *Proc., 2013 ASCE Int. Workshop on Computing in Civil Eng.*, Los Angeles, California, 637–644.

Bae, H., Golparvar-Fard, M., and White, J. (2012). "Enhanced HD4AR (Hybrid4-Dimensional Augmented Reality) for Ubiquitous Context-aware AEC/FM Applications." *Proc., CONVR 2012*, Taipei, Taiwan, 253–262.

Bowden, S., Dorr, A., Thorpe, T., and Anumba, C. (2006). "Mobile ICT support for construction process improvement." *Automation in Constr*, 15(5), 664-676.

Chen, Y. and Kamara, J. (2011). "A framework for using mobile computing for infomation management on construction sites." *AutoCon*, 15, 605–613.

Engineering News Record (ENR) (2012). "New Survey Shows That Mobile Tools Boost Jobsite Productivity," http://bit.ly/INCqlf, Last accessed 10/31/13.

Golparvar-Fard, M., Peña-Mora, F., and Savarese, S. (2012). "Automated model-based progress monitoring using unordered daily construction photographs and IFC as-planned models." *J. of Computing in Civil Eng.* doi: 10.1061/(ASCE)CP.1943-5487.0000205.

Golparvar-Fard, M., Peña Mora, F., and Savarese, S. (2011). "Integrated sequential as-built and as-planned representation with D^{4}AR tools in support of decisionmaking tasks in the AEC/FM industry." *ASCE J.of Construction Eng and Management*, 137(12), 1099–1116.

Kim, C., Park, T., Lim, H., and Kim, H. (2013). "On-site construction management using mobile computing technology." *Automation in Constr*, 35, 415-423.

Navon, R. and Sacks, R. (2007). "Assessing research issues in Automated Project Performance Control (APPC)." *AutoCon*, 16(4), 474-484.

Norušis, M. J. (2011). *IBM SPSS Statistics 19 Stat Procedures Companion*, Addison Wesley.

Retik, A., Clark, N., Fryer, R., Hardiman, R., McGregor, D., Mair, G., and Revie, K. (2002). "Mobile hybrid virtual reality and telepresence for planning and monitoring of engineering projects." *AutoCon*.

Son, H., Park, Y., Kim, C., and Chou, J. S. (2012). "Toward an understanding of construction professionals' acceptance of mobile computing devices in South Korea: An extension of the technology acceptance model." *AutoCon*, 28, 82-90.