

RESEARCH ARTICLE

Open Access

High-precision vision-based mobile augmented reality system for context-aware architectural, engineering, construction and facility management (AEC/FM) applications

Hyojoon Bae^{1*}, Mani Golparvar-Fard² and Jules White³

Abstract

Background: Many context-aware techniques have been proposed to deliver cyber-information, such as project specifications or drawings, to on-site users by intelligently interpreting their environment. However, these techniques primarily rely on RF-based location tracking technologies (e.g., GPS or WLAN), which typically do not provide sufficient precision in congested construction sites or require additional hardware and custom mobile devices.

Method: This paper presents a new vision-based mobile augmented reality system that allows field personnel to query and access 3D cyber-information on-site by using photographs taken from standard mobile devices. The system does not require any location tracking modules, external hardware attachments, and/or optical fiducial markers for localizing a user's position. Rather, the user's location and orientation are purely derived by comparing images from the user's mobile device to a 3D point cloud model generated from a set of pre-collected site photographs.

Results: The experimental results show that 1) the underlying 3D reconstruction module of the system generates complete 3D point cloud models of target scene, and is up to 35 times faster than other state-of-the-art Structure-from-Motion (SfM) algorithms, 2) the localization time takes at most few seconds in actual construction site.

Conclusion: The localization speed and empirical accuracy of the system provides the ability to use the system on real-world construction sites. Using an actual construction case study, the perceived benefits and limitations of the proposed method for on-site context-aware applications are discussed in detail.

Background

Automated, on-demand, and inexpensive access to project information on-site has significant potential to improve decision-making during construction or facility management activities. This information, which is usually in the form of specifications, drawings, or schedule information, enables prompt identification, processing, and communication of discrepancies between actual and expected performance. Fast access to this information also helps project managers to proactively decide on corrective actions and minimize the cost and delays due to performance discrepancies (Golparvar-Fard et al. 2012).

Despite the importance of on-site information access, most of current approaches to jobsite progress monitoring include manual and time consuming data collection, non-systematic analysis and visually/spatially complex reporting (Golparvar-Fard et al. 2012; Navon and Sacks 2007). As a part of data collection and analysis, field personnel have to carry large stacks of specifications and drawings on jobsites and spend significant amount of time to write down an actual progress on paper and compare it to relevant cyber-information (Khoury and Kamat 2009). Such inefficiencies in site analysis and information gathering can cause downtime or rework and ultimately lead to schedule delays or cost overruns. In addition, the quality and timing of information access and exchange can either delay or facilitate successful execution of on-site activities (Chen and Kamara 2011).

*Correspondence: hjbae@vt.edu

¹Bradley Department of Electrical and Computer Engineering, Virginia Polytechnic Institute and State University, Blacksburg, VA, USA
Full list of author information is available at the end of the article

To minimize these inefficiencies, we have proposed a new context-aware vision-based mobile augmented reality system, Hybrid 4-Dimensional Augmented Reality (HD⁴AR), which identifies location and orientation of field personnel solely based on a site photograph (Bae et al. 2012). As described by Bae et al. (2012), HD⁴AR allows field personnel to query and access semantically-rich 3D cyber-information and see it precisely overlaid on top of real-world imagery. HD⁴AR does not need RF-based tracking technologies or inertial measurements to find a user's position. Rather, the system takes a photograph from jobsite as input and computes the location and orientation of the user's camera using a set of computer vision algorithms. As a result, HD⁴AR can be used with any camera-equipped mobile device, such as a smartphone or tablet, to provide accurate on-site localization of a field engineer and thus is practical and inexpensive to use on a construction site. As shown in Figure 1, the proposed vision-based system can successfully localize a photograph even with large changes in the viewpoint of a user. Moreover, as shown in the bottom-right screenshot, the system successfully recognizes the target building from the photograph, which includes the cellular phone displaying the building, and overlays the cyber-information precisely. This photograph has different color histogram and pixel values from the photograph taken at the actual site, which can simulate the different illumination conditions of the target scene.

This paper extends our prior work on HD⁴AR (Bae et al. 2012) in the following ways: 1) the localization speed is further increased using direct 2D-to-3D

matching, 2) different image feature description methods are implemented and tested to investigate the impact of those descriptors on performance of 3D reconstruction and localization, 3) a new homography-based 3D content-creation (annotation) method, which allows a field engineer to draw and mark any building elements within the photograph, is described. The enhanced localization speed and impact of feature descriptors will be further discussed in Section 'Results and discussion', while 3D annotation functionality will be examined in Section 'High-precision augmentation with HD⁴AR'.

The remainder of this paper is organized as follows: After brief summary of related works in Section 'Related work', Section 'Method: Hybrid 4-dimensional augmented reality' presents an overview and technical approach of the HD⁴AR system. The details of the 3D reconstruction process that generates a 3D point cloud model from a set of unordered photographs are discussed in Section '3D reconstruction with HD⁴AR'. Section 'High-precision augmentation with HD⁴AR' presents the localization and augmentation process using a generated 3D point cloud. The new feature of the HD⁴AR, e.g., 3D annotation, is also discussed in this section. Section 'Results and discussion' presents empirical results from experiments with HD⁴AR and also compares the performance to other state-of-the-art Structure-from-Motion based (SfM-based) localization solutions. Finally the perceived benefits and limitations are described in Section 'Conclusion'. Video demos and detailed performance data of HD⁴AR can be found at <http://www.magnum.ece.vt.edu/index.php/research-projects/100-hd4ar> and <http://raamac.cee.illinois.edu/hd4ar>.

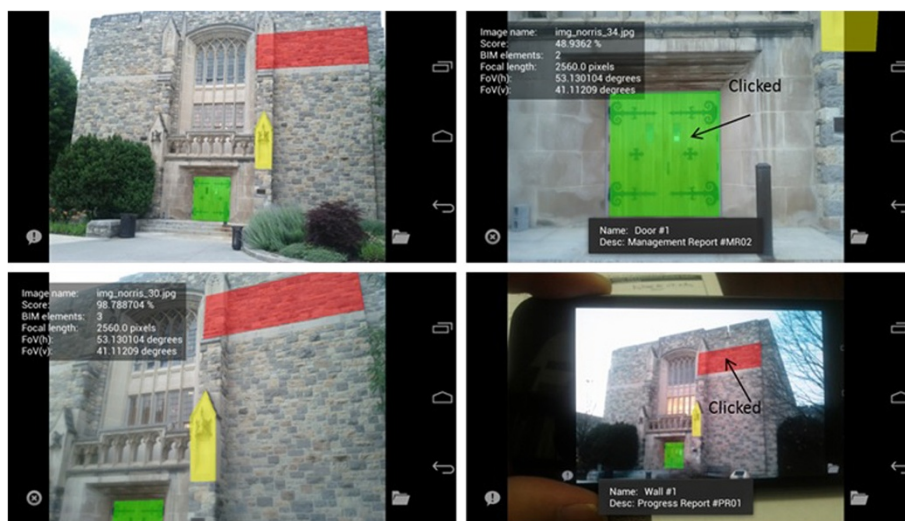


Figure 1 Screenshots of the Android HD⁴AR client. 3D BIM information is precisely overlaid on photos From different viewpoints. (adopted from Bae et al. 2012).

Related work

The advantages of using augmented reality system on Architectural, Engineering, Construction (AEC) application has been extensively studied by many researchers (e.g., Behzadan and Kamat 2007; Behzadan et al. 2008; Chen et al. 2011; Chi et al. 2013; Dunston et al. 2003; Golparvar-Fard 2009a, 2009b; Hammad et al. 2009; Hou and Wang 2010; Kuula et al. 2012; Schall et al. 2009; Wang 2008; Wang and Dunston 2006; Woodward et al. 2010; Yeh et al. 2012). They have shown that augmented reality indeed improve physical task performance and can reduce mental workload of engineers for AEC tasks (Wang and Dunston 2006). They have also indicated that augmented reality improves design activities as well as design visualization by providing better spatial cognition (Dunston et al. 2003). On-site building information retrieval using a wearable device, proposed by Yeh et al. (2012), also validates that the proper displaying of user-required information on-site leads to shorter task completion time and higher correctness than traditional approach.

To exploit these benefits, many research projects have focused on providing cyber-information to field personnel through mobile devices and/or augmented reality systems (e.g., Akula et al. 2011; Anumba and Aziz 2006; Behzadan et al. 2008; Shin and Dunston 2008; Hakkarainen et al. 2008; Khoury and Kamat 2009; Irizarry et al. 2012; Pisman and Woodward 2003). These works have primarily focused on using Global Positioning Systems (GPS), Wireless Local Area Networks (WLAN), or Indoor GPS for accurately positioning the user within congested construction environments. The main drawback of these Radio Frequency based (RF-based) location tracking technologies is their high degree of dependency on pre-installed infrastructure, which makes their application either difficult or impractical for construction sites. Application of fiducial markers is also suggested by several researchers (e.g., Feng and Kamat 2012; Hakkarainen et al. 2008; Lee and Akin 2011; Yakubi et al. 2011). The systems are also infrastructure-dependent and require the markers to be attached to various surfaces on construction sites, which challenges their applications for large-scale implementations.

On the other hand, some researches have focused on developing infrastructure-independent location tracking systems (e.g., Akula et al. 2011; Ojeda and Borenstein 2007). These systems are typically based on inertial measurements and make use of highly accurate accelerometers and gyroscopes. Given their independence from an existing infrastructure, however, their application may result in accumulated drift error which grows with the distance traveled by the users.

Recent advances in image processing and computer vision have led to new research on the application of

image-based reasoning for various construction management tasks and techniques that can manually, semi-automatically, and automatically interpret them (Cheng and Chen 2002; Carozza et al. 2012; Golparvar-Fard et al. 2010,2010; Kiziltas et al. 2008). These researches have shown that a set of overlapping images can be used to extract accurate 3D geometry of stationary objects such as buildings under construction. After the physical models (e.g., the generated 3D point cloud models) and the cyber models (e.g., Building Information Modeling (BIM)) are aligned, they can be compared to determine the actual state of the physical elements on construction site versus the expected state. Some researchers have proved that the fused (aligned) cyber-physical model is accurate to within millimeters (Golparvar-Fard et al. 2011) and can be used to predict the actual construction progress versus the planned cyber model with high accuracy, even when visual obstructions are present (2010). Recent works such as Carozza et al. (2012) extend marker-less augmented reality systems for urban planning purposes. Despite the fact that these systems rely on tracking the camera position and orientation, and does not require additional infrastructure, such systems still require a large amount of matching to be conducted at each step. As reported by Carozza et al. (2012), the overall tracking is likely not as efficient as tracking image features.

Although this body of computer vision research has shown the potential and high-accuracy of image-based reasoning, the speed of 3D reconstruction/localization and the lack of on-site localization methods make these systems difficult to use on worksites. Generating a 3D point cloud model from a set of construction photographs requires non-linear multi-dimensional optimization as well as exhaustive matching of the photographs in the data set and can take hours or days. A specific aim of HD⁴AR was to overcome these challenges, speeding up overall time of 3D reconstruction and localization by optimizing and enhancing each process.

Method: Hybrid 4-dimensional augmented reality

Overview

HD⁴AR combines user localization and AR visualization to target on-site query and view of project information on top of real-world imagery. For user localization, HD⁴AR uses a computer vision-based and model-based method, which obtains detailed information from pre-reconstructed 3D point cloud models built from daily construction photos, and estimates the location of a field engineer's camera using these models. Using 3D point cloud models additionally permits the system to estimate the complete pose of the camera and therefore can support high-accuracy applications such as construction progress monitoring where millimeter-level precision is needed. Because HD⁴AR relies on 3D point cloud models,

it requires that users first take overlapping photos of the target scene to produce the initial 3D point cloud used for localization. This initial 3D reconstruction is based on a Structure-from-Motion (SfM) algorithm that triangulates the 3D position of image features in photographs through feature extraction, matching, and an optimization process called Bundle Adjustment.

Once the 3D reconstruction is done, a field engineer can take a new photo at a random location and his/her location and orientation are determined by comparing the new image to the generated 3D point cloud. Specifically, the system attempts to estimate extrinsic camera parameters, e.g. rotation matrix and translation vector of the camera, to find the relative position of the camera. After recovering the complete pose of the user's camera, HD⁴AR decides what cyber-information, such as elements of the BIM, should appear in the field engineer's photograph.

Finally, HD⁴AR allows a field engineer to select physical objects in the photograph by touching on them in order to retrieve more information associated with each object. Moreover, a field engineer can create new BIM elements by simply drawing a polygon on the photograph. The user-created 2D BIM elements are then automatically back-projected to cyber 3D space and attached to the existing cyber-physical model. Once user-created elements are successfully back-projected, they can be accurately overlaid on other photographs, which are taken from significantly different viewpoints. This simple 3D annotation functionality is one of the distinct features of HD⁴AR. Figure 2 summarizes the overall procedures of HD⁴AR, from initial 3D reconstruction process to localization/augmentation process.

Technical approach

As aforementioned, HD⁴AR is based on a set of computer vision algorithms. However, due to exhaustive

computations including non-linear multi-dimensional optimization processes in the SfM algorithm, model-based approaches are often considered impractical solutions for user localization. For example, the Bundler package (Snavely et al. 2007), a widely-used software package that implements SfM for 3D reconstruction, takes from hours to days to generate a 3D point cloud even for small number of base images. In addition, it uses the SIFT (Scale Invariant Feature Transformation) descriptor (Lowe 2004) for feature extraction, which have good invariance properties but require multiple layers of computation for each spatial scale, and thus is time consuming. Therefore, we designed and implemented a new parallelized 3D reconstruction module that operates across cores in a multi-core CPU and GPU. HD⁴AR uses a client-server architecture with the mobile phone as the client that uploads photos to the server for localization and the major image processing load located on the server-side. The entire system consists of following components:

- A 3D reconstruction component runs on the server on a multi-core CPU and GPU. This component generates a 3D point cloud from the initial base images through feature extraction, matching, and the SfM procedure.
- A user localization component runs on the server. This component takes a single photograph taken from a mobile device as input and derives the 3D position and orientation of the mobile device with respect to the 3D point cloud by solving a Direct Linear Transform equation followed by a Levenberg-Marquardt optimization against the underlying point cloud model.
- A client component, which is a small program that runs on Android and iOS smartphones, sends user-captured images to the server. It also has the

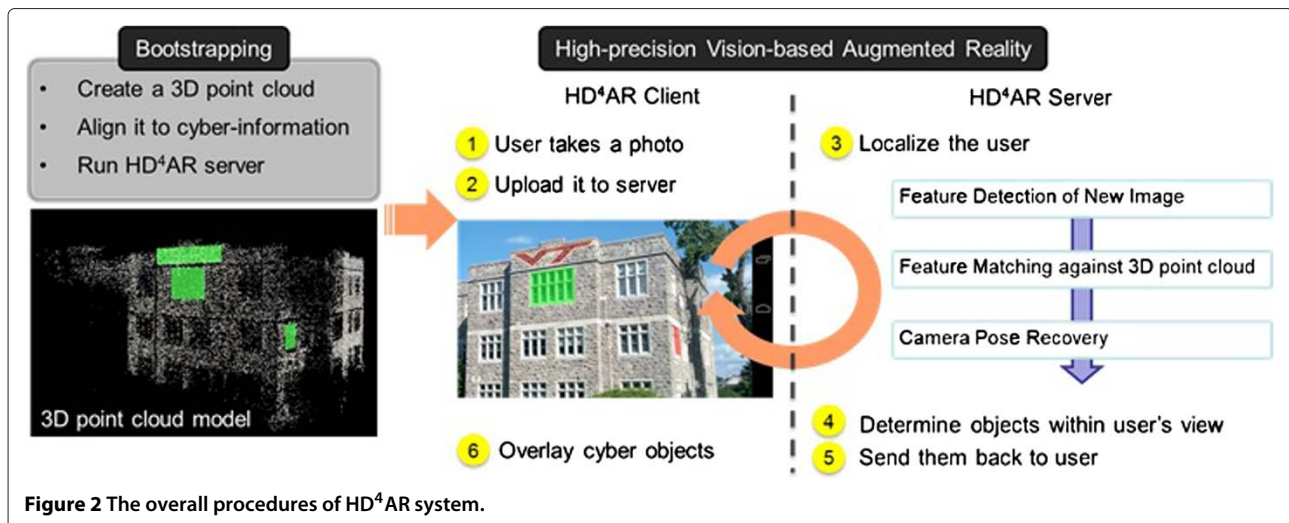


Figure 2 The overall procedures of HD⁴AR system.

capability of drawing cyber objects on top of the photograph once it gets localization results from the server.

3D reconstruction with HD⁴AR

Most engineering workstations today have a multi-core CPU with 2–16 cores and a GPU with anywhere from 4 to 128 cores. Exploiting this hardware parallelism is key to the performance and scalability of HD⁴AR. We parallelize all the steps for 3D point cloud generation to obtain performance gains and implement GPU-based SURF (Speeded-Up Robust Features) descriptors (Bay et al. 2008) and CPU-based FREAK (Fast REtinA Keypoint) descriptors (Alahi et al. 2012) for fast feature extraction. The 3D reconstruction procedure in HD⁴AR mostly follows the original steps in the SfM algorithm of the Bundler package except that it 1) uses different feature detectors and descriptors, 2) introduces new optimization parameters for reducing noise in the 3D point cloud to improve localization accuracy, and 3) exploits multi-core CPU and GPU hardware for faster processing speeds. Figures 2 and 3 show the steps of image-based 3D reconstruction from a high-level perspective. Each step can be summarized as follows:

Feature Detection and Extraction

To find a set of image feature points, a feature detection and extraction algorithm is executed on each base image. Two different state-of-the-art feature descriptors, e.g., SURF and FREAK, are implemented and tested in the HD⁴AR 3D reconstruction pipeline. In contrast to SIFT, SURF creates a stack of integral images without down-sampling for higher levels in the pyramid and it filters the stack using a box filter approximation of second-order Gaussian partial derivatives to speed up the processing

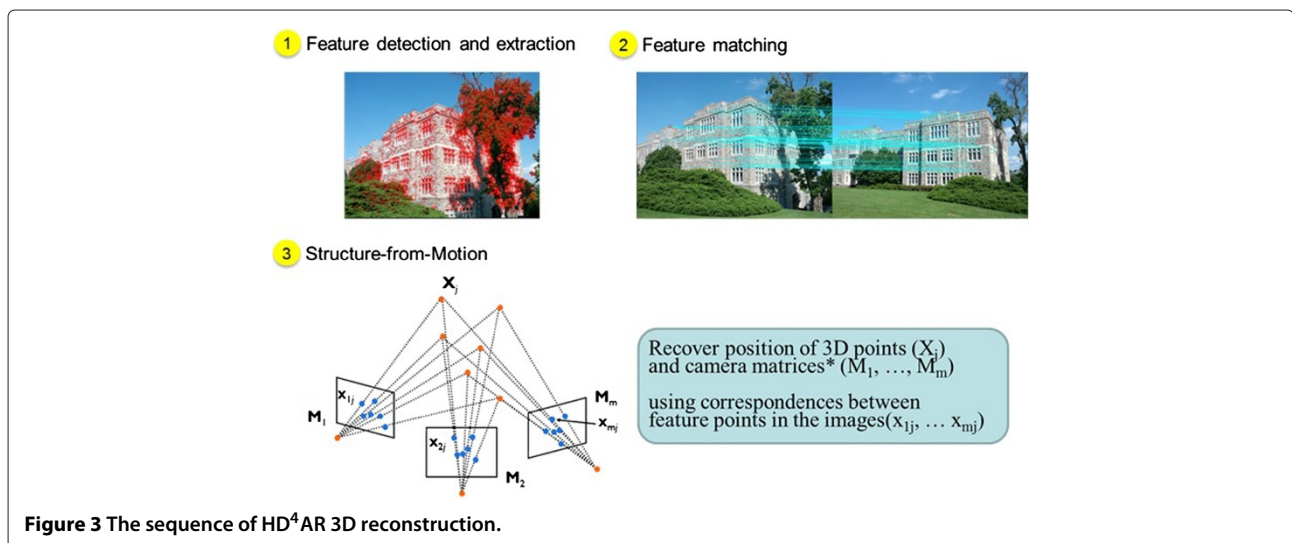
(Bay et al. 2012). On the other hand, FREAK uses retinal sampling patterns to compare image intensities and produces a cascade of binary strings (Alahi et al. 2012). Both SURF and FREAK are invariant to image scale and rotation, but provide faster feature extraction than SIFT. Therefore, the HD⁴AR pipeline now supports SURF and FREAK, in addition to SIFT, to speed up feature extraction.

Feature Matching

The next step is finding correspondences between each image pair (e.g., pair-wise matching). For each image pair, HD⁴AR creates a *kd*-tree of the descriptors and runs the Approximate Nearest Neighbors (ANN) algorithm (Arya et al. 1998) to find the two nearest neighbors of each descriptor. Then the HD⁴AR performs a distance ratio-test (Lowe 2004) to remove erroneous matches. In addition, if more than one feature descriptor matches the same feature in the opposite image, the HD⁴AR removes all of those matches. Finally, the HD⁴AR robustly estimates a Fundamental matrix with the eight-point algorithm (Hartley and Zisserman 2004) loop and removes matching outliers for every image pair. This filtering process removes false matches using an epipolar geometry constraint given by the estimated Fundamental matrix. To shorten the overall matching time, each image pair is processed on different CPUs with parallelized I/O tasks.

Structure-from-Motion (SfM)

- 1) *Camera Registration and Point Triangulation*: The SfM algorithm estimates a set of camera parameters, such as the focal length, rotation matrix, and translation vector, for each image and triangulates 3D positions of feature points observed in each image. Similar to the Bundler package, the HD⁴AR



uses an incremental approach, e.g., recovering a few cameras at a time. The HD⁴AR starts with initial image pair to recover camera parameters using Nistér's five-point algorithm (Nistér 2004), and triangulates their feature points. As discussed by Snavely et al. (2007), the initial pair should have a large number of matched feature points, but also have a long separation distance between the cameras to avoid converging in a local minimum during the optimization process. After estimating the camera parameters of the initial image pair, the HD⁴AR attempts to calibrate the camera parameters of each additional base image using the already triangulated 3D points and matching information between the images. If the system successfully recovers camera parameters of an additional base image, it registers the new camera and triangulates the points seen by the newly registered camera. This registration fails in the event that an additional base image does not have any matched feature points against the previously registered images. In the HD⁴AR, these camera registration and point triangulation steps are well-parallelized to exploit multi-core CPUs.

- 2) *Incremental Bundle Adjustment*: While the base images are being added (registered), the 3D reconstruction pipeline is run through a GPU-based sparse Bundle Adjustment module to minimize the overall re-projection error, e.g., the difference between predicted 2D positions of the feature points in the photographs given their triangulated 3D positions and the locations of where the feature points are actually extracted in the images. The HD⁴AR adopts Parallel Bundle Adjustment (Wu et al. 2011) to significantly enhance the speed of this optimization.
- 3) *Noise Reduction*: Bundle Adjustment is an optimization process that tries to minimize the overall re-projection error of all 3D points at the same time. It is possible that some 3D points have high re-projection error while other 3D points have a very small re-projection error, resulting in an overall small Minimum Mean-Square Error (MMSE). Since the ultimate purpose of the 3D point cloud generation is user localization, not the visual representation of target scene in 3D, it is very important to reduce the noise in the 3D point cloud by removing 3D points with high re-projection errors. To achieve this, HD⁴AR uses a double-threshold scheme. The first threshold is for controlling the target MMSE of Bundle Adjustment. We set this threshold as 1.0 pixel² so that the average re-projection error of entire 3D point cloud is not greater than 1.0 pixel. Another threshold, which we call an absolute re-projection threshold, is for removing individual

3D points from point cloud. This threshold is set to be 4.0 pixels so that no 3D points in final point cloud have a re-projection error greater than 4.0 pixels.

Due to our algorithmic enhancements and parallelization, 3D reconstruction with HD⁴AR is up to 35 times faster than the Bundler package. In Section 'Results and discussion', the experimental results of 3D reconstruction are discussed in detail. Figure 4 shows some examples of 3D point clouds generated by HD⁴AR using real-world construction site photos and existing building photos in Blacksburg, VA.

High-precision augmentation with HD⁴AR Localization and augmentation

Once the HD⁴AR has the 3D point cloud of the target construction site or building, the system can accurately localize and augment new photographs captured on a mobile device. Figures 2 and 5 summarize this process from a high-level perspective. In this use case, a field engineer first takes a picture of the building elements, which he/she wishes to query the information about, and uploads the photograph to the HD⁴AR server. Upon receiving the photo from user's device, the server starts to run feature detection on the received image, finding correspondences between the image and 3D point cloud, and camera calibration to identify the relative pose of the camera. If the server successfully estimates the camera pose information, it determines what cyber-information is within the camera's field of view and where the information should appear. This decision is done by first projecting each vertex of cyber-information:

$$s \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} [R | T] \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (1)$$

where $[X, Y, Z]$ is a 3D vertex point of cyber-information, $[R|T]$ is the estimated rotation matrix and translation vector, $[f_x, f_y]$ is the focal lengths expressed in pixel units, $[c_x, c_y]$ is a principal point, and $[x, y]$ is the resulting projected point in pixels. Then the simple visibility test is performed to determine whether the cyber-information appears in current image or not:

$$V(x, y) = \begin{cases} 1, & 0 \leq x \leq W, 0 \leq y \leq H \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

where W is image width and H is image height. The visible cyber-information is then sent back to user's device with

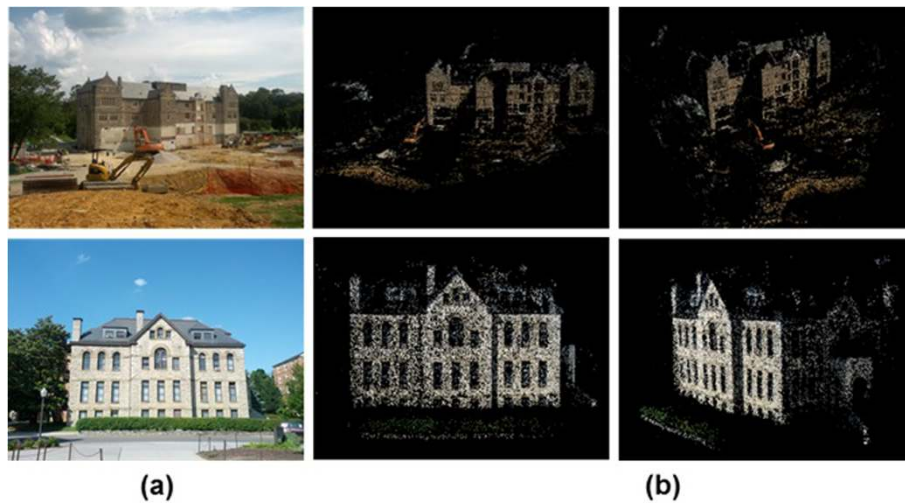


Figure 4 Example of HD⁴AR 3D reconstruction. (a) Initial base images. (b) 3D point-cloud from HD⁴AR. Resulting 3D point clouds well-represent the target construction site and building.

positional information and semantics. Finally, the user's device renders the returned visible cyber-information on the top of captured-image. As shown in Figure 6, HD⁴AR can precisely localize and augment photographs with various test cases and it implies that HD⁴AR remains stable under different illumination conditions and totally different viewpoint of the user's device.

One of the improvements over our previous work is that finding correspondences between image and the 3D point cloud is further accelerated using direct 2D-to-3D matching. HD⁴AR only compares feature descriptors of the image to that of each 3D point in the point cloud to find 2D-to-3D correspondences. With our previous approach, however, HD⁴AR matched feature descriptors of the image to an entire set of feature descriptors from all base images, which incurs unnecessary descriptor comparisons. Consequently, the localization time did depend on the number of base images as we discussed in our previous work (Bae et al. 2012). As we will see in Section 'Results and discussion', the new direct 2D-to-3D matching approach further speeds up the localization by an average factor of 2.79.

3D annotation

Upon successful localization of a new photograph, a field engineer can easily create and add a new 3D BIM element with HD⁴AR by drawing a polygon on the localized photograph. Since HD⁴AR keeps all the base images that are registered during the 3D reconstruction, it can estimate homography matrices between the localized photograph and each base image using a RANSAC algorithm. HD⁴AR then utilizes these estimated homographies to find correspondences of a user-created element for each base image. As shown in Figure 7a and 7b, a window drawn by the user is correctly found in base images by the system. Finally, HD⁴AR triangulates each vertex of the user-created polygon using camera parameters of localized photograph and registered base images. After running Bundle Adjustment to minimize the reprojection error of a triangulated polygon, the resulting 3D element is well-aligned with the existing 3D point cloud as shown in Figure 7c. Once the user-created element has 3D positional information, it can be precisely overlaid on other photographs from different viewpoints as shown in Figure 7d. This simple and robust 3D annotation/tagging functionality makes

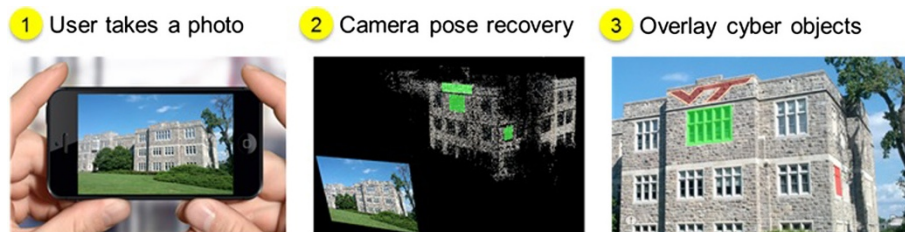


Figure 5 The sequence of HD⁴AR localization and augmentation.



Figure 6 Example of HD⁴AR localization and augmentation. Cyber-information is precisely overlaid on user's photograph despite the significant change of viewpoints.

it easier to create 3D content associated with building elements on-site, and is one of distinct features of HD⁴AR.

Results and discussion

This section presents experimental results of 3D reconstruction and user localization with HD⁴AR. To assess the ability of HD⁴AR to produce the initial 3D point clouds, 3D reconstruction is performed on several data sets, which were randomly collected from actual construction sites and existing buildings. For user localization, test images were taken at random locations and localized on-site for validating the correctness. The details of data set specification and experimental results will be discussed in following subsections.

Platform specification and data sets

The server side of HD⁴AR 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 Ubuntu version 12.04. The NVIDIA GeForce GTX 560 Ti graphic card was used for GPU computations. The image data sets used to create 3D point clouds came from the actual construction sites and existing buildings on Virginia Tech campus. Table 1 shows the summary of data sets that we used for 3D reconstruction.

Several Android smartphones were used to run the HD⁴AR client for localization tests. For fast data transfer, the client-side communication was based on Wi-Fi 802.11g connection rather than using the cellular network.

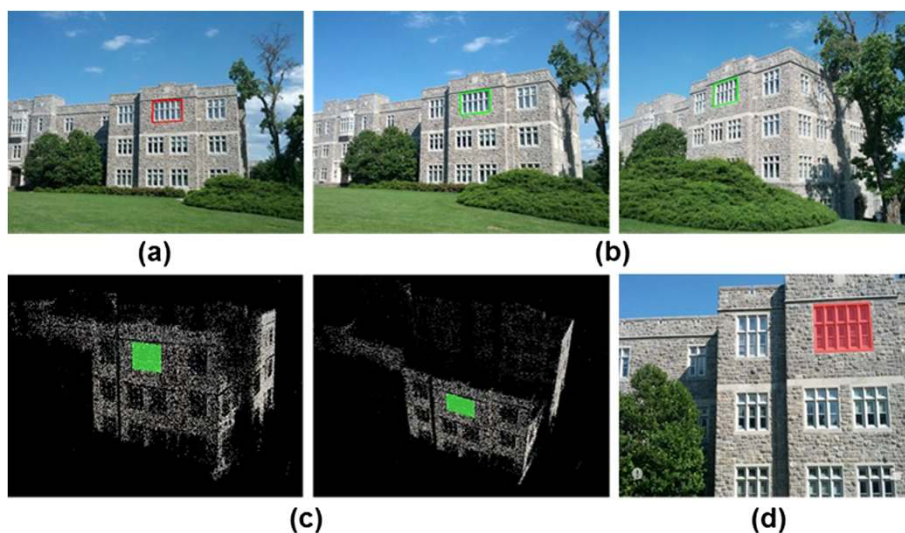


Figure 7 Example of HD⁴AR 3D annotation. (a) User marks a window on the localized photograph, (b) HD⁴AR automatically finds correspondences of window for each base image, (c) The system triangulates the window using camera information of base images and the localized photograph, (d) A user-created window element is precisely overlaid on other photographs from different viewpoints.

Table 1 Data sets for 3D reconstruction

Data set	Camera	Number of base images	Image size
Parking garage	Google Nexus S	143	2560 × 1920
Center for the arts	Nikon D300S	125	2144 × 1424
Norris hall	Google Nexus S	50	2560 × 1920
Patton hall	Samsung Galaxy Nexus	44	2592 × 1944

Performance of 3D reconstruction

An entire 3D reconstruction procedure with HD⁴AR was run on each data set to produce the initial 3D point clouds. The performance of the Bundler package was also measured and compared to that of HD⁴AR to demonstrate the performance gains of HD⁴AR's optimizations. In addition, we tested two different descriptors, i.e. GPU-based SURF and CPU-based FREAK, to investigate the impact of feature descriptors on the performance of 3D reconstruction. Table 2 compares the overall elapsed time and number of recovered cameras for 3D reconstruction on each data set. The results show that HD⁴AR with FREAK descriptor obtains the maximum performance gain of 3,471%. However, this only illustrates the tendency of time cost since there are many factors that influenced on the performance, such as number of base images, image sizes, and the texture of target scenes. Nevertheless, HD⁴AR with FREAK outperformed HD⁴AR with SURF and the Bundler package in all tested data sets. This result is due to the fact that the FREAK descriptor is a binary descriptor which uses simple Hamming distance calculations for descriptor matching, while SURF and SIFT descriptors are a vector of real numbers and must be compared using Euclidean norms. Compared to HD⁴AR with SURE, however, HD⁴AR with FREAK has fewer registered images in

the Parking Garage and Patton Hall data sets. This outcome implies that the FREAK descriptor may not be as robust as SURF or SIFT descriptors for 3D reconstruction. Having a smaller number of registered images means that there is less 3D camera information in a point cloud and therefore it may affect the localization success-ratio. The success-ratio will be discussed further in the next subsection. Figure 8 shows results of HD⁴AR 3D point cloud reconstruction for all test cases.

Performance of localization

In order to measure the reliability of the reconstructed 3D point clouds for localization of new photographs, localization tests were performed on each 3D point cloud. In this paper, the success in localization means that HD⁴AR was able to solve the camera calibration equation, e.g. Direct Linear Transform equation followed by a Levenberg-Marquardt optimization, using a given image and 3D point cloud. More extensive quantification and measurement of the accuracy of localization with different feature descriptors will be investigated in future work to determine the feasibility of precise measurement using HD⁴AR. As observed in Figure 6 and 9, however, the augmented photographs show that recovered camera parameters were accurate enough to precisely

Table 2 Performance of 3D point cloud reconstruction

Data set	System	Number of registered images/ Number of base images	Elapsed time
Parking garage	HD ⁴ AR with FREAK	125 / 143	36.73 mins (×17.25)
	HD ⁴ AR with SURF	138 / 143	42.01 mins (×15.08)
	Bundler package*	143 / 143	633.63 mins (×1)
Center for the arts	HD ⁴ AR with FREAK	125 / 125	9.25 mins (×19.67)
	HD ⁴ AR with SURF	125 / 125	12.49 mins (×14.57)
	Bundler package*	125 / 125	181.95 mins (×1)
Norris hall	HD ⁴ AR with FREAK	50 / 50	2.62 mins (×27.53)
	HD ⁴ AR with SURF	50 / 50	3.27 mins (×22.06)
	Bundler package*	50 / 50	72.13 mins (×1)
Patton hall	HD ⁴ AR with FREAK	37 / 44	3.78 mins (×34.71)
	HD ⁴ AR with SURF	43 / 44	4.80 mins (×27.34)
	Bundler package*	44 / 44	131.22 mins (×1)

*A widely-used software package for SfM (Snavely et al. 2007).

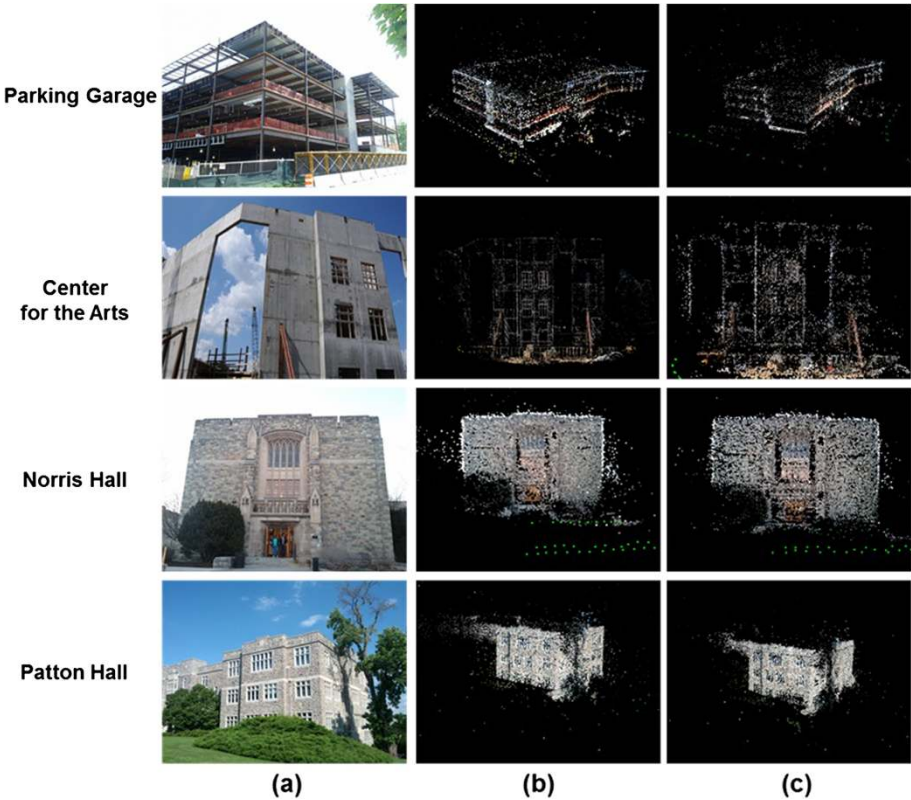


Figure 8 Results of HD⁴AR 3D reconstruction. (a) Initial base images. (b) HD⁴AR with SURF. (c) HD⁴AR with FREAK.

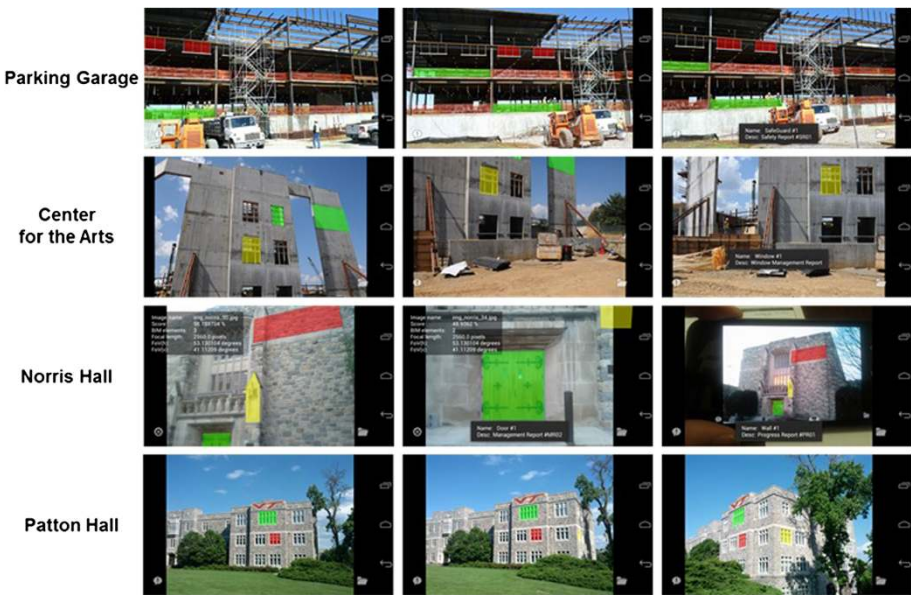


Figure 9 Results of HD⁴AR localization. From construction sites to existing building, HD⁴AR provides high-precision of cyber-information visualization.

overlay the cyber-information on photographs from different viewpoints.

Table 3 shows the localization success-ratio and average localization time for each data set. As we expected, HD⁴AR with FREAK has the lowest localization success-ratio, but the fastest localization speed. The localization success-ratio of 95.38% in worst case can be considered as a high-level of success for a construction site and therefore HD⁴AR with FREAK is a good candidate for fast 3D reconstruction and localization that provides a reasonable level of localization success.

Another interesting result is that we achieved an additional speedup in localization, compared to our previous work. In our previous work, HD⁴AR matched a given image to an entire set of base images for finding correspondences between image feature points and 3D points. As a result, the localization time depends on number of base images and increases as the number of base images increases (Bae et al. 2012). With the improvements presented in this paper, HD⁴AR now directly compares feature descriptors of the image to that of 3D points in a point cloud, which reduces matching time significantly. The localization time does not depend on the number of base images anymore, but depends on the number of 3D points or texture of the given image. As shown in Table 3, our new approach further increases the speed of localization by an average factor of 2.79. Compared to the Bundler package, HD⁴AR with FREAK is now up to 30 times faster in user localization.

All experimental results as observed in Table 3 and Figure 9 prove that HD⁴AR can successfully localize a user solely based on an image and within few seconds. With HD⁴AR using FREAK descriptor, a field engineer has to wait only 3–6 seconds after he or she takes a photo of target objects to retrieve related cyber-information. This is very promising for the use of HD⁴AR in practice. Information retrieval using HD⁴AR now can be done in much less time than our previous work or the traditional means (traveling back to trailer to lookup cyber-information or carrying large stacks of drawings on site and looking up for information on demand).

Discussion and research challenges

This paper presented a high-precision vision-based mobile augmented reality system for context-aware applications. The experimental results demonstrate the applicability of the proposed system to construction sites and existing buildings. The system can successfully localize the user solely based on an image, without using any external location tracking modules. Once the user's camera is accurately localized, the overlays or cyber-information can be overlaid on top of real-world imagery. The results, shown in Figures 6 and 9, indicate the robustness of the method to dynamic changes of illumination, viewpoint, camera resolution, and scale in the image, which are typical for unordered construction photo collections. While this paper presented the initial works toward vision-based localization and AR visualization for the purpose

Table 3 Performance of localization

Data set	System	Localization success-ratio	Average localization time
Parking garage	HD ⁴ AR with FREAK	100% (50 / 50)	5.42 sec (×19.94)
	HD ⁴ AR with SURF	100% (50 / 50)	6.42 sec (×16.84)
	Bundler package*	100% (50 / 50)	18.45 sec (×5.86)**
Center for the arts	HD ⁴ AR with FREAK	95.38% (62 / 65)	3.17 sec (×29.40)
	HD ⁴ AR with SURF	98.46% (64 / 65)	3.47 sec (×26.86)
	Bundler package*	100% (65 / 65)	16.02 sec (×5.82)**
Norris hall	HD ⁴ AR with FREAK	100% (25 / 25)	93.20 sec (×1)
	HD ⁴ AR with SURF	100% (25 / 25)	4.98 sec (×7.51)
	Bundler package*	100% (25 / 25)	10.74 sec (×3.48)
Patton hall	HD ⁴ AR with FREAK	100% (25 / 25)	13.31 sec (×2.81)**
	HD ⁴ AR with SURF	100% (25 / 25)	37.38 sec (×1)
	Bundler package*	100% (25 / 25)	6.33 sec (×5.10)
Patton hall	HD ⁴ AR with FREAK	100% (25 / 25)	10.07 sec (×3.20)
	HD ⁴ AR with SURF	100% (25 / 25)	24.56 sec (×1.31)**
	Bundler package*	100% (25 / 25)	32.26 sec (×1)

*A widely-used software package for SfM (Snavely et al. 2007).

**The result using our previous approach (Bae et al. 2012).

of context-aware applications, several challenges remain. Some of the open research problems include:

- Quantifying the accuracy of image-based localization in terms of re-projection error to validate how cyber objects are precisely overlaid on top of real-world photograph.
- Quantifying the quality of 3D point cloud, which will guide users to take a minimal number of images from various sites for initial bootstrapping, e.g. 3D reconstruction.
- Further increasing the speed of localization by using supplemental information such as mobile GPS available in mobile devices to reduce data set to be matched. Minimizing the image resolution to reduce matching time is also in our focus.

Conclusion

The current practice of construction progress monitoring still has significant opportunities for improvement through the integration of cyber-information into regular site operations. The HD⁴AR system was designed to provide such cyber-information on worksites using existing and already available camera-equipped mobile devices. HD⁴AR takes vital project information, such as the expected quality of building elements, or location of elements, project schedule, and cost information, which traditionally has been difficult to access on a jobsite, and makes it mobile, accessible to on-site users. In addition, it provides an easy and intuitive method to create 3D information using 2D jobsite photographs. This content authoring capability may further facilitate the accurate exchange of project information among field personnel.

Using a set of computer vision algorithms, HD⁴AR allows users to leverage any camera-equipped mobile device to take pictures for accurate on-site localization. This vision-based and location tracking-free system can support a range of promising context-aware AEC/FM applications since it does not require the installation of new technological components on the jobsite. HD⁴AR uses image feature points as the basis for user localization and a SfM algorithm to build and match a 3D geometric model from regular smartphone camera images. Users can use a smartphone outfitted with a camera, screen, and wireless communication to upload a captured image, localize it, and then overlay the returned cyber-information on the physical objects in the photograph to which it pertains. The performance of HD⁴AR, with a localization success-ratio of 95.38% (in worst case), implies that it is possible to develop a near real-time augmented reality systems using site photographs. It takes 3–6 seconds for localization and less than an hour for point cloud generation. With everyday data collection and application of HD⁴AR, 3D point clouds can be produced very

quickly, allowing AEC/FM practitioners to easily monitor construction progress by quickly and accurately accessing relevant information. In future work, we plan to use the full IFC-based (Industry Foundation Class) BIM rather than using manually created elements to completely test HD⁴AR in an actual construction site. Enhancing localization speed to real-time with the aid of GPS information available in smartphone is also one of our focuses.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

All authors contributed extensively to the work presented in this paper. H.B. designed and implemented infrastructure-independent mobile augmented reality system (MARS), gathered and analyzed experimental data, and prepared the manuscript. M.G and J.W. designed the concept and study of infrastructure-independent MARS, supervised the overall project, and edited the manuscript. All authors read and approved the final manuscript.

Acknowledgment

This material is based upon work supported by the National Science Foundation under NSF CMMI-1200374 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.

Author details

¹Bradley Department of Electrical and Computer Engineering, Virginia Polytechnic Institute and State University, Blacksburg, VA, USA. ²Department of Civil and Environmental Engineering, Department of Computer Science, University of Illinois at Urbana-Champaign, Urbana, IL, USA. ³Department of Electrical Engineering and Computer Science, Vanderbilt University, Nashville, TN, USA.

Received: 9 February 2013 Accepted: 18 April 2013

Published: 12 June 2013

References

- Akula, M, Dong, S, Kamat, VR, Ojeda, L, Borrell, A, Borenstein, J. (2011). Integration of infrastructure based positioning systems and inertial navigation for ubiquitous context-aware engineering applications. *Automation in Construction*, 25(4), 640–655.
- Alahi, A, Ortiz, R, Vanderghyest, P (2012). FREAK: Fast Retina Keypoint, *Proceedings of 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2012)* (pp. 510–517).
- Anumba, C, Aziz, Z. (2006). Case studies of intelligent context-aware services delivery in AEC/FM. *Lecture Notes in Computer Science*, 4200, 23–31.
- Arya, S, Mount, DM, Netanyahu, NS, Silverman, R, Wu, AY. (1998). An optimal algorithm for approximate nearest neighbor searching in fixed dimensions. *Journal of the ACM*, 45(6), 891–923.
- Bae, H, Golparvar-Fard, M, White, J (2012). Enhanced HD⁴AR (Hybrid 4-Dimensional Augmented Reality) for Ubiquitous Context-aware AEC/FM Applications, *Proceedings of 12th international conference on construction applications of virtual reality (CONVR 2012)* (pp. 253–262).
- Bay, H, Ess, A, Tuytelaars, T, Gool, LV. (2008). Speeded-up Robust Features (SURF). *Computer Vision and Image Understanding*, 110(3), 346–359.
- Behzadan, A H, Kamat, VR. (2007). Georeferenced registration of construction graphics in mobile outdoor augmented reality. *Journal of Computing in Civil Engineering*, 21(4), 247–258.
- Behzadan, A H, Aziz, Z, Anumba, CJ, Kamat, VR. (2008). Ubiquitous location tracking for context-specific information delivery on construction sites. *Automation in Construction*, 17(6), 737–748.
- Carozza, L, Tingdahl, D, Bosché, F, Van Gool, L. (2012). Markerless Vision-Based Augmented Reality for Urban Planning. *Journal of Computer-Aided Civil and Infrastructure Engineering*, 2012. doi:10.1111/j.1467-8667.2012.00798.x.
- Chen, Y, Kamara, J. (2011). A framework for using mobile computing for information management on construction sites. *Automation in Construction*, 15, 605–613.

- Chen, Y C, Chi, HL, Hung, WH, Kang, SC. (2011). Use of tangible and augmented reality models in engineering graphics courses. *Journal of Professional Issues in Engineering Education & Practice*, 137(4), 267–276.
- Cheng, M Y, Chen, J.C. (2002). Integrating barcode and GIS for monitoring construction progress. *Automation in Construction*, 11, 23–33.
- Chi, H L, Kang, SC, Wang, X. (2013). Research trends and opportunities of augmented reality applications in architecture, engineering, and construction. *Automation in Construction*. doi:10.1016/j.jbbr.2011.03.031.
- Dunston, P, Wang, X, Billinghamurst, M, Hampson, B, et al. (2003). Mixed reality benefits for design perception. *National Institute of Standards (NIST) Special Publication*, 191–196.
- Feng, C, Kamat, VR (2012). Augmented reality markers as spatial indices for indoor mobile AEC/FM applications. *Proceedings of 12th international conference on construction applications of virtual reality (CONVR 2012)* (pp. 235–242).
- Golparvar-Fard, M, Peña Mora, F, Savarese, S. (2012). Automated model-based progress monitoring using unordered daily construction photographs and IFC as-planned models. *ASCE Journal of Computing in Civil Engineering*, 2012. in press.
- Golparvar-Fard, M, Peña Mora, F, Savarese, S. (2011). Integrated sequential as-built and as-planned representation with D⁴AR tools in support of decision-making tasks in the AEC/FM industry. *Journal of Construction Engineering and Management*, 137(12), 1099–1116.
- Golparvar-Fard, M, Savarese, S, Peña Mora, F (2010). Automated model-based recognition of progress using daily construction photographs and IFC-based 4D Models. *Proceedings of 2010 construction research congress* (pp. 51–60).
- Golparvar-Fard, M, Peña Mora, F, Arboleda, CA, Lee, S. (2009a). Visualization of construction progress monitoring with 4D simulation model overlaid on time-lapsed photographs. *Journal of Computing in Civil Engineering*, 23(6), 391–404.
- Golparvar-Fard, M, Peña Mora, F, Savarese, S. (2009b). Application of D4AR–A 4-Dimensional augmented reality model for automating construction progress monitoring data collection, processing and communication. *ITcon, Special Issue on Next Generation Construction II: Technology Foresight, Future Studies, Roadmapping, and Scenario Planning*, 14, 129–153.
- Hakkarainen, M, Woodward, C, Billinghamurst, M (2008). Augmented assembly using a mobile phone. *Proceedings of 7th IEEE/ACM international symposium on mixed and augmented reality (ISMAR 2008)* (pp. 167–168).
- Hammad, A, Wang, H, Mudur, SP. (2009). Distributed augmented reality for visualizing collaborative construction tasks. *Journal of Computing in Civil Engineering*, 23(6), 418–427.
- Hartley, R, Zisserman, A (2004). *Multiple view geometry in computer vision*, Cambridge, UK: Cambridge University Press .
- Hou, L, Wang, X (2010). Application of augmented reality technology in improving assembly task proficiency. *CD proceedings of the 10th international conference on construction applications of virtual reality* (pp. 4–5).
- Irizary, J, Gheisari, M, Williams, G, Walker, BN. (2012). InfoSPOT: A mobile Augmented Reality method for accessing building information through a situation awareness approach. *Automation in Construction*, 2012. in press.
- Khoury, H, Kamat, VR. (2009). High-precision identification of contextual information in location-aware engineering applications. *Advanced Engineering Informatics*, 23(4), 483–496.
- Kiziltas, S, Akinci, B, Ergen, E, Tang, P, Gordon, C. (2008). Technological assessment and process implications of field data capture technologies for construction and facility/infrastructure management. *Journal of Informan Technology in Construction (ITcon), Special Issue on Sensors in Construction and Infrastructure Management*, 13, 134–154.
- Kuula, T, Piira, K, Seisto, A, Hakkarainen, M, Woodward, C (2012). User requirements for mobile AR and BIM utilization in building life cycle management. *Proceedings of 12th international conference on construction applications of virtual reality (CONVR 2012)* (pp. 203–211).
- Lee, S, Akin, Ö. (2011). Augmented reality-based computational fieldwork support for equipment operations and maintenance. *Automation in Construction*, 20(4), 338–352.
- Lowe, D G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2), 91–110.
- Navon, R, Sacks, R. (2007). Assessing research issues in Automated Project Performance Control (APPC). *Automation in Construction*, 16(4), 474–484.
- Nistér, D. (2004). An efficient solution to the five-point relative pose problem. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(6), 756–777.
- Ojeda, L, Borenstein, J (2007). Personal dead-reckoning system for GPS-denied Environments. *Proceedings of the 2007 IEEE international workshop on safety, security and rescue robotics (SSRR): 27–29 September; Rome, Italy* (pp. 1–6).
- Pasman, W, Woodward, C (2003). Implementation of an Augmented Reality System on a PDA. *Proceedings of the second IEEE and ACM international symposium on mixed and augmented reality (ISMAR 2003)* (pp. 276–277).
- Schall, G, Mendez, E, Kruijff, E, Veas, E, Junghanns, S, Reitingner, B, Schmalstieg, D. (2009). Handheld augmented reality for underground infrastructure visualization. *Personal and Ubiquitous Computing*, 13(4), 281–291.
- Shin, D H, Dunston, PS. (2008). Identification of application areas for augmented reality in industrial construction based on technology suitability. *Automation in Construction*, 17(7), 882–894.
- Snaveily, N, Seitz, SM, Szeliski, R. (2007). Modeling the world from internet photo collections. *International Journal of Computer Vision*, 80(2), 189–210.
- Wang, X (2008). Improving human-machine interfaces for construction equipment operations with mixed and augmented reality. In Balaguer, C, Abderrahim, M (Eds.), *Robotics and Automation in Construction* (pp. 349–362). Vienna, Austria: InTech.
- Wang, X, Dunston, PS. (2006). Compatibility issues in Augmented Reality systems for AEC: An experimental prototype study. *Automation in Construction*, 15(3), 314–326.
- Woodward, C, Hakkarainen, M, Korkalo, O, Kantonen, T, Aittala, M, Rainio, K, Kähkönen, K (2010). Mixed reality for mobile construction site visualization and communication. *Proceedings of 10th international conference on construction applications of virtual reality (CONVR 2010)* (pp. 35–44).
- Wu, C, Agarwal, S, Curless, B, Seitz, SM (2011). Multicore bundle adjustment. *Proceedings of 2011 IEEE conference on computer vision and pattern recognition (CVPR 2011)* (pp. 3057–3064).
- Yabuki, N, Miyashita, K, Fukuda, T. (2011). An invisible height evaluation system for building height regulation to preserve good landscapes using augmented reality. *Automation in Construction*, 20(3), 228–235.
- Yeh, K C, Tsai, MH, Kang, SC. (2012). On-site building information retrieval by using projection-based Augmented Reality. *Journal of Computing in Civil Engineering*, 26(3), 342–355.

doi:10.1186/2213-7459-1-3

Cite this article as: Bae et al.: High-precision vision-based mobile augmented reality system for context-aware architectural, engineering, construction and facility management (AEC/FM) applications. *Visualization in Engineering* 2013 **1**:3.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com