# LIVE MOBILE PANORAMIC HIGH ACCURACY AUGMENTED REALITY FOR ENGINEERING AND CONSTRUCTION

***Stéphane Côté,*** Bentley Systems, Québec, Canada, stephane.cote@bentley.com
***Philippe Trudel*[1]*,*** Bentley Systems, Québec, Canada
***Marc-Antoine Desbiens*[1]*,*** Bentley Systems, Québec, Canada
***Mathieu Giguère*[1]*,*** Bentley Systems, Québec, Canada
***Rob Snyder,*** Bentley Systems, Lexington, KY, USA

***ABSTRACT:*** *Augmented reality finds many potential uses in the infrastructure world. However, the work done by architects and engineers has potential impacts on people's lives. Therefore, the data they base their decisions upon must be accurate and reliable. Unfortunately, so far augmented reality has failed to provide the level of accuracy and robustness that would be required for engineering and construction work using a portable setup. Recent work has shown that panorama based augmentation can provide a level of accuracy that is higher than standard video-based augmentation methods, because of its wider field of view. In this paper, we present a live mobile augmentation method based on panoramic video. The environment is captured live using a high resolution panoramic video camera installed on top of a tripod, and positioned in the area to be augmented. The system is first initialized by the user, who aligns the 3D model of the environment with the panoramic stream. The live scene is then augmented with a 3D CAD model, the augmenting elements being properly occluded by live moving objects in the scene. To augment the scene from a different vantage point, the user grabs the tripod and carries it to the new location. During that time, the system calculates the camera position by tracking optical features identified on the panoramic video stream. When the user places the tripod back on the ground, the system automatically resumes augmentation from the new position. The system was tested in indoor and outdoor conditions. Results demonstrate high tracking accuracy, jitter free augmentation, and that the setup is sufficiently portable to be used on site.*

## INTRODUCTION

Augmented reality, which consists of overlaying virtual data with the physical world, has an enormous potential in the AEC world. By aligning model data with reality, AR could enable a wide range of potentially very useful applications including: building site monitoring & planning, asset identification and query, systems monitoring, remote site work planning, surveying, safety warning systems, etc. Since these involve assets of the built environment as well as virtual data related with those assets, AEC tasks are actually ideal candidates for the implementation of AR applications.

Decisions taken by architects, engineers and builders have a direct impact on public safety. They must therefore be supported by accurate and reliable data. Augmented reality applications in the AEC world would therefore need to be very accurate. Unfortunately, while approximate, low accuracy augmentations are easy to obtain, accurate AR is very hard to achieve.

For a long time, the main difficulty with augmented reality has been (and still is) registration: the capacity to align properly the 3D model and data with the corresponding physical objects. That capacity is extremely important: if an engineer uses an AR app on site to "click" on a valve box cover to query its maintenance information, he most likely wants information about that specific box cover, and not the one located 30 cm next to it. In the AEC world, inaccurate AR applications could lead to incorrect interpretations and therefore bad decisions.

For accurate AR to be possible, one must know the exact position and orientation (the "pose") of the camera. While an approximate camera pose can be obtained relatively easily using basic and inexpensive sensors generally available on tablets and smart phones (GPS, orientation sensor, and accelerometer), an accurate pose is extremely difficult to obtain without a complex an expensive setup. Poirier (2011) estimated that to augment an object located 2 meters away with a 1-pixel accuracy using a 640 × 480 pixel resolution camera, the exact camera pose needs to be known within 0.09 degree for orientation and 3.5 mm for position. Such a level of

[1] Current affiliation : Dept. Electrical & Computer Engineering, Sherbrooke University, QC Canada.

accuracy can be obtained using a complex setup (for example: limited range tracking systems used for virtual reality), which unfortunately is incompatible with outdoor mobile augmentation.

Considering the various limitations of pose measurement hardware, the problem of image-based tracking has received a lot of attention. By identifying and matching features that appear on sequential frames of a live video stream and matching those with a 3D model of the environment, it is possible to calculate the camera pose. Such methods have now reached a point where they can be used to capture limited size environments and track a moving camera in real time. However, such methods are still far from perfect. One of the main limitations of those techniques is the fact that most cameras have a limited field of view – typically about 60 x 30 degrees, up to 120 degrees diagonally. Since image-based tracking is dependent on tracking identified features, tracking will only be possible if such features exist in the first place. Although visual features are omnipresent in our world, features are not always suitable, or not always present in a sufficient number for tracking, for instance: moving targets (vehicles, tree leaves blown by wind) or repetitive patterns (brick wall), etc. Sometimes features may just be undetectable: low contrast areas (shadow), uniform surfaces (painted walls, sky), etc. Naturally, the use of narrow field of view cameras makes the situation even worse, as it increases the chances of capturing zones of the physical environment that are unsuitable for tracking. In addition, such cameras limit the capacity to view features over long distances, which limits the accuracy of the resulting pose (Lemaire and Lacroix, 2007). Another problem is related with user's movements: A tablet is held in user's hands. It is therefore subjected to constant movement – making accurate tracking even more challenging to achieve in real time.

Over the past few years, work has been done on the augmentation of static panoramic images (Côté 2011a, 2011b, 2012; Wither et al., 2011; Hill et al., 2011). Static images can be augmented more easily than video: since the camera position is fixed, they require no position tracking. In addition, a 3D model can be aligned more accurately with panoramic images because the alignment can be done on features distributed around the $360^{o}$ field of view, increasing the chance of capturing areas that are suitable for tracking (Argyros et al., 2001). Unfortunately, static images become out of date from the moment they are captured. Moreover, because they are static, the augmentation of such images can incorporate no dynamic event. What would be nice would be to develop an augmentation method that shares the advantages of panoramic images (accurate and stable augmentation) with those of live cameras (real time augmentation).

In this paper, we propose an augmentation system that circumvents the limitations of standard aperture cameras and of static panoramic images by providing a stable and accurate, yet mobile and live augmentation experience. We propose an augmentation system based on panoramic video streams. The system augments a live panoramic scene in real time. Features identified on sequential frames of a panoramic stream are tracked as the camera moves. The location of each feature in the panoramic stream enables the system to calculate the camera position, as it is being moved in the environment. We implemented and tested our method in a real environment, both indoor and outdoor. Our qualitative results confirm that our system can track the camera position in real time, and provides stable augmentations that show no jitter. We envision that such a system could be used to implement high accuracy augmentation systems.

## RELATED WORK

Live augmented reality has been invested by a large number of investigators. However, only a few of them worked with panoramic imagery. Panoramic tracking has been studied by Jogan and Leonardis (2000) who present a method for robust localization using panoramic images in a pre-learned environment, and Fiala and Basu (2004) who show a robot navigation system based on panoramic landmark vertex and line tracking. Langlotz et al. (2011) built a system where 3 DOF camera tracking can be obtained using a pre-registered panoramic environment.

Static panoramic augmented environment have been described by Côté (2011a, 2011b, 2012) and Wither et al., (2011). Langlotz et al. (2011) demonstrated live augmentation of pre-recorded video from a fixed position. Hill et al. (2011) showed a mirror world augmentation system in which pre-captured panoramic images of the environment were augmented when the user stood approximately at their image capture position. In these systems, static panoramic images were used. Those offer the advantage of providing precise augmentation (since no camera tracking is required). Augmentation based on panoramic media also has the potential of being much more accurate because of the numerous points of control located all around the camera that can be tracked over long distances (Lemaire and Lacroix, 2007) and because of the increased chance of capturing areas of the environment that are suitable for tracking (Wither et al., 2011).

## METHOD

## Data

Our panorama-based augmentation system requires 3 types of data: live panoramic video stream, tracking model and augmentation model.

**Live panoramic video stream**

The live panoramic video stream is used as a representation of the physical world. It is displayed on screen with the overlaid augmentation. The streams were captured live using a Ladybug 3 panoramic camera from Point Grey Research (see Figure 1A). The camera was connected to a laptop computer (see Figure 1B), used as a "server", through an IEEE 1394b FireWire 800 connection allowing 800 Mbps of data transfer. The video streams received from each of the 5 individual camera sensors were processed in real time into a single stitched and color corrected equirectangular panoramic stream. The live stitching and color processing program was developed in C++ using the Ladybug SDK. On a quad core laptop, our panorama processing program could achieve 15 FPS for panorama resolution of 3500 × 1750 pixels. The panoramic frames were then transferred live to a second laptop (see Figure 1C), used as a "client" that processed the stream and displayed the augmentation, via a 1 Gbps Ethernet connection, at a rate of about 5 images per second (uncompressed). We used 2 laptops because the capture and processing of live panoramic video occupied the first laptop full time, leaving no processing time available for tracking and augmentation.

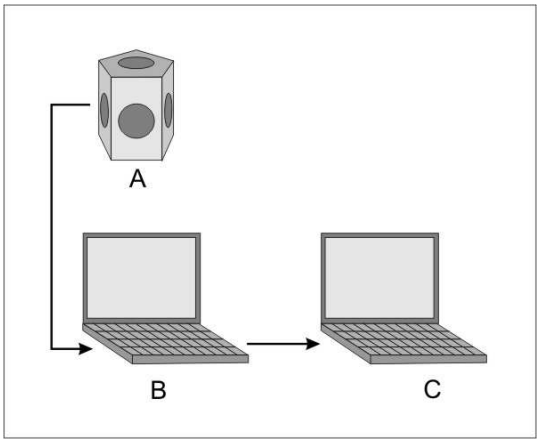


Figure 1: Hardware setup used for the experiment. A panoramic camera (A) is connected to a first (server) laptop (B) via a Firewire 800 connection. Laptop (B) is connected to a second (client) laptop (C) via a 1 Gbps Ethernet connection.

**Tracking model**

The tracking model is used as a basis for camera tracking. In this experiment, the tracking algorithm relies on a very simple tracking model obtained from a CAD model of the test area. It is composed of flat or poorly detailed surfaces that represent building walls, floors, and road surface. A tracking model containing only few details helps keep the tracking process fast.

**Augmentation model**

The augmentation model contains the 3D data to use for augmentation. It could contain, for instance, elements that represent hidden assets such as pipes, cables, structure, etc. The tracking and augmentation models are aligned and share the same georeference. In our experiment, the augmentation model is a detailed CAD model of the test area. Initially stored in DGN format, it was exported to our augmentation application that is based on Ogre 3D. The augmentation model is kept invisible in the augmented view until augmentation is required by the user.

## Video augmentation

### Initialization

The first part of the augmentation session is an initialization phase, in which the tracking model is aligned with the first frame of the panoramic stream. This step is required once at the beginning of the augmentation session, or when the system has lost track of the camera position. Although that step could be made automatic, in our system it is achieved manually. The camera is installed on a tripod at a fixed location, the augmentation application is loaded, and the panoramic video stream displayed on the client laptop. The camera's approximate position is also located using a GPS or manually selected on a map by the user. The georeferenced 3D tracking model is then displayed on screen overlaid to the panoramic stream, at approximately the same location (see Figure 2, left). The user has then the possibility to rotate the model, to roughly align model features with corresponding image features (e.g. building vertices). Then, he enters a set of correspondences, clicking on a model feature first, then clicking on the corresponding image feature. A minimum of 4 correspondences is required, while 7 or more, well distributed around the camera, is ideal. The correspondences are then used to calculate the camera pose with respect to the tracking model using the method proposed in (Poirier, 2011). The resulting pose is then used to accurately align the tracking model to the panoramic image (see Figure 2, right). It takes approximately 30 seconds to 1 minute for a trained user to find and select the required correspondences. Pose calculation time is less than 1 sec.

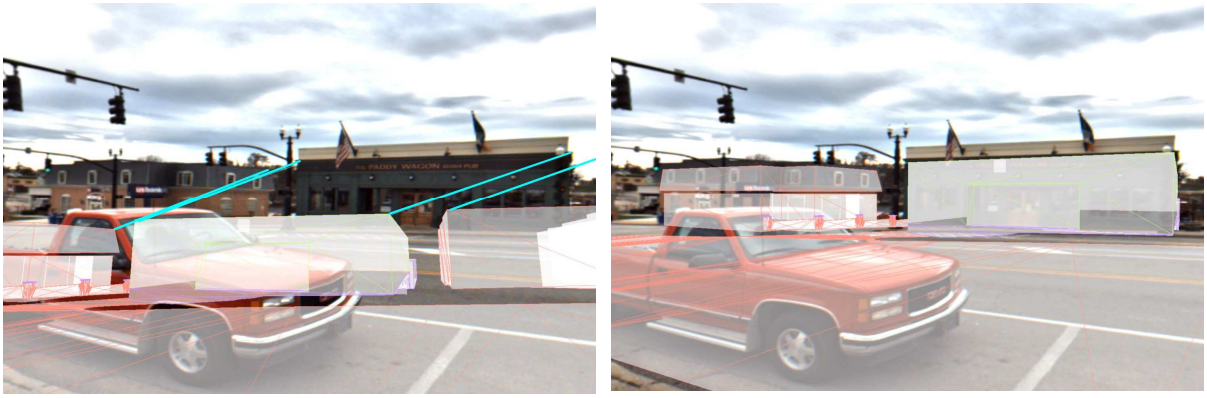

Figure 2: Initialization process. Left: Correspondences selected by user. Right: The pose calculated from those correspondences is used to align the panorama and the tracking model.

### Tracking

Once the initial camera position was obtained through initialization, the live camera pose must be obtained as it is being moved in the environment. That can be achieved through image tracking. We proposed and implemented a very basic image tracking algorithm: while the camera is still at its initial position, SURF features are extracted from a first frame of the panoramic stream using SURF GPU implemented on OpenCV 2.4. Those features are then projected onto the 3D tracking model from the camera position obtained through initialization. The projected location of each feature is considered as being the most likely 3D location of those image features in the physical world. A second frame of the video stream is then analyzed: SURF features are first identified, then matched with those of the first image. Only the best matches are used. The new camera pose is then calculated using the same method used in the initialization (Poirier 2011), but this time the correspondences are generated automatically based on those matches. Features captured on moving targets, or badly matched features are identified using their reprojection error, and eliminated from the pose calculation. Each feature of the second frame is then associated with a 3D position via projection, and the process restarts for a third frame. Keyframes were used to minimize drift. The algorithm used is very simple, but sufficient for us to prove the concept of panoramic tracking and augmentation.

### Augmentation

In our current prototype, augmentation is achieved through a virtual excavation feature that lets the user see through walls, floor and ceiling. The augmentation technique is similar to the one described by (Schall *et al.*, 2010; Côté, 2011b) for augmenting subsurface utilities. In these projects, a virtual excavation is drawn on the surface of the road, revealing hidden infrastructure (see Figure 3). The technique basically consists of creating a virtual hole in an object, by clipping all but some of the elements composing it. In this project, we used the

same technique but applied it on walls, floors, and ceilings.

## Use of the system

In a typical use of the system, the camera is placed on a tripod at a fixed position, and the system is initialized by the user. The camera is then carried to the first augmentation location. As it is being moved, the system tracks the camera location in real time. When the user puts the camera and its tripod back on the ground, the system knows the location of the camera, and augmentation can start right away, without the need of a new initialization step. Once the task that required augmentation from that location is complete, the user can then move the camera somewhere else and augment the world from that new location. Although the system can track the camera and augment the scene at the same time, the tracking feature can be stopped during the augmentation if the camera remains at a fixed position – that leaves more CPU power available to the augmentation application, and avoids any potential tracking error due to occlusion or some other dynamic event. The whole augmentation system introduced a lag in the video steam. Out tests revealed that on average, the augmentation was displayed about 1 to 1.5 seconds after the live events occurred.

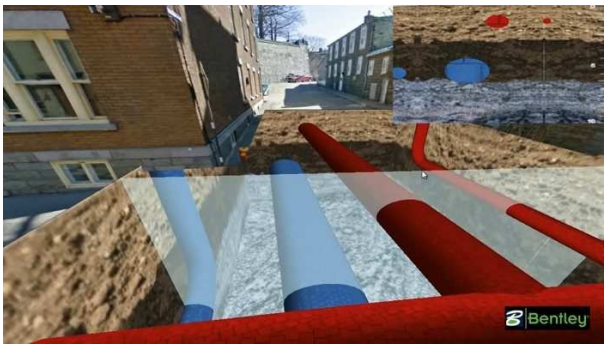


Figure 3: Virtual excavation for subsurface utilities as shown in (Côté, 2011b).

# RESULTS

## Improvements made to the model

We tested our method around and inside the Paddy Wagon Irish Pub located in Richmond, Kentucky, USA (see Figure 4, left). We chose that site because we also had a detailed CAD model (BIM) of that building (Figure 4, right), created by *McKay Snyder Architects, James McKay, Architect*, using MicroStation® and had permission to use it given by the building owner. Both our tracking and augmentation models are based on that model. The tracking model is a simplified version of the original CAD model, while the augmentation model contains only some of the invisible elements of the CAD model (structure, pipes, etc.).


Figure 4: The test site (left). The detailed 3D CAD model of the pub (right).

Initial tracking tests using pre-recorded video showed a low quality tracking, characterized by drifting augmentation as the camera was being moved, both indoor and outdoor. Our investigation with the outdoor

scene revealed that the CAD model did not cover enough of the scene surrounding the camera - we would have needed a model for many of the neighbor buildings to enable better tracking. Indoor tracking was also very deficient, and a close examination of the model and captured panoramic videos revealed several differences between the model and the actual building. Those differences could have explained the tracking difficulties we experienced. We realized we needed to make some corrections to the design model to account for changes that were made during construction.

We therefore acquired a detailed point cloud of the area using a Leica C10 scanner. A total of 25 high density scans were completed indoor and outdoor the pub, and merged together into a point cloud containing over 750 million points (see Figure 5). The point cloud was used to create a block model of the surrounding buildings and a basic surface model of the road surface. The superposition of the 3D model and the point cloud revealed major differences between the 2 (see Figure 6). For instance, it turned out that the outer walls of the building do not represent a perfect rectangle, the building being slightly “skewed”. The actual differences between the model outer walls and the actual physical wall were, in some areas, as large as 8.2 cm. That difference could explain some of the augmentation discrepancies we had observed. The point cloud was therefore used to fine tune the indoor model to fit with the actual physical walls and bar structure. It was also used to add new buildings and road surface to the outdoor model to enable more stable outdoor tracking.


Figure 5: Point cloud acquired in the test area. Neighbor buildings (top left); Close-up of façade (top right). Interior (bottom).

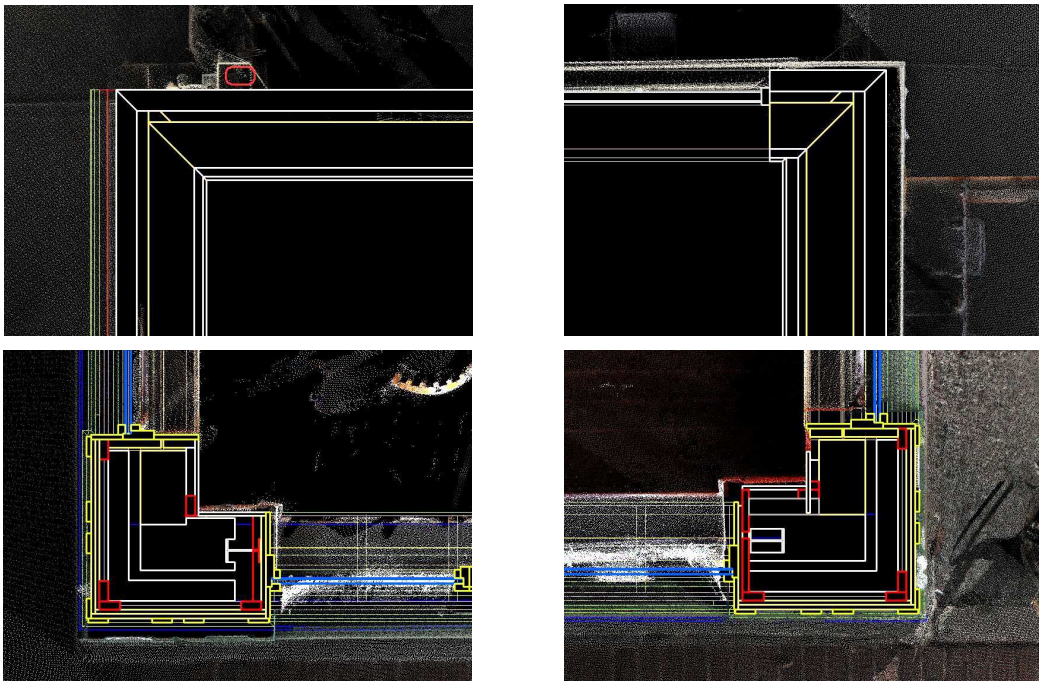

Figure 6: Detailed view of the 4 corners of the building, in a top view, showing the difference between the model and the point cloud. Differences observed in the top right (8.2 cm) and bottom left corners (7.6 cm) cannot be fixed by model rotation or translation.

## Experimental tests

### Outdoor tracking and augmentation

The method was tested on site using 2 quad core Lenovo W520 laptops equipped with 12 Gb or RAM, and installed on top of each other on a harness worn by the tester (see Figure 7). The panoramic camera was installed on top of a tripod and transported around the building. Although the whole setup could be carried by one user, in practice it was much easier when assisted by another user.

Figure 7: Setup for carrying the computers and camera on site.

Results for basic camera tracking were excellent: an outdoor test where the camera was moved and rotated like a reversed pendulum showed no jitter and only a small (not quantified) amount of drift (see Figure 8). The model remained well attached to the physical world during camera movement. We could achieve a tracking rate of 2-3 fps while moving the camera around the building. We also tested augmentation quality: on the opposite side of the building, we displayed a virtual excavation on the wall surface, that reveals model elements located inside the wall, as well as other objects located inside the pub model (see Figure 9). The excavation could be moved freely, live, along the wall surface.

The augmentation being displayed on top of the live video stream, appeared on top of everything, even objects

and people located between the camera and the wall being augmented. To avoid that undesired effect, we implemented a basic occlusion detection algorithm based on object movement. It allowed the superposition of moving objects on top of the augmentation. This way, a user can point at and draw the location of hidden pipes, as his own image is not occluded by the augmentation (see Figure 10).

Figure 8: Two frames extracted from the camera rotation experiment. Tracking produced no augmentation jitter, but a small amount drift accumulates over time (see rightmost part of right image).

Figure 9: Two frames extracted from the wall augmentation experiment. Virtual excavation reveals pipes hidden inside the wall and other elements inside the pub.

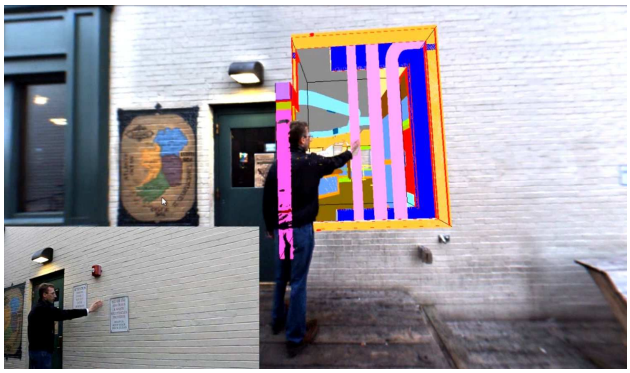

Figure 10: Dynamic occlusion detection, based on user's movements, allows proper occlusion between user and augmentation.

**Indoor tracking and augmentation**

Inside the pub, the panoramic camera was installed on top of a tripod and a dolly, for smooth movement. The pub interior is exceptionally rich in features: wall decoration, tables, bar, bottles, etc. (see Figure 11). Therefore, the tracking was exceptionally stable. Results from our indoor tracking experiment showed very stable tracking, without jitter, for both camera translation and rotation (see Figure 12).

Although the model was carefully manually aligned with the panorama at the beginning of the augmentation

session, we observed an increasing offset between the panoramic stream and the model. The exact origin of that drift is unknown, but we presume it is related with the tracking technique. It will be the subject of a future investigation.

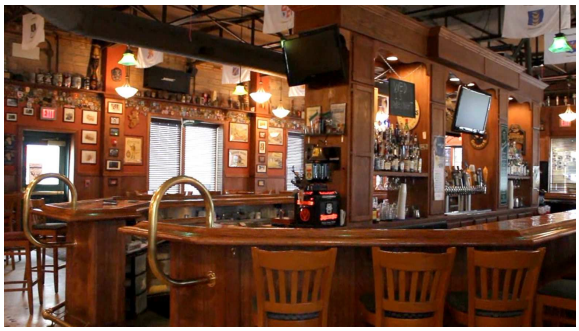

Figure 11: Indoor environment.

Figure 12: Two frames extracted from our indoor tracking experiment.

## CONCLUSION & FUTURE WORK

Our experiment showed that it is possible to obtain accurate and live augmentation in a building environment, both outdoor and indoor, using a panoramic video camera. Our results showed a stable tracking, probably because of the panoramic camera's large field of view that increases the chance of capturing areas suitable for tracking. The experiment also helped us identify the conditions that make such a stable tracking possible. In particular, our results highlighted the importance of having an accurate and detailed model of the building environment.

Our results open the door to future augmented reality applications where high accuracy is required. They also highlight the importance of further studying some aspects of panorama-based augmentation. Future research efforts could be put on:

- Minimizing the lag between live events and augmented display. This could be achieved for instance by using only one, faster computer and improving parallelism between processes.
- Improving the tracking algorithm, which currently accumulates drift and is too slow on large images.
- Obtaining good tracking without requiring a laser-based model.
- Detecting camera movement and starting the tracking feature automatically.

## ACKNOWLEDGEMENTS

Many thanks to NSERC for financial support via the Industrial Undergraduate Student Research Award program.

## REFERENCES


Argyros A.A., Bekris K.E. and Orphanoudakis S.C., 2001. Robot Homing based on Corner Tracking in a Sequence of Panoramic Images. In Proceedings of the CVPR 2001 conference.

Côté S. (2011a). Augmented reality for infrastructure: a first step. Published electronically on *BE Communities*.

Côté S. (2011b). Augmented reality for underground infrastructure: the problem of spatial perception. Published electronically on *BE Communities*.

Côté S. (2012). Augmented reality for building construction and maintenance: augmenting with 2D drawings. Published electronically on *BE Communities*.

Fiala M. and Basu A., 2004. Robot Navigation Using Panoramic Landmark Tracking. *Pattern Recognition*, Vol. 37 No. 11, Nov 2004.

Hill A., Barba E., MacIntyre B., Gandy M., Davidson B. (2011). Mirror Worlds: Experimenting with Heterogeneous AR. 2011 International Symposium on Ubiquitous Virtual Reality, Jeju-si, Republic of Korea.

Jogan M. and Leonardis A., 2000. Robust Localization Using Panoramic View-Based Recognition. ICPR '00 Proceedings of the International Conference on Pattern Recognition.

Langlotz T., Degendorfer C., Mulloni A., Schall G., Reitmayr G., Schmalstieg D., 2011. Robust detection and tracking of annotations for outdoor augmented reality browsing. http://dx.doi.org/10.1016/j.cag.2011.04.004

Lemaire T. and Lacroix S. (2007). SLAM with panoramic vision, *Journal of Field Robotics*, Vol. 24.

Poirier S. (2011). Estimation de pose omnidirectionnelle dans un contexte de réalité augmentée. M.Sc. Thesis, Université Laval, 2011.

Schall G., Junghanns S. and Schmalstieg D. (2010). VIDENTE - 3D Visualization of Underground Infrastructure using Handheld Augmented Reality, in GeoHydroinformatics: Integrating GIS and Water Engineering *(CRC press)*.

Wither J., Tsai W-T., Azuma R., 2011. Indirect augmented reality. *Computers & Graphics* 35, 810–822.