Computers & Graphics 35 (2011) 854–868

Contents lists available at ScienceDirect

Computers & Graphics

journal homepage: www.elsevier.com/locate/cag

Mobile Augmented Reality

A topometric system for wide area augmented reality

Andrew P. Gee [a,*], Matthew Webb [b], Jorge Escamilla-Ambrosio [c,1], Walterio Mayol-Cuevas [a], Andrew Calway [a]

[a] University of Bristol, Department of Computer Science, Merchant Venturers Building, Woodland Road Bristol BS8 1UB, UK
[b] University of Bristol, Department of Electrical and Electronic Engineering, Merchant Venturers Building, Woodland Road, Bristol BS8 1UB, UK
[c] Instituto Nacional de Astrofísica Óptica y Electrónica, Departamento de Electrónica, Puebla, México

ARTICLE INFO

Article history:
Received 11 February 2011
Received in revised form
21 April 2011
Accepted 27 April 2011
Available online 7 May 2011

Keywords:
Topometric
Augmented reality
Visual SLAM
UWB
GPS

ABSTRACT

We describe a system designed to facilitate efficient communication of information relating to the physical world using augmented reality (AR). We bring together a range of technologies to create a system capable of operating in real-time, over wide areas and for both indoor and outdoor operations. The central concept is to integrate localised mapping and tracking based on real-time visual SLAM with global positioning from both GPS and indoor ultra-wide band (UWB) technology. The former allows accurate and repeatable creation and visualisation of AR annotations within local metric maps, whilst the latter provides a coarse global representation of the topology of the maps. We call this a 'Topometric System'. The key elements are: robust and efficient vision based tracking and mapping using a Kalman filter framework; rapid and reliable vision based relocalisation of users within local maps; user interaction mechanisms for effective annotation insertion; and an integrated framework for managing and fusing mapping and positioning data. We present the results of experiments conducted over a wide area, with indoor and outdoor operations, which demonstrates successful creation and visualisation of large numbers of AR annotations over a range of different locations.

1. Introduction

There is little doubt that augmented reality (AR) has the potential to change radically the way in which information is communicated, whether it be for entertainment, navigation, logistics, or one of its many other applications. Even in its simplest form, the ability to associate virtual content directly with 3D physical structure and to display that content *in situ* opens up a vast range of novel possibilities for storing, presenting and analysing data relative to the world around us. In entertainment, distributed gaming takes on a whole new meaning when the virtual and real worlds are merged, and annotating our surroundings with virtual signposts and routing information offers the possibility of both flexible and tailored navigation well beyond that provided by current GPS devices.

What is less clear is how this potential is to be realised in practice. Although huge progress has been made in the research and development of key elements, notably in the areas of tracking and display, it remains an open question as to how these techniques are to be best integrated into complete systems, capable of robust and scalable operation over large areas by multiple mobile users. If AR is to become commonplace and an everyday tool for a wide spectrum of users, then such operational performance will be essential. In other words, AR systems need to be moved out of the laboratory and to be developed into usable applications.

Of fundamental concern, particularly in the case of mobile AR, is how to provide a scalable and efficient tracking and mapping infrastructure. AR systems rely critically on localising the 6D pose of display devices to allow annotations to be aligned with the video from a viewing camera. For convincing display this tracking needs to be both accurate and robust. Efficient mechanisms are also needed to map the 3D environment, allowing annotations to be associated directly with physical structure. Finally, global representations are required for effective navigation and context awareness.

Many techniques have been developed to address these issues, using a variety of sensors and for both calibrated and uncalibrated scenarios. Examples are discussed in Section 2. An important concern in all these systems is the relative accuracy achieved in global positioning and local tracking. On the one hand, accurate tracking is essential for realistic annotation. However, maintaining that accuracy over wide areas for positioning is impractical without extensive calibration and instrumentation of the environment, hence severely limiting both the flexibility and scalability of the system. This is a key limiting factor in many systems. The work described in this paper seeks to address this issue.

[*] Corresponding author. Tel.: +44 117 95 45629; fax: +44 117 95 45208.
E-mail addresses: gee@cs.bris.ac.uk (A.P. Gee), mww23@cantab.net (M. Webb), jescami@inaoep.mx (J. Escamilla-Ambrosio), wmayol@cs.bris.ac.uk (W. Mayol-Cuevas), andrew@cs.bris.ac.uk (A. Calway).
[1] This work was carried out while this author was at the University of Bristol, Department of Computer Science, UK.

0097-8493/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.cag.2011.04.006

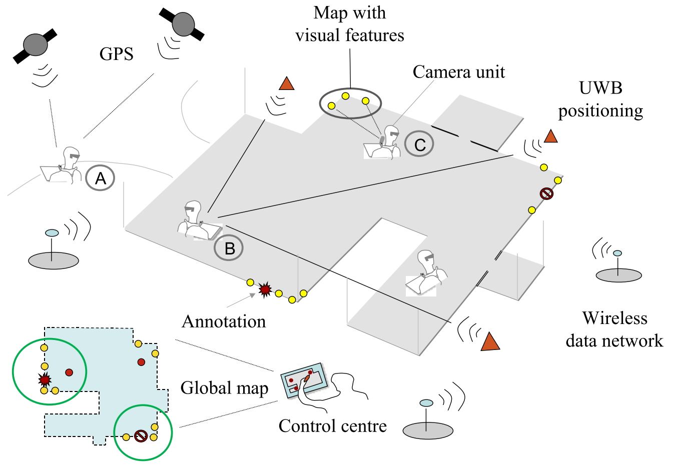

Fig. 1. Schematic overview of the ToMAR system showing multiple users authoring a scene with AR annotations based on visual SLAM and using indoor (UWB) and outdoor (GPS) absolute positioning technologies. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

We adopt a *topometric* approach, in which topological global mapping, using relatively low accuracy outdoor and indoor positioning, is integrated with higher accuracy local metric tracking and mapping (see Fig. 1). By topological, we mean a global mapping system in which the arrangement of features and places is coarsely correct but also potentially distorted by disturbances such as reflections and occlusions as commonly found in positioning systems such as GPS or UWB. For local metric operation we use vision based simultaneous localisation and mapping (SLAM) to enable the on-line creation of local feature maps, which facilitates both accurate tracking and reliable relocalisation for the insertion and visualisation of annotations. Overall, this yields a scalable system, with the global topological map providing context and coarse navigation for both indoor and outdoor scenarios, and visual SLAM allowing robust and mobile AR creation and display. Moreover, this is achieved using a relatively small number of sensors.

In the next section we review related systems and techniques, followed by an overview of our topometric system in terms of positioning and mapping in Section 3 and interactive operation in Section 4. Experimental results illustrating system performance for both indoor and outdoor operations, including a user evaluation study, are then presented in Sections 5 and 6. Initial findings of the work were previously reported in [40,41].

2. Related work

An ideal wide area AR tracking system would be capable of simultaneously providing: (i) high frequency estimates of absolute position in a global reference frame for navigation and contextual awareness and (ii) locally accurate six DoF pose for the 3D display of visual annotations. However, it has long been realised that no single sensor is currently capable of providing such a complete tracking solution over indoor and outdoor locations [1,2]. Instead, hybrid tracking approaches that combine single sensor tracking technologies, such as GPS, UWB, inertial sensors and cameras, are used to provide a combination of absolute positioning and local accuracy at the cost of increased system complexity. This flexibility in the choice and combination of sensors has led to the proposal of many different types of wide area AR systems for different applications.

Early wide area AR systems have suffered from various limitations which have prevented their wide acceptance and usage. Rekimoto et al. [3] propose a system that attaches annotations to locations in the environment prepared with physical markers that can be used to estimate the pose of a camera. This provides wide area AR capabilities but has the fundamental limitation of requiring physical modification of the environment and lacks a global overview of the locations of annotated content. Similarly, many other wide area AR systems have been developed that rely on visual markers or known models of the environment [4–7]. This is fine in locations where models are readily available but severely limits scalability due to the cost of constructing and distributing these models, although one recent system has begun to make use of fire-exit maps which are available in many buildings [8].

Other systems [9,10] have attempted to tackle these problems by introducing GPS and inertial sensors to provide a global position and orientation estimate without markers or models, and the related problem of managing multiple different tracking systems has also been considered, for example the Kalman filter 'meta-tracker' proposed to bridge the gaps between tracking systems [11]. These systems are capable of annotating unprepared areas of the environment, providing tracking at any location and producing a complete global map. However, they are limited by the typically low accuracy of global positioning sensors and the difficulties of accurately estimating global orientation indoors [11,12].

At the other end of the spectrum, systems have emerged which construct sets of highly accurate independent local maps using visual SLAM techniques. For example, Castle et al. [13] present a visual SLAM system that creates small submaps that are kept disjoint and then compared against an input image to detect that the user is in the same area once again. Instead of continuously tracking the global location of the user, such a system switches between a 'tracking' mode when it is viewing a mapped part of the environment and a 'relocalisation' mode when it is lost and searching for recognisable mapped features. Providing that the relocalisation method is robust and reliable, this approach can be highly efficient, since only the areas which are to be annotated need to be mapped. However, it does not provide the user with a global overview of the area of interest and raises the question of how the user might be expected to navigate between annotated regions or even know that an annotation will be available to view in a specific location. For this reason, the scalability of a system that completely disregards any global reference appears unrealistic.

Intermediate solutions also exist where the single global map is broken into local submaps with known global poses, although these techniques are primarily found in the robotics literature and have not yet been widely applied to wide area AR systems. Fritz et al. [14] use visual feature descriptors to enable building recognition and 2D annotation and GPS provides gating based on location. Schleicher et al. [15] link local metric visual SLAM maps together using GPS to build a global topological graph which can then be optimised using loop-closure constraints. Tracking within the local maps is highly accurate but the global position estimate is still limited by the typically low accuracy of the global positioning and orientation sensors. This can be improved by using visual odometry to provide more accurate estimation of the relative transformations between nodes in the graph or by ensuring that local maps share visual features which can be used to constrain the global graph [16]. However, maintaining an accurate global map over a large area can incur significant overheads to maintain global consistency and detect and manage loop-closures and may be a waste of effort in applications which only require the environment to be sparsely annotated.

3. Topometric mobile AR

Our proposed system falls somewhere in the middle of the above spectrum, combining global topological positioning with local metric mapping. We call this topometric mobile AR (ToMAR). Fig. 1 shows an overview of the system. Using a method similar to Castle et al. [13], we use visual SLAM to build highly accurate local maps in locations we wish to annotate but, in contrast to their system, we also build a global map based on GPS and UWB positioning to assist users in locating and navigating to annotations. We use a gating mechanism on global position combined with ranking based on visual appearance to relocalise users within the local maps.

It is important to stress, however, that unlike Schleicher et al. [15] and many other authors of large area visual SLAM systems, we do not attempt to build a highly accurate global map. Instead we aim simply to build a global map with sufficient resolution to enable a user to identify locations where there is interesting annotated content and navigate to them. For this reason, we value contextual information, such as the locations of walls and doors, over and above global positioning accuracy and we provide tools that allow the user to annotate the global map with this type of information, similar to the techniques presented by Baillot et al. [9]. This enables us to completely ignore the problems of global orientation estimation and optimisation of the global map. In this way, we avoid much of the overhead and complexities of combining different types of sensors to build a full global map whilst still providing the essential functionality of a scalable wide area AR system. A limitation of the approach is that annotations can only be visualised inside their own local maps, and not between maps, since we do not estimate the global orientation and scale of the local maps. However, since multiple annotations can be added to each local map, this only becomes a significant drawback when co-visibility of annotations in widely separated locations is required.

3.1. Global topological positioning

As shown in Figs. 1 and 2, there are three different modes of global positioning within the ToMAR system: (A) GPS, (B) floor plan maps and (C) UWB. A communications infrastructure (in our case using WiFi) links users and allows them to share their maps and annotations with each other and with a control centre in order to generate a global map as shown in the bottom left of Fig. 1. Local SLAM maps (green circles) are positioned in the global reference frame at creation using a 3D position estimate from the GPS, UWB or user interaction. The accuracy of this registration will depend upon the positioning method in use at the time of authoring. Again, we are satisfied if the local maps are only roughly positioned in the global map, since visible annotations will always be displayed with local accuracy relative to the camera thanks to the automatic relocalisation and full 6D camera pose estimation in the local map. This also means that we do not need to refine the initial estimates of the local map positions in the global map and also that we are able to treat them independently and avoid the need to maintain consistency between the local map contents.

We employ GPS positioning when available which provides an accepted alignment with an absolute frame of reference. Although some work has shown that GPS may be usable in some indoor setups [17], we also employ a UWB indoor positioning system composed of multiple transponders [18]. We assume that the UWB system is installed in the building and has already been calibrated with respect to the GPS reference frame. In a typical indoor environment, the UWB system provides 3D positioning to at least metre-level accuracy. However, the accuracy of both GPS and UWB is affected by obstructions in the lines of sight between units and reflective surfaces in the environment that add multipath effects and make accurate absolute positioning a challenging problem. Switching between the UWB and GPS systems is completely transparent to users, since the ToMAR system is left to determine at any instance if there is coverage by one or the other system and to make a decision as to which system will be used with priority (in our case it is the more accurate UWB system). Recall that our global map is not intended to be highly accurate and is used primarily to locate and navigate to annotated regions of the environment, so unfused absolute position estimates are perfectly adequate for our requirements.

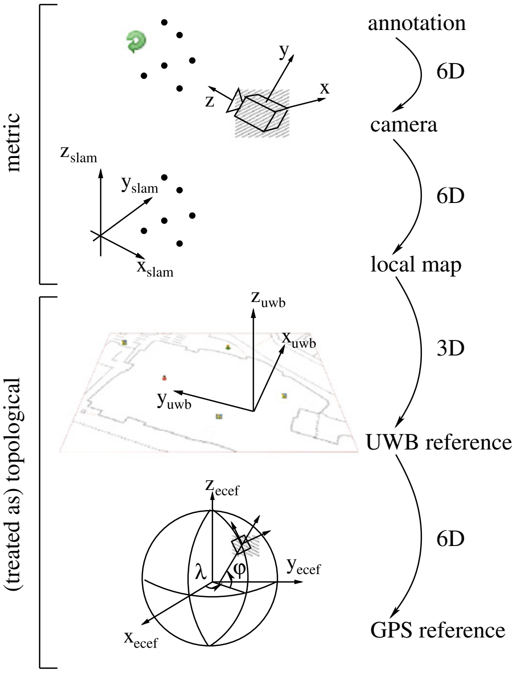

Fig. 2. Hierarchy of reference frames. The transformation between the topological and metric layers is purely translational and initialised for the origin of each local map using a 3D position estimate from GPS, UWB or user interaction. The global map displays the 3D position of the local map origins and enables navigation between annotated regions. Annotations and camera pose cannot be transformed into the UWB or GPS reference frames because the global scale and orientation of the local maps is not estimated.

If the user wishes to create an annotation when neither UWB nor GPS are available, the system prompts the user to refine location interactively on a 2D map shown centred on the last trusted position fix. The user can then simply select an approximate location in this map. Our system uses street maps showing only the outlines of buildings but nothing prevents the use of more detailed map representations. The maps can also potentially be extended to include architectural floor plans, if available, which can be further enhanced by user input of contextual information as described in Section 4. Once more, the end result is a topological frame of reference. By combining automatic referencing with the interactive user input we are able to perform authoring over a wide area.

3.2. Local metric mapping

The choice of visual SLAM to facilitate local authoring and display of AR annotations is an important one. It provides real-time 3D camera pose tracking and structure mapping in previously unseen environments, making it flexible and mobile, operating without the use of markers or additional instrumentation. In our case we only require to annotate in localised regions and hence we are content with mapping over small areas, something which current systems are particularly proficient at. There are a number of possible systems that could be selected for the task [19–21], using different feature matching and estimation methods. We have chosen to make use of the approach described by Chekhlov et al. [20], primarily because of its robustness and that it complements an effective mechanism for relocalisation [22]. We provide a brief overview of the method below; readers should refer to the above references for further details.

In visual SLAM the aim is to estimate the 3D position and orientation of the camera whilst simultaneously estimating depth information about features in the scene. The approach by Chekhlov et al. [20] utilises an extended Kalman filter (EKF) framework for this, in which the filter state consists of the 3D pose and a map of 3D point features. The measurements for the filter are the positions of the projected point features in successive frames captured by the camera, related to the state via perspective projection (we assume a calibrated camera). A useful interpretation is that of iterative predictor–corrector cycles as illustrated in Fig. 3. Given a current state and an assumed motion model, predictions for feature positions in the next frame can be made along with their associated uncertainty derived from the state covariances maintained in the filter. The latter provides spatial gating on image measurements as shown in Fig. 3, hence minimising computation. The most likely match for the corresponding 3D point is then used as the measurement to update the pose and 3D map using the Kalman filter equations [23].

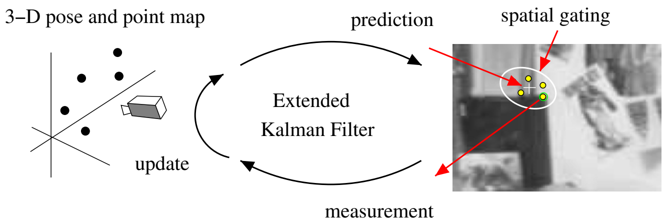

Fig. 3. Predictor–corrector operation within the extended Kalman filter for feature matching and map update in visual SLAM.

Feature matching is based on image descriptors computed around the points of interest and at multiple scales. These are histograms of spatial gradients, suitably normalised with respect to the dominant orientation, similar to the descriptors used in the SIFT algorithm [24]. These are known to provide a good degree of viewpoint invariance [25] and robustness to erratic motion [20]. Scale gating based on the estimated camera pose then enables matching of descriptors at correct scales [20]. The 3D map features are initialised using the inverse depth method [26] and potential features are determined in each frame using combined application of the FAST [27] and Shi and Tomasi [28] salient point detectors. We use a combined constant position and constant velocity motion model, switching between modes according to the success or otherwise of feature matching, hence adapting to the camera motion.

The above system allows users to build local maps in the environment and then insert AR annotations into the map using an interactive process as described in Section 4.1. Fig. 4 shows an example of one such map, with the 3D estimates of camera pose and point map shown on the left and feature matching in the current frame with the projected annotation shown on the right. The image regions denote the spatial gatings, with green indicating a successful feature match and red a failed match. Further examples of inserted annotations at 16 locations are shown on the right in Fig. 13 with their positions in a global map shown on the left (further details are given in Section 5.2). Having built and stored such maps, a key element in ToMAR is then to relocalise users in a given map as and when they are in its vicinity. We consider this in the next section.

3.3. Relocalisation

As users approach an area within which annotations have been previously created, we require the camera to be relocalised within the associated local map, enabling tracking to recommence and

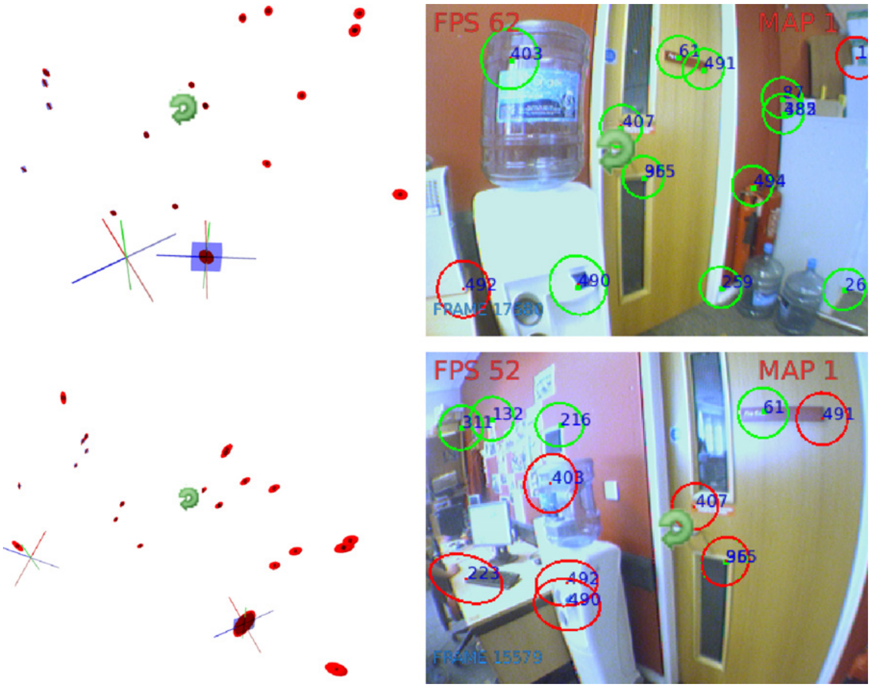

Fig. 4. Example of a local map created using visual SLAM: (left) estimated camera pose and 3D point map with covariances shown in red; (right) view through the camera, showing spatial gatings for matched (green) and unmatched (red) features along with the projected AR annotation (green rotation arrow). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

hence allow the display of the annotations. Since the maps are defined in terms of visual features, this process needs to be based primarily on appearance—essentially, features detected as the camera views the mapped area need to be rapidly matched against those associated with the stored local maps. For effective on-line operation this needs to be as robust and as fast as possible, with few or no false positive detections. There are a number of existing approaches to address this type of problem, many of which have been developed in the context of visual SLAM systems.

For example, Williams et al. [29] use a randomised trees classifier for re-detecting features, combined with a RANSAC verification step for pose initialisation. The trees are generated off-line and use a relatively large storage space—about 1.3 MB per map point [13]. The PTAM system [21] uses a method based on matching low resolution keyframes and this formed the basis of the relocalisation over multiple maps [13]. This approach is better from the point of view of data storage, however, in our experience, keyframe based localisation is prone to false positives, in particular when operating in roughly similar areas. Another popular alternative is to use visual codebooks [30], such as that used by Cummins and Newman [31] for wide area location recognition. The codebooks are usually determined off-line using an optimised clustering process and are therefore not easily updated on the fly, something which is addressed by Eade and Drummond [16].

3.4. Single map relocalisation

For ToMAR we have adapted the method described by Chekhlov et al. [22], primarily because of its efficiency in terms of memory requirements and its low false positive rates. It combines geometric consistency checks with robust visual descriptor matching, where the latter are the histograms of spatial gradients used in our visual SLAM algorithm, and fast library indexing using Haar coefficients. The latter is based on a quantisation table which is small in comparison to other approaches (e.g. using randomised trees) and can be updated on the fly. The method described by Chekhlov et al. [22] was designed to work on a single map but we extend that approach to work more efficiently with multiple maps. Furthermore, our method differs from the previous multiple map relocalisation work by Eade and Drummond [16] both in the smaller size of the descriptors used and in our use of a relatively small quantisation table created only from the 3D features in our local maps. Again, we provide a brief overview of the method here.

We assume that a map M_i of features has been built previously and the 3D geometry of features together with their visual descriptors is available. To attempt to relocalise, salient points are detected in the incoming frame. Around each point, a fixed-size patch is extracted, aligned with the dominant orientation, to give rotation invariance, and the first three Haar coefficients are extracted. These encode the rough appearance variation of the patch in the horizontal, vertical and diagonal directions, and are used to index a quantisation table Q_i. Cells in the latter contain descriptors along with their 3D positions from the stored local map for which the associated patches have similar Haar coefficients, i.e. a cell c_{ij} in Q_i contains a list of features $F=\{f_k,\ldots,f_m\}$ generated at the time M_i was created. Thus, matching is achieved by comparing the descriptor for the incoming patch with those in the cell c_{ij} and neighbouring cells as illustrated in Fig. 5.

In other words, the Haar coefficients act as a hashing mechanism to reduce the number of comparisons, hence speeding up the matching process. This is similar to their use for image matching [32]. Once candidate matches have been found, then a RANSAC method is used to compute a consistent camera pose. If successful, and if an annotation linked to M_i is visible in the current frame, it will then be displayed as an AR object. In our tests, this approach uses only about 3% of the comparisons needed by an exhaustive search. The whole process is also fast, usually relocalising within 50–300 ms, and with less than 1% false positive results and between 3% and 40% false negative results in a typical office scenario [22]. Generally, as changes occur in the environmental conditions and local scene structure, the false negative rate of the system will increase but the false positive rate remains low, due to the robustness of the combination of geometry checks with distinctive visual feature matching.

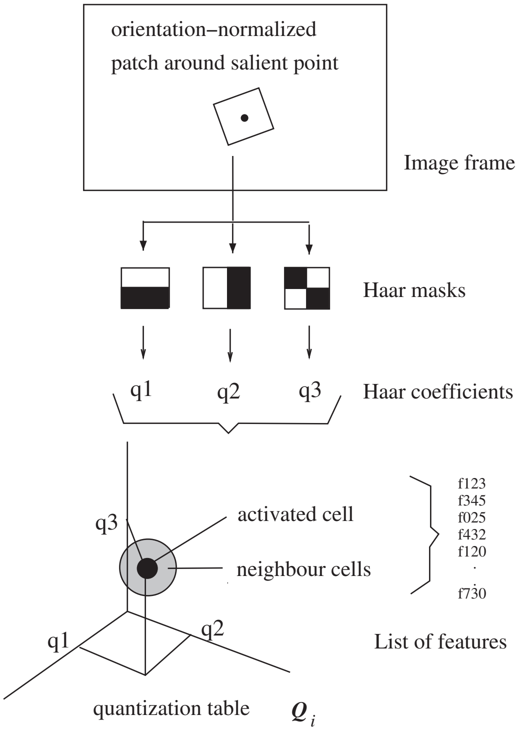

Fig. 5. Fast relocalisation using an incoming frame in a single local map is based on the computation of three Haar coefficients per saliency point which provide an index into cells in a quantisation table which contain likely matching features.

3.5. Relocalisation in multiple maps

When considering many local maps, the naive approach would be to run the above process for every map M_i individually, perhaps gated by location. When the number of maps in a vicinity is small, that process may be sufficient, but in general we would need to be prepared to run relocalisation on many maps to ensure robustness. To this end, we have developed a system of map ranking based on the single map method by combining the information of the individual quantisation tables Q_i as follows.

We create a master table QM based on all the individual tables Q_i. Indexing into QM is also via the three Haar coefficients extracted around each salient point in an incoming frame. However, the cells in QM contain a list of all the maps that have features in that cell. Therefore if a cell in QM is activated by an input patch, a list of all possible maps that have to be searched is obtained. In addition, each cell is weighted by the 'term frequency-inverse document frequency' (*tf-idf*) measure to reflect the uniqueness of a cell [30]. In this way, cells that activate for every map will have a lower weight than those that activate for fewer maps. By combining the weighted lists generated for every patch on the image, it is possible to rank all maps according to the

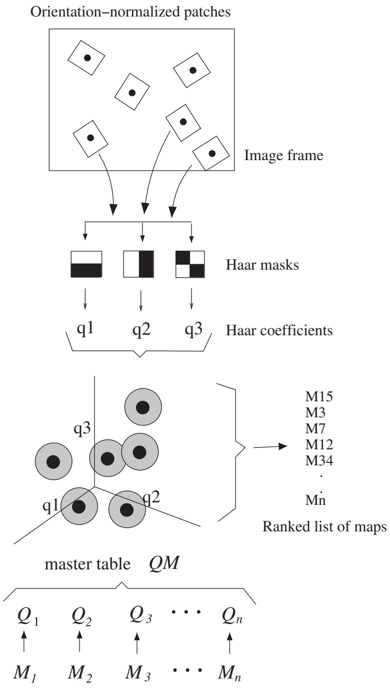

Fig. 6. When multiple maps have to be searched to attempt relocalisation, a master quantisation table *QM* assists in the ranking of the maps to speed up the process.

cosine similarity score between the *tf-idf* vectors for each map and the current image.

The process is illustrated in Fig. 6 and is very fast as we only need to look at the weighted frequency of *i* indices and rank them. The rank establishes the order in which relocalisation in the individual maps is to be attempted as per Section 3.4. When the first relocalisation is successful the process stops and switches to AR visualisation, since in our experience the single map relocalisation method very rarely produces false positives in real applications [22].

4. Interactive mapping

4.1. Defining 3D points with a touchscreen

In order to annotate points of interest in the local maps we require a suitable method for defining 3D points to which the annotations can be attached. A recent user study [33] evaluates different techniques for defining a 3D point with a touchscreen device and recommends a triangulation technique using 3D rays defined from two different viewpoints and a fixed crosshair in the centre of the touchscreen. We modify this technique by extending the number of 3D rays from two to *N*, where $N \geq 2$, allowing the user to refine the 3D position and improve accuracy. This is a well known technique from multiple view geometry [34] and is similar to the construction methods proposed by Baillot et al. [9]. However, we do not intend users to attempt the more complicated *in situ* modeling described in that work or later publications [35,36,33]. Instead, the 3D points are used to position annotations selected from a predefined library of 3D models and 2D sprites.

The 'N clicks' (NCs) method is illustrated in Fig. 7. Assuming that we know the local camera pose from visual SLAM, we begin by defining an initial view of the point of interest by aligning a fixed crosshair in the centre of the touchscreen with the corresponding point in the camera image. This defines a 3D ray in space that links the camera with the point of interest. This process can be repeated multiple times to generate a set of *N*3D rays. The 3D position of the point is then estimated by applying outlier detection and least squares optimisation [34]. Furthermore, once at least two 3D rays have been defined, the estimated 3D position of the point can be overlaid on the camera image to continuously provide visual feedback of the current 3D accuracy. This allows the user to continue refining the position until they are happy with the accuracy.

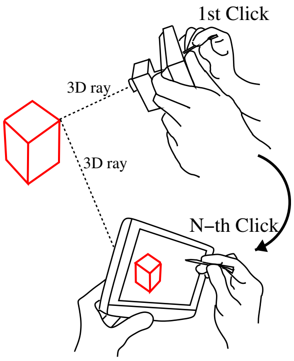

Fig. 7. The technique used to construct 3D features. In the NCs method, 3D points are constructed by triangulating multiple 3D rays.

4.2. Contextual information

The system also provides tools to sketch the internal structure of buildings. The aim of this process is not to generate a highly accurate model of the building but to provide important contextual information which can help users to navigate around the environment. Fig. 8 shows an example of the type of higher level map that can be created on top of a background street map and a comparison between the generated map and an architectural plan of our test site.

Internal structural features, such as walls and doors, can be added semi-automatically to the map by the users from a predefined library. A user stands next to the point of interest and selects the type of structure to be inserted in the map. This records the current absolute position of the user, as measured by the most accurate available absolute positioning system (UWB, GPS or user defined), and inserts the chosen structure at this location. In the case of linear structures, such as walls, this process is repeated to define the location of each endpoint in turn.

A team of users working collaboratively together can rapidly generate a map containing the essential indoor features. The accuracy of the map is limited by two main factors: the accuracy of the absolute positioning system and the accuracy of the user's real position with respect to the structures being mapped. However, since the map is primarily used to provide contextual information to human users, metre-level accuracy is sufficient to generate a map with useful information. Fig. 8 shows a comparison between the generated map and an architectural plan of our test site. The access map is correctly limited to the balcony and stairs regions of the plan (see Fig. 16 for reference image). Offsets in the structure occur where the user was unable to stand at the precise location of the feature or where the coverage from the UWB was poor.

In addition to annotating the map with physical structures, users are also able to annotate it with images captured from the camera. Images can be stamped with the 3D location of the operation taking the photograph and placed into the map.

4.3. Multiple users

A central control centre (CCC) manages the flow of data between users over a wireless LAN, enabling each user to view

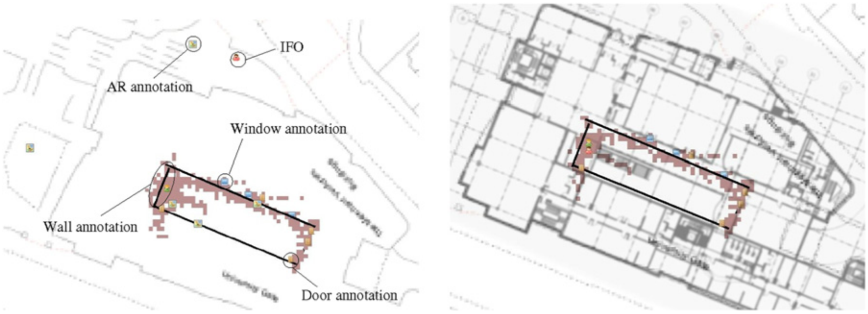

Fig. 8. Example virtual map and comparison with architect's plan of test site. (left) Structural annotations provide context and a low resolution occupancy grid access map is generated as the users move around the floorspace. (right) Alignment of the structural elements is within the expected margin of error of the UWB tracking system.

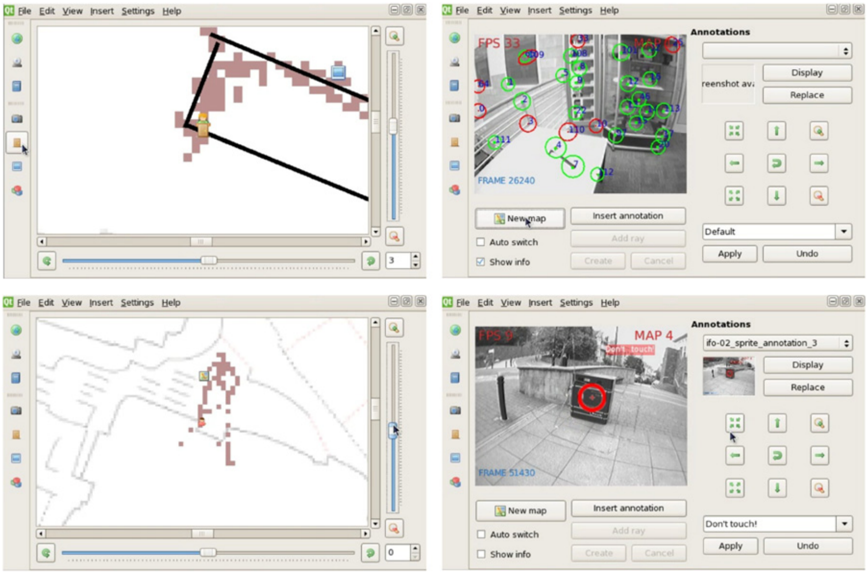

Fig. 9. Switching map contexts: (top-left) adding contextual information in the global map view; (top-right) creating a new SLAM map and annotation in the AR view; (bottom-left) navigating to an annotation marker using the global map; (bottom-right) viewing an annotation in the AR view after relocalisation.

the data collected by the other users. The workload of building and annotating the maps can also be shared between users, facilitating rapid mapping of the environment. Users who enter the scene are immediately provided with the latest version of the map and are able to view annotations that have been added by other users in their correct context. The OGC Sensor Observation Service (SOS) standard [37] provides a standardised interface for sharing observations throughout the system. This enables observations to be exposed to other devices and applications outside the system and simplifies the process of defining new sensors and technologies. The CCC acts as a central interface for external access to observations and also takes responsibility for controlling the flow of information around the system to enable it to adapt to changing network conditions.

4.4. Switching between map contexts

During normal operation, users are free to manually switch between the global map view and an AR view through the camera. The global map view provides situational awareness by showing the locations of all the tracked users and local maps alongside any additional contextual information that has been added to the map. The AR view through the camera displays the AR content associated with the local maps and a thumbnail screenshot taken when the AR content was created. Whenever the system successfully relocalises itself in a local map, the AR annotations are rendered over the camera image with full 6D pose relative to the tracked camera, allowing the user to view the content in real-time. When the system is lost, i.e. when no local maps are visible, the AR view displays the raw camera image and indicates that the system is attempting to relocalise.

When the user is in the global map view, the visual relocalisation is disabled to avoid unnecessary computation. In order to view a specific local map and its attached annotations, the user simply has to tap on the corresponding AR annotation icon in the global map. At this moment, the user is taken to the AR view, the specified local map is loaded and the system begins to attempt to relocalise in that map alone, i.e. single map relocalisation. If relocalisation is successful, then the local AR content becomes visible in the camera view. If the relocalisation is failing, then the user can return to the global map view and use it to navigate closer to the selected local map.

Alternatively, the user may also switch to the AR view without selecting a specific local map. In this case, the system attempts to relocalise in the full set of local maps, i.e. multiple map relocalisation with location gating. If any of the local maps is visible and relocalisation is successful, then that map's AR content is automatically loaded and becomes visible in the camera view. Similarly, if the user walks around the environment, then the system will switch between local maps automatically as it relocalises in different areas. When the user switches back to the global map view, the automatic relocalisation is disabled again to avoid unnecessary computation.

Fig. 9 provides an example of the switching that occurs between different views as a user builds a map and explores the annotations inside it.

5. Experiments

All of the system components have been built into a wearable backpack, which is shown in Fig. 10. Each hardware unit integrates components around a dual core Centrino laptop worn on the backpack. The interface with the user is displayed on a handheld touchscreen which has a firewire camera with a horizontal FOV of 80° rigidly attached to a 3D orientation sensor (which is not used in this work). The camera is calibrated in terms of focal length and radial distortion parameters. The touchscreen also has the UWB antenna attached to it, so that the most accurate sensors are close together. The GPS antenna is worn on the backpack's shoulder strap to enhance reception strength.

5.1. *Relocalisation performance*

We performed experiments to assess the performance of relocalisation in multiple maps. For this, we assume the worst case where no location gating is available. Experiments were conducted for an indoor scenario with 20 maps and an outdoor scenario with five maps, as shown in Figs. 11 and 12 respectively.

Although we ignore location gating for these experiments, the effectiveness of location gating in the final system will be dependent on two factors: the density of the local maps and the accuracy of the absolute positioning system that is being used. We assume that applications using the ToMAR system will typically require relatively sparse annotation of a large environment, hence a low density of local maps. We also assume that the availability of UWB, GPS and user input means that we are always able to achieve a global positioning accuracy of less than 20 m and that we can know which floor we are on inside a building. Under these assumptions, location gating should always yield a low bound on the number of local maps that need to be tested, so our choice of 20 maps indoors and five maps outside for this experiment seems reasonable.

The performance of the map ranking was evaluated using camera tracking and exhaustive single-map relocalisation to provide a ground-truth estimate of the correct map for each frame. This was matched against the multiple map relocalisation ranking computed at each frame and used to plot the cumulative distribution function of the ranking. The results of the ranking method were then compared against the baseline case of a randomised sort of the maps. The plots are shown on the left in Figs. 11 and 12.

Five different cell sizes for *QM* were tested. Although average performance was better than the baseline in all cases, the results showed that no single cell size gave good results for all maps. Sorting the maps by their mean rank over the five different cell sizes improved the average performance and reduced the number of individual maps that performed worse than the baseline. Alternative methods of combining the ranks from the different

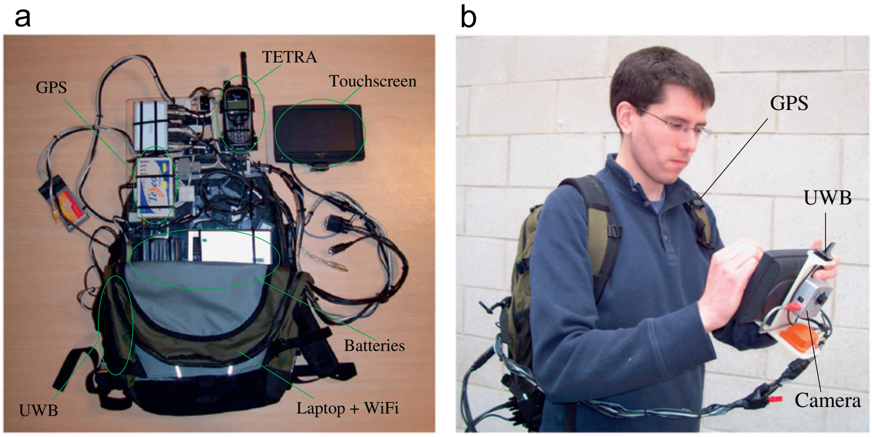

Fig. 10. Hardware components of the ToMAR system: (a) the contents of the backpack; (b) the system worn by a user showing interaction with touchscreen device.

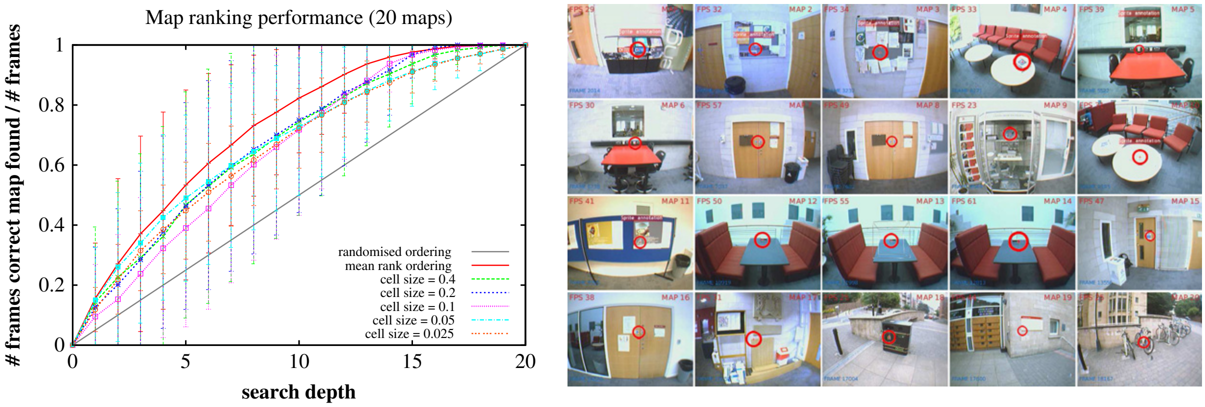

Fig. 11. Twenty maps were generated over a large indoor space incorporating many similar areas (several tables with red chairs). The cumulative distribution function of the ranking shows the improvement in performance achieved by the proposed method. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

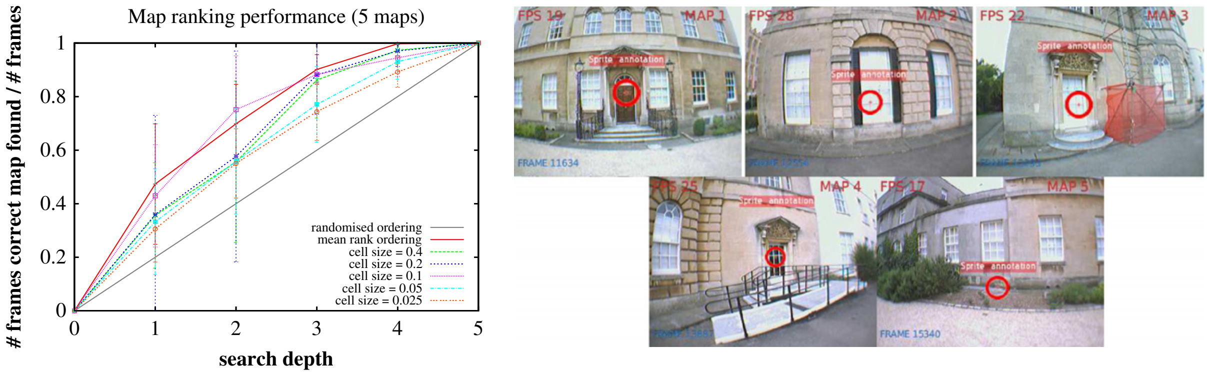

Fig. 12. Five maps were generated over a local outdoor area within a 10 m radius representative of GPS accuracy. The cumulative distribution function of the ranking shows the improvement in performance achieved by the proposed method.

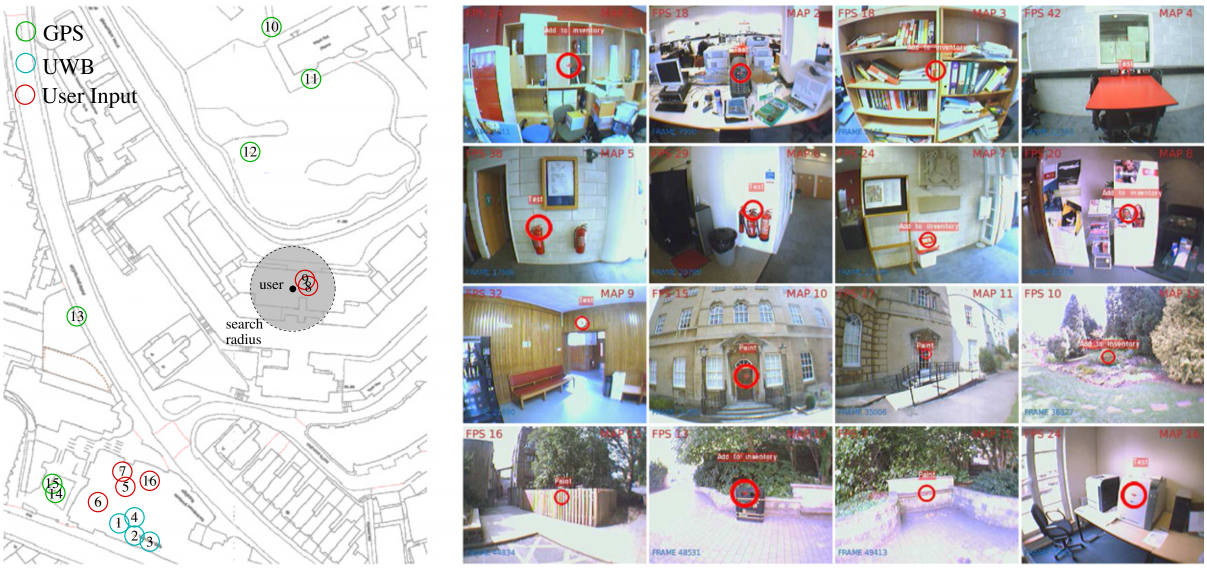

Fig. 13. An example of a global map generated by the ToMAR system. Sixteen local maps were created over an area of 0.1 km^2 containing a mixture of indoor and outdoor locations and with a mixture of GPS, UWB and User Input positioning. The 20 m search radius reflects that the user is currently in an area with no GPS or UWB coverage and is using interactive input positioning.

cell sizes, such as the median, minimum or maximum rank, were also considered but provided less performance improvement than the mean rank method.

In all cases, exhaustive relocalisation over all maps returned a false positive rate of 0%. This is despite the fact that the test sequences contain several instances of maps with very similar appearance. This supports the results by Chekhlov et al. [22] that show the single map relocalisation method producing very low false positive rates in real scenes. This low false positive rate means that we can confidently cut off our tests as soon as we find the first positive match. Therefore, the ranking procedure allows us to test fewer local maps, on average, per frame during relocalisation. In the worst case, when no match is found, we can either test the full set of maps, as normal, or set an upper bound on the depth to search in the ranking list. This choice of upper bound will be a trade-off between bounded computation and a higher overall false negative rate on the relocalisation, resulting from not testing all of the possible local maps. The graphs in Figs. 11 and 12 can be used to inform this decision. For example, testing only the top 50% of the ranked maps would successfully find the relocalisation in 80% of the frames in our two test cases.

The main limitation of the relocalisation approach occurs when the false negative rate of the single map relocalisation system becomes high, for example when lighting conditions have changed significantly or the camera is viewing the map from a novel viewpoint where the affine transformation is large enough to prevent feature descriptors from being successfully matched [25]. In these cases, the relocalisation will simply fail to return a match and the user will need to adjust their position inside the local map's viewing volume. In practice, we can assist this process by providing screenshots of the annotation to guide the initial alignment of the camera with the map.

5.2. Demonstration

To illustrate the wide area operation of the system we assessed performance over a 0.1 km^2 area containing a mixture of indoor and outdoor locations. The scenario mimics a maintenance task where users label multiple objects to be revisited by other users at a later time. In some indoor locations a UWB positioning system was available to provide absolute position.

A total of 16 local maps with associated annotations were built, as shown in Fig. 13, in a range of different natural and man-made environments. Map building took a couple of minutes on average per local map and the creation of annotations required just a few seconds to triangulate each 3D position. The main consideration when building the local maps was to initialise features covering all of the desired viewpoints of the map and ensure that their 3D position estimates were sufficiently well converged to give a good estimate of camera pose. This required a relatively smooth camera trajectory with plenty of translational motion for best results. The use of single map relocalisation meant that it was easy to recover and resume map building after any temporary losses of tracking caused by erratic motion or strong occlusions.

In areas with UWB coverage, a 2 m distance threshold was used and the separation of the constructed maps was such that a maximum of one candidate map was returned for relocalisation after location based gating. In one of the maps (map 3), the UWB accuracy was degraded by the surrounding furniture, producing position measurements outside the expected distance threshold and preventing automatic relocalisation. However, single map relocalisation was successful when the map was selected manually from the user interface.

Areas with GPS coverage used a 10 m distance threshold and returned a maximum of two candidate maps for relocalisation. In areas requiring interactive input to define absolute position, i.e. locations where both GPS and UWB were unavailable, a 20 m distance threshold was used and returned between two and six candidate maps. The multiple map relocalisation method found the correct map within the first two maps tested on each of the six occasions it was used.

The time required for relocalisation was very dependent on the visual complexity of the environment. In outdoor areas containing foliage, the number of visual features was comparatively high and single map relocalisation times could be as long as several hundred milliseconds. In these cases, it was sometimes necessary to hold the camera stationary for one or two seconds in order to find a successful relocalisation match. In less complex areas, relocalisation times were significantly shorter and it was often possible to relocalise from the continuously moving camera as it passed through the mapped area.

Our tests were carried out over the course of a single day and we have not performed any extensive evaluation of the system over longer periods of time. The system has some robustness to changes in the local structure of the environment because it uses a joint compatibility test to discard outliers during SLAM data association [38]. However, an additional complication during extended periods of operation may be the effect of varying lighting conditions, since the visual descriptors that we use are only partially invariant to illumination changes [24,20]. We expect that these changes would lead to an increased false negative rate in the relocalisation.

6. User evaluation

Wide area AR systems have received some attention in the literature, as discussed in Section 2. However, there have been relatively few attempts to assess this type of system in terms of its speed and accuracy for annotating a previously unknown environment. In this section, we present results from a user study evaluating the mapping and searching capabilities of our topometric mobile AR system (ToMAR) against a baseline set by a low-tech pen and paper (P+P) method.

It should be emphasised that the primary aim of this evaluation is not to identify one system as being better than the other but to place the performance of the ToMAR system in a meaningful context. In fact, we would expect the P+P system to generate good results in the trial, since the tools should be familiar to the users and provide a great degree of flexibility in creating the global map and adding contextual information. It should also be remembered that the ToMAR system is capable of providing advanced functionality that the P+P system cannot easily handle, such as viewing live AR content and sharing maps and data with multiple users. This advanced functionality is not assessed in this evaluation but provides a strong motivation for using the ToMAR system in real applications.

We have also constrained the trials to be run in an indoor location over a relatively small area. Our reason for doing this is that it provides us with full control over the test environment, which would not have been possible if we attempted to run the evaluation over a larger area or outdoors. Since we do not go outdoors, the GPS sensor was not used in this evaluation, but the UWB and visual SLAM components were both active and used for absolute positioning and local map building and annotation respectively.

6.1. Topometric mobile AR (ToMAR)

The ToMAR system used in the evaluation implements the design described in Section 3. Since we have calibrated the alignment of the UWB coordinate system with the global reference frame, the ToMAR system can display the global map data superimposed on a background map showing the building outline and surrounding streets. In this case, we use a low resolution rasterised version of Ordnance Survey MasterMap vector data [39]. Fig. 14 shows an example of the global map and AR annotations generated during the evaluation.

It should be noted that this evaluation only considers the single-user mode of the ToMAR system. In the full ToMAR system, annotations and maps can be shared between multiple users as they are created and updated. In addition, we disabled the ability to manually

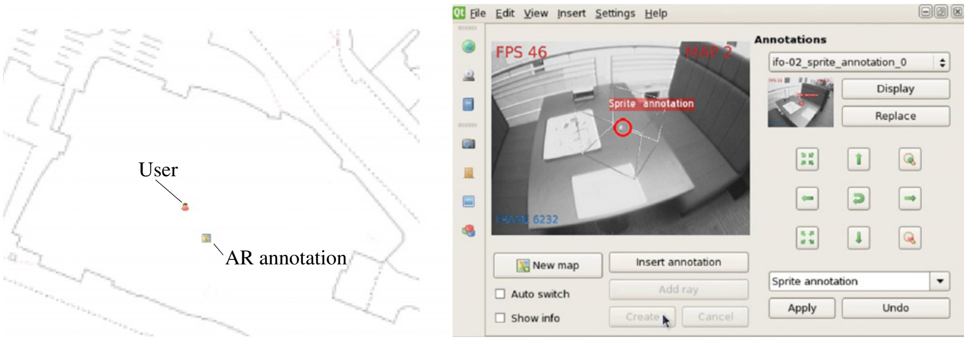

Fig. 14. Example of ToMAR maps: (left) global map showing the location of an annotated point of interest; (right) view of the annotation (red circle) via the touchscreen as it is tracked in the local map using visual SLAM. The GUI also displays a reference screenshot of the annotation which can be used to align the camera if visual SLAM initially fails to relocalise. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

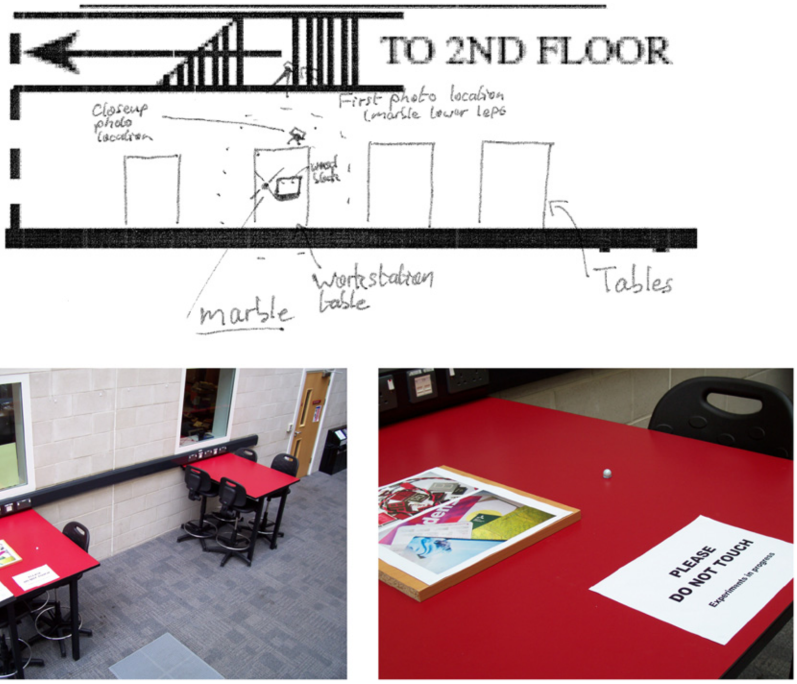

Fig. 15. Example of an annotated pen and paper (P+P) map containing contextual information (e.g. locations of tables along lower wall) and local annotations (e.g. position of marble on table) with the accompanying photographs taken by the user.

add contextual information to the global map, such as the locations of doors and furniture, in order to clearly focus on the capabilities of the UWB and visual SLAM components of the system.

6.2. *Pen and Paper (P+P)*

The P+P system was chosen to provide a low-tech baseline that the ToMAR system could be evaluated against. The system consisted of a paper map with a room plan of the building interior, a pen to add contextual information and local annotations to the paper map, and a camera to capture photographs of the local areas of interest associated with the annotations. Fig. 15 shows an example of a map created with this system during the trial. In broad terms, the annotated paper map corresponds to the ToMAR global map, and the photographs provide a locally accurate representation of the scene corresponding to the ToMAR local maps.

6.3. *Task description*

The two systems were set up in an indoor environment covering two floors of an open atrium area of a building, shown in Fig. 16, and containing 10 fixed tables (six on the upper floor and four on the lower floor). UWB base units were positioned to give good coverage over the whole environment and selected tables were set up with textured boards and marbles. The marbles acted as well-defined points of interest in the scene that could be used for annotation.

Using textured boards on visually featureless tables ensured that all of our annotated locations contained broadly equivalent sets of local visual features. This allowed us to make a meaningful comparison of annotation accuracy between the different locations. If we had used more natural locations with differing appearances, then this type of direct comparison would be meaningless due to the strong dependence between local accuracy and the set of visible features.

The trial itself was split into two stages in order to test the single-user mapping and searching capabilities of the systems independently.

6.3.1. *Mapping stage*

Four tables (two on the upper floor, two on the lower floor) were set up with a textured board and a marble. The marble was placed at a different location relative to the board on each table. Users were instructed to use each system (ToMAR and P+P) to create a map of two of the marble locations (one on each floor). They were told that the map should be detailed enough to enable another user to accurately place the marbles back in their correct locations if they were removed from the environment.

The completion time for the task was measured from the instant at which the ToMAR system was initialised, or the user was given the paper map and camera, until the instant when the user returned to the starting position of the trial and announced that they had finished mapping.

6.3.2. *Searching stage*

All 10 tables were set up with a textured board and a marble. The marble was placed at a different location relative to the board on each table and the local ground-truth location of each marble was recorded relative to one corner of each table. An expert user mapped two pairs of marble locations with each system (ToMAR and P+P) to generate four sets of maps. All of the marbles were then removed from the scene.

Users were instructed to use each system to place a pair of marbles back in the scene as accurately as possible using one of the sets of pre-generated maps created by the expert user. This provided an indirect way of measuring the accuracy of the registration of the AR annotation with the local map. Each marble was only handed to the user once they had located the table and announced that they were ready to place the marble.

The overall completion time for the task was measured, and also the individual times to position each marble and the local

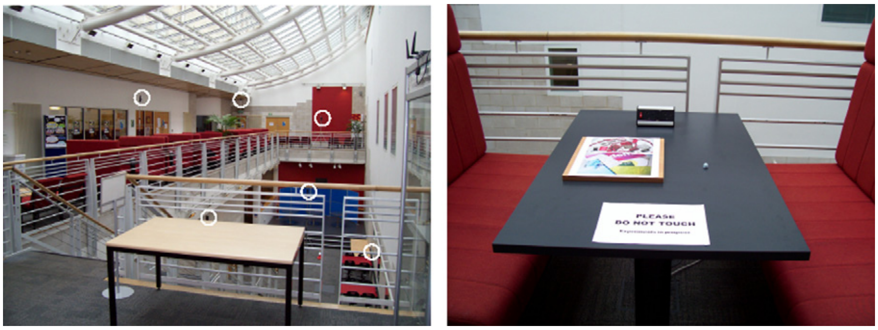

Fig. 16. Reference images of trial site: (left) locations of UWB base units are indicated by white circles; (right) marbles were placed in the scene as points of interest to be used in the mapping and searching tasks.

Table 1
The questions asked at the end of each stage of the trial. The questions relating to accuracy and search were tailored to the stage of the trial (questions on the left relate to the mapping stage, and those on the right relate to the searching stage).

Parameter	Question	
Effort	Which system required the least effort to use?	
Speed	Which system enabled you to complete the task in the fastest time?	
Performance	Which system was the least likely to lead to mistakes?	
Accuracy	Which system generated the most accurate map?	Which system allowed you to place the object most accurately?
Search	–	Which system allowed you to most easily find the general area to place the object?
P+P Rating	Overall, how well do you feel you completed the task using pen and paper?	
ToMAR Rating	Overall, how well do you feel you completed the task using the ToMAR system?	

accuracy of each marble's location. The overall completion time was the duration between the user being given their set of maps and the second marble being placed in the scene. The individual positioning times were defined from the instant at which the marble was given to the user until the instant when they finished placing the marble in the scene. The local accuracy of the marble position was calculated as the mean squared error with respect to the a priori ground-truth and measured manually using a tape measure.

6.4. Procedure and design

The experiment used 12 participants (two female and 10 male) between the ages of 20 and 35. The participants were researchers with a background in computer science or electrical engineering. Some had prior familiarity with the visual SLAM system, but none were familiar with the ToMAR system. All participants were tested individually and took part in both stages of the trial (mapping and searching). The mapping and searching stages of the trial were conducted on consecutive days and lasted approximately 30 and 15 min respectively.

We designed a within-subject experiment to evaluate the two different systems for each task. Over the course of the mapping phase, each user created one map with each system, each map containing two annotated marble locations. This gave a total of 12 trials measuring mapping time with each system. During the searching phase of the trial, each user positioned two pairs of marbles in the scene using a different system and pre-generated map for each pair. This gave a total of 12 trials measuring overall search time with each system, 24 trials measuring local positioning time with each system, and 24 trials measuring local positioning accuracy with each system.

Participants were briefed individually on the way to operate the two systems and given a practice trial to familiarise themselves with each task immediately prior to completing the real task. The order of presentation of the systems and the selection of tables for the mapping and searching tasks was counterbalanced between participants for learning and location effects. The participants were instructed to complete the tasks as quickly as possible.

At the end of each stage of the trial, the participants were asked to complete a brief questionnaire to provide feedback on their impressions of the two systems. Table 1 shows the questions that were asked to each participant for the mapping and searching phases of the trial. Participants responded by ticking one of five boxes, where the boxes defined a range of preference between the two systems. There was also space on the questionnaires for participants to add any additional comments which they wished to make.

6.5. Results

Fig. 17 shows the speed and accuracy results for the different systems for each task. The mean time to complete the mapping task was quicker with the ToMAR system (320 s) than with the P+P system (457 s) (two-sample t-test, $p=4\times10^{-6}$). However, the mean time to complete the searching task and position marbles back in the scene was quicker with the P+P system (115 s searching, 14 s positioning) than with the ToMAR system (242 s searching, 41 s positioning) (two-sample t-test, $p=2\times10^{-6}$ searching, $p=4\times10^{-12}$ positioning). The mean error in the locations of the objects added back into the scene was smaller with the ToMAR system (2.0 cm) than with the P+P system (3.1 cm) (two-sample t-test, $p=5\times10^{-11}$). The number of gross positioning errors, where the marble was placed in completely the wrong location, was the same for both systems (four out of 24 trials in each case), suggesting that users were similarly prone to making mistakes with either system.

Fig. 18 shows the questionnaire responses and user ratings for the different systems for each task. The rating responses indicate that the participants felt that they completed both tasks better using the ToMAR system (two-sample t-test, $p=7\times10^{-3}$ mapping, $p=0.04$ searching). Questionnaire responses indicate that the P+P system was perceived to be faster (one-sample t-test, $p=3\times10^{-3}$) and less effort to use (one-sample t-test, $p=0.01$) for the searching task, but that the ToMAR system was less likely to lead to mistakes

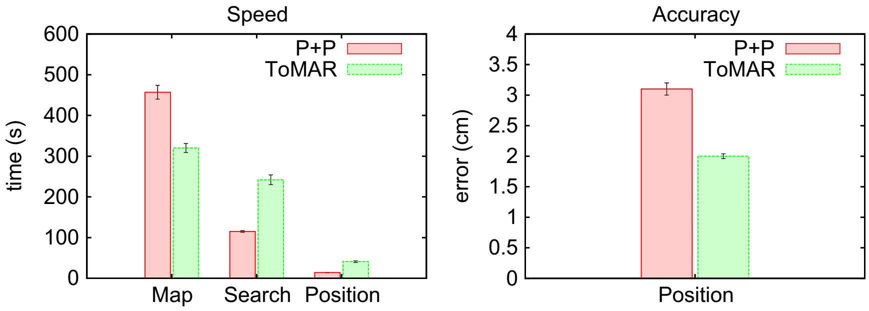

Fig. 17. Speed and accuracy results for the different systems. The timings for mapping and searching were taken over the full duration of the corresponding stage of the trial. The timings for positioning correspond to the time taken to position a single object during the searching stage of the trial. See the main text for more details.

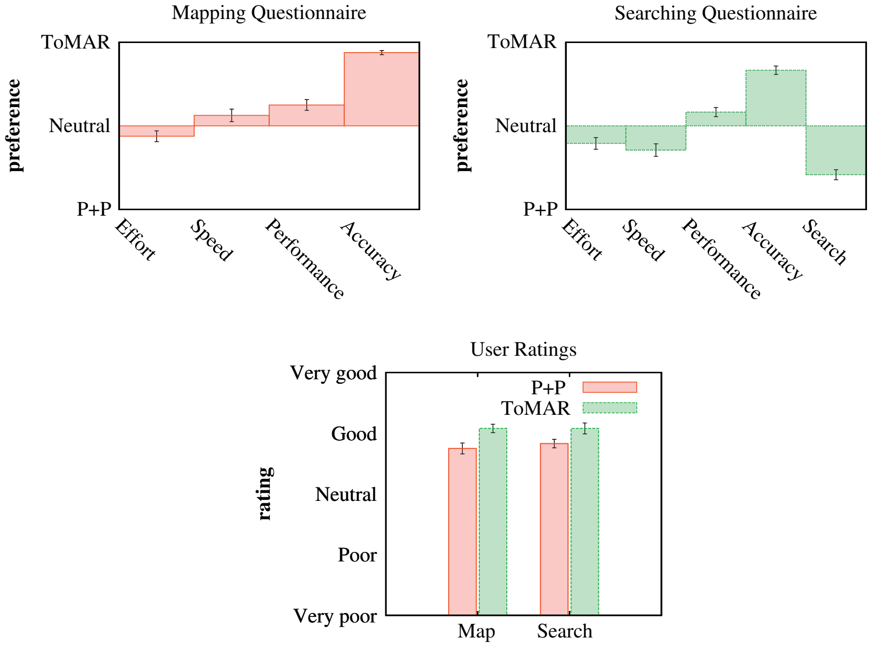

Fig. 18. Questionnaire responses for the mapping and searching trials. Refer to Table 1 for the corresponding questions. In each case, users were asked to express their preferences using a 5 point rating scale.

(one-sample t-test, $p=2\times10^{-3}$ mapping, $p=0.01$ searching) and provided the most accurate map (one-sample t-test, $p=2\times10^{-12}$) and object positioning (one-sample t-test, $p=3\times10^{-8}$). In contrast, the P+P system was perceived to be better for finding the general location of objects during the searching task (one-sample t-test, $p=8\times10^{-7}$). There was no perceived difference in effort (one-sample t-test, $p=0.09$) or speed (one-sample t-test, $p=0.10$) for the mapping task.

6.6. Discussion

In the mapping phase of the trial, the ToMAR system was faster than the P+P system. This was expected due to the constrained nature of the possible interactions with the ToMAR map. In contrast, in the P+P system, users were free to annotate their maps with as much contextual information as they wished, and often spent time creating detailed maps. This is supported by some of the comments from the trial questionnaires: “pen and paper requires more thought to determine the most useful set of annotations”, “pen and paper is straightforward but does take time to get it right”. In many applications, limiting the users creative freedom and ensuring a uniform appearance to the global maps and presentation of information could be seen as an advantage of a formalised system such as ToMAR.

In the searching tasks, the P+P system was faster than the ToMAR system. Although this result might be expected, given the greater flexibility and dexterity of a pen and paper interface over a touchscreen, we observed that the most significant factor during the trial was actually the lack of contextual information in the ToMAR global map. The trial environment was set up using identical tables which were separated by just 1–2 m from each other. This distance meant that the accuracy of the UWB positioning system was sometimes not sufficient to distinguish between two nearby locations and in these cases the participants had to visit both locations in turn and rely on the visual SLAM relocalisation to identify the correct one. In contrast, the additional contextual information present in the P+P maps provided an important cue to enable the participants to disambiguate the general location to place their marble and allowed them to navigate quickly and directly to the correct table. It also prevented the initial confusion that some users experienced when they went to the wrong table with the ToMAR system and the SLAM system (correctly) failed to relocalise. This observation was supported by some of the comments made by users on the trial questionnaires: “the pen and paper [map] had a clear distinction [between] confusing/ambiguous landmarks”, “very difficult to find the correct area [with ToMAR]”.

This observation demonstrates the benefits that contextual information provides for navigation and location tasks. We took the decision to use visually similar locations in the trial, which inevitably affected the ToMAR system more than P+P. The ToMAR system can easily be modified to allow users to add more contextual information to the global map, such as the location of tables, which we believe provides a more scalable approach to improving navigation performance than trying to increase the metric accuracy of absolute positioning sensors.

7. Conclusion

Topometric AR systems promise the dual benefits of being able to operate in wide areas whilst still providing the locally accurate mapping and tracking required for displaying AR annotations. In this paper we have presented a topometric mobile AR system that combines GPS and UWB positioning technologies with user interaction, providing a topological reference for navigation, and which uses local visual SLAM for accurate presentation of AR annotations in local areas of interest. The strength of the approach is in the scalability of the global topological map to cover wide areas and the robustness of the visual SLAM relocalisation which ensures that the system is always able to display local annotations in the correct location independently of the accuracy in the global map.

The efficiency of the relocalisation has been improved by the development of a method for the efficient ranking of visual maps. An evaluation of this method has demonstrated the system operating over various areas in a maintenance-like scenario where multiple users are able to find and label objects throughout the environment.

A user study has evaluated the basic mapping and searching capabilities of our topometric AR system against a low-tech pen and paper system and sets the stage for future evaluations of the more sophisticated capabilities. The pen and paper system provides a challenging baseline, since it is familiar to users and provides a great degree of flexibility in adding contextual information to the global map. The results of the evaluation suggest that the contextual information added to the pen and paper maps is more useful than the absolute positioning data from the UWB system in assisting a human to navigate between the local workspaces. This supports the idea of moving towards a topological representation of the global map which contains rich contextual information instead of attempting to improve the absolute accuracy of the global map. Given the additional scope for extending the system to multi-user applications and more complex tasks, we believe that the topometric mobile AR approach we present has considerable potential.

Acknowledgments

This work was funded by the UK Technology Strategy Board and the UK Engineering and Physical Sciences Research Council. The authors wish to thank all partners in the ViewNet project for their discussions and participation in this work. We are grateful to Sriram Subramanian and the Bristol Interaction and Graphics Group for assistance in the user evaluation pilot study, and Michael Collett and Tian Loh of the UK National Physical Laboratory for assistance collecting the user evaluation data. Ordinance Survey mapping © Crown copyright.

References

[1] Azuma R. The challenge of making augmented reality work outdoors. In: Mixed reality: merging real and virtual; 1999. p. 379–90.

[2] Welch G, Foxlin E. Motion tracking: no silver bullet, but a respectable arsenal. IEEE Computer Graphics and Applications 2002;22(6):24–38.

[3] Rekimoto J, Ayatsuka Y, Hayashi K. Augment-able reality: situated communication through physical and digital spaces. In: International symposium on wearable computers; 1998. p. 68–75.

[4] Thomas B, Close B, Donoghue J, Squires J. ARQuake: an outdoor/indoor augmented reality first person application. In: International symposium on wearable computers; 2000. p. 139–46.

[5] Reitmayr G, Schmalstieg D. Location based applications for mobile augmented reality. In: Australasian user interface conference; 2003. p. 65–73.

[6] Reitmayr G, Schmalstieg D. Collaborative augmented reality for outdoor navigation and information browsing. In: Symposium on location based services and telecartography; 2004. p. 31–41.

[7] Reitmayr G, Drummond T. Going out: robust model-based tracking for outdoor augmented reality. In: International symposium on mixed and augmented reality (ISMAR); 2006. p. 109–18.

[8] Löchtefeld M, Gehring S, Schöning J, Krüger A. PINwI—pedestrian indoor navigation without infrastructure. In: Nordic conference on human computer interaction; 2010. p. 731–4.

[9] Baillot Y, Brown D, Julier S. Authoring of physical models using mobile computers. In: International symposium on wearable computers; 2001. p. 39–46.

[10] Höllerer T, Wither J, Diverdi S. Anywhere augmentation: towards mobile augmented reality in unprepared environments. In: Location Based Services and TeleCartography; 2007. p. 393–416.

[11] Hallaway D, Feiner S, Höllerer T. Bridging the gaps: hybrid tracking for adaptive mobile augmented reality. Applied Artificial Intelligence 2004;25: 477–500.

[12] Kourogi M, Sakata N, Okuma T, Kurata T. Indoor/outdoor pedestrian navigation with an embedded GPS/RFID/self-contained sensor system. In: International conference on artificial reality and telexistence; 2006. p. 1310–21.

[13] Castle R, Klein G, Murray D. Video-rate localization in multiple maps for wearable augmented reality. In: IEEE International symposium on wearable computers (ISWC); 2008. p. 15–22.

[14] Fritz G, Seifert C, Paletta L. A mobile vision system for urban object detection with informative local descriptors. In: IEEE International conference on computer vision systems; 2006. p. 30–9.

[15] Schleicher D, Bergasa L, Ocaña M, Barea R, López E. Real-time hierarchical GPS aided visual SLAM on urban environments. In: IEEE International conference on robotics and automation; 2009. p. 4381–6.

[16] Eade E, Drummond T. Unified loop closing and recovery for real time monocular SLAM. In: British machine vision conference (BMVC); 2008. p. 1–10.

[17] Kjaergaard M, Blunck H, Godsk T, Toftkjaer T, Christensen D, Gronbaek K. Indoor positioning using GPS revisited. In: International conference on pervasive computing; 2010. p. 38–56.

[18] Harmer D, Russell M, Frazer E, Bauge T, Ingram S, Schmidt N, et al. EUROPCOM: emergency ultrawideband radio for positioning and communications. In: IEEE conference on ultra-wideband; 2008. p. 85–8.

[19] Davison A, Mayol W, Murray D. Real-time localisation and mapping with wearable active vision. In: IEEE/ACM international symposium on mixed and augmented reality (ISMAR); 2003. p. 18–28.

[20] Chekhlov D, Pupilli M, Mayol-Cuevas W, Calway A. Real-time and robust monocular SLAM using predictive multi-resolution descriptors. In: International symposium on visual computing; 2006. p. 276–85.

[21] Klein G, Murray D. Parallel tracking and mapping for small AR workspaces. In: IEEE/ACM international symposium on mixed and augmented reality (ISMAR); 2007. p. 1–10.

[22] Chekhlov D, Mayol-Cuevas W, Calway A. Appearance based indexing for relocalisation in real-time visual SLAM. In: British machine vision conference (BMVC); 2008. p. 363–72.

[23] Bar-Shalom Y, Li X, Kirubarajan T. Estimation with applications to tracking and navigation. John Wiley and Sons; 2001. ISBN: 9780471416555.

[24] Lowe D. Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision 2004;60(2):91–110.

[25] Mikolajczyk K, Schmid C. A performance evaluation of local descriptors. IEEE Transactions on Pattern Analysis and Machine Intelligence 2005;27(10): 1615–30.

[26] Civera J, Davison A, Montiel J. Inverse depth parametrization for monocular SLAM. IEEE Transactions on Robotics 2008;24(5):932–45.

[27] Rosten E, Drummond T. Machine learning for high-speed corner detection. In: Proceedings of the European conference on computer vision (ECCV); 2006. p. 430–43.

[28] Shi J, Tomasi C. Good features to track. In: IEEE international conference on computer vision and pattern recognition (CVPR); 1994. p. 593–600.

[29] Williams B, Klein G, Reid I. Real-time SLAM relocalisation. In: IEEE international conference on computer vision (ICCV); 2007. p. 1–8.

[30] Sivic J, Zisserman A. Video google: a text retrieval approach to object matching in videos. In: International conference on computer vision (ICCV); 2003. p. 1470–8.

[31] Cummins M, Newman P. FAB-MAP: probabilistic localization and mapping in the space of appearance. International Journal of Robotics Research 2008;27(6):647–65.

[32] Brown M, Szeliski R, Winder S. Multi-image matching using multi-scale oriented patches. In: Proceedings of the IEEE international conference on computer vision and pattern recognition (CVPR); 2005. p. 510–7.

[33] Bunnun P, Damen D, Subramanian S, Mayol-Cuevas W. Interactive image-based model building for handheld devices. In: ISMAR workshop on augmented reality super models; 2010. p. 1–4.

[34] Hartley R, Zisserman A. Multiple view geometry in computer vision. 2nd ed. Cambridge University Press; 2004. ISBN: 0521540518.

[35] Piekarski W, Thomas B. Augmented reality working planes: a foundation for action and construction at a distance. In: Proceedings of the international symposium on mixed and augmented reality (ISMAR); 2004. p. 162–71.

[36] van den Hengel A, Hill R, Ward B, Dick A. In situ image-based modeling. In: Proceedings of the international symposium on mixed and augmented reality (ISMAR); 2009. p. 107–10.

[37] Na A, Priest M. OpenGIS sensor observation service. Technical Report OGC 06-009r6, Open Geospatial Consortium, Inc.; 2007.

[38] Neira J, Tardos J. Data association in stochastic mapping using the joint compatibility test. IEEE Transactions on Robotics and Automation 2001;17(6):890–7.

[39] 〈http://www.ordnancesurvey.co.uk/oswebsite/〉, 2011.

[40] Gee AP, Escamilla-Ambrosio PJ, Webb M, Mayol-Cuevas W, Calway A. Augmented crime scenes: virtual annotation of physical environments for forensic investigation. ACM international workshop on multimedia in forensics, security and intelligence, 2010. p. 105–10.

[41] Gee AP, Calway A, Mayol-Cuevas W. Visual mapping and multi-modal localisation for anywhere {AR} authoring. ACCV international workshop on application of computer vision for mixed and augmented reality, 2010. p. 31–41.