17 Applications of Augmented Reality for the Automotive Industry

Vincent Gay-Bellile, Steve Bourgeois, Dorra Larnaout, and Mohamed Tamaazousti

CONTENTS

17.1 Introduction .................................................................................................................................. 434
17.2 Potential Benefits of Augmented Reality for the Automotive Industry ........................................ 434
  17.2.1 Vehicle Design and Conception ................................................................................................. 434
  17.2.2 Factory Planning ......................................................................................................................... 435
  17.2.3 Vehicle Production ...................................................................................................................... 435
  17.2.4 Sales Support ............................................................................................................................ 436
  17.2.5 Driving Assistance ...................................................................................................................... 436
  17.2.6 User Manual and Maintenance Support ..................................................................................... 438
17.3 Technological Challenges for a Large-Scale Deployment ................................................................ 438
17.4 Tracking a Vehicle or One of Its Component ................................................................................. 440
  17.4.1 State of the Art ........................................................................................................................... 440
  17.4.2 Our Solution: VSLAM Constrained by a CAD Model ................................................................. 443
    17.4.2.1 Principle ............................................................................................................................. 443
    17.4.2.2 Implementation .................................................................................................................... 444
  17.4.3 Discussion .................................................................................................................................. 445
17.5 Vehicle Localization for Aided Navigation in an Urban Context ..................................................... 446
  17.5.1 State of the Art ........................................................................................................................... 446
  17.5.2 Constrained VSLAM for Large-Scale Vehicle Localization ....................................................... 447
    17.5.2.1 Constraining VSLAM with GPS .......................................................................................... 448
    17.5.2.2 Improving the In-Plane Accuracy with GIS Constraint ....................................................... 449
    17.5.2.3 Improving the Out-Plane Accuracy with a SIG Constraint ................................................ 452
    17.5.2.4 Overview of Our Complete Solution for Vehicle Localization ........................................... 452
  17.5.3 Discussion .................................................................................................................................. 453
17.6 Conclusion ....................................................................................................................................... 454
Acknowledgments .................................................................................................................................... 454
References .............................................................................................................................................. 455
17.1 INTRODUCTION

For a number of reasons, the automotive industry is heavily involved in the development of augmented reality (AR) and its applications. Since the late 1990s, car manufacturers and original equipment manufacturers (OEMs) have explored the benefits of AR through major collaborative projects, such as ARVIKA with Daimler Chrysler, Ford, Audi, Volkswagen, and Siemens AG (ARVIKA Project 1999–2003); ARTESAS with Siemens AG and BMW (ARTESAS Project 2004–2006); EFA2014 with Audi, BMW, Continental, and Bosch (EFA2014 Project n.d.); AVILUS with Daimler, Siemens, and Volkswagen; and finally, EGYPT with PSA. During these projects, the use of AR for the whole product life cycle has been studied, from car conception until customer assistance services. Still, the deployment of AR applications in the automotive industry has remained limited due to various technical challenges.

In this chapter, we first provide an overview of the different uses of AR in the automotive industry and their relative requirements. Then, we focus on recent tracking technology advances that may lead to large-scale deployment of AR solutions in the automotive industry.

17.2 POTENTIAL BENEFITS OF AUGMENTED REALITY FOR THE AUTOMOTIVE INDUSTRY

In the following sections, we briefly illustrate the potential benefits of AR during the different stages of the life cycle of a vehicle. However, the resulting list of applications are not meant to be exhaustive but representative of the uses of AR in the automotive industry.

17.2.1 VEHICLE DESIGN AND CONCEPTION

Design and ergonomics of a vehicle are the key elements in the purchase decision of many customers. Therefore, in the vehicle conception phase, engineers must carefully consider customer assessments of various design features as early as possible. However, due to the multiple variations of a vehicle (equipment, material, etc.), production of a real-size physical mock-up for each variant suggested by a purchaser would be both expensive and slow. In this context, virtual reality (VR) provides an efficient solution to allow a potential buyer to observe many variants of a vehicle using the same display device (e.g., VR glasses, CAVE). However, compared to the use of VR, in many steps of the design process, physical mock-ups are still being used since they provide an accurate evaluation of the human factor, which includes the ability to touch the real object and to better visualize the object’s shape and dimensions.

Compared to VR as a visualization tool, AR offers additional benefits as it allows virtual elements of an automobile to be visualized while also allowing the end user to remain in his or her natural environment (Menk et al. 2010; Porter et al. 2010). Indeed, a mock-up of the target object can be virtually customized in real time through an AR device (video stream of a camera, AR glasses or projection onto
the mock-up). Further, because it uses a tangible prototype, AR (compared to VR) provides a better perception of the dimensions and the volume of an automobile. The use of AR also allows the potential customer to move naturally around the object and even interact with it. Finally, since the fine design elements of an automobile (e.g., buttons, handles, prints) can be added virtually to the real object, with AR only a low-fidelity mock-up is required. Compared to the traditional mock-up approach, the overall cost of production using AR is reduced since one model can be used to assess many design configurations.

### 17.2.2 FACTORY PLANNING

To remain competitive, that is, maintaining reduced production costs and adhering to consumer taste evolution, the automotive industry needs to have a quick renewal pace for its vehicles. This shorter product life cycle implies that the manufacturer must upgrade its installation frequently. However, suspending a production chain to check if the validity of a new process is expensive since it reduces the productivity of the current production chain. Maintaining a high availability of the production while providing flexibility is the target of factory planning. Therefore, virtual simulation is widely used in factory planning since it allows engineers to simulate the entire production system, detect discrepancies (e.g., collision between a robotic arm and its environment), and evaluate the impact of the new workspace without disturbing the current production chain. However, this solution requires a complete and accurate 3D model of the existing environment, which is difficult to obtain for a plant in operation.

In the factory planning context, AR can provide an in situ visualization of the virtual simulation (Pentenrieder et al. 2007; Park et al. 2008) where the operator is invited to move through the plant while observing the virtual simulation integrated into the real environment. The consistency of the simulation with respect to the real environment can thus be visually assessed.

### 17.2.3 VEHICLE PRODUCTION

Although the production lines are increasingly automated, tasks that require complex manipulations, or whose accessibility is limited, are still performed by human operators. Also, when a vehicle is produced in limited quantity, such as the case for customized utility vehicles, the use of robots in the production cycle can be prohibitively expensive. In these cases, automobile assembly is achieved by a human technician following a specific order and with dedicated tools. Because the use of operation documents is time consuming, operators tend to rely on their memory instead of visually checking the step order. This results in a higher error rate, and a loss of both quality and time. AR offers an alternative to paper documentation. By directly superimposing the instruction sequence into the operator’s field of view, AR can provide the required information at the right time, at the right place, without requiring any significant mental effort from the operator. Moreover, the information provided by AR is not limited to just the description of what to do, but also how to do it: the tool to use and the correct physical gesture. AR can also improve the perception the
operator forms of the design by showing him or her hidden elements that he or she otherwise could not directly see, such as the position of an electrical wire behind a metal panel. This virtual information can be generated from existing CAD data of the product, or also from data arising from nondestructive testing (e.g., ultrasonic imagery, tomography).

Summarizing, AR can reduce the risk of human mistakes, but also increase the efficiency of the operator. Indeed, with AR the operator can remain focused on an important area of intervention since he or she no longer needs to switch between the documentation and its work area. This approach, the use of AR within the automotive industry, was evaluated for tasks such as picking (Reif and Günthner 2009), assembly (Reiners et al. 1998; Regenbrecht et al. 2005), and quality control (Zhou et al. 2012). In addition, if AR technology can be used to make complex operations easier, it can also be used for training purposes, in which case AR could free senior staff from basic training duties or accelerate the replacement of an absent worker.

### 17.2.4 Sales Support

During the process of a sale, an automotive salesperson faces various challenges. First, he or she may have to present to the potential customer the various models and options of a vehicle; yet only one or two models may be available in the showroom. Moreover, even if the vehicle is available, some specificity of the vehicle cannot be easily observed, such as the turning radius or the braking distance. One current solution to this problem is to use a VR configurator. However, this approach has problems since the virtual representation is limited given that the user perception of dimensions and volumes of the automobile are distorted.

AR provides an interesting alternative to this problem. Similar to a virtual configurator, with AR the vehicle can be customized interactively (Gay-Bellile et al. 2012). Importantly, with this approach, since the augmentation is achieved on a real vehicle, the perception of dimensions and volumes are preserved. With AR the end user experience can be further improved by integrating some interactions between the real environment and the virtual elements. For example, if the customer chooses to change the color of the bodywork, it is important to provide a restitution that respects the current lighting conditions, but that also keeps the reflections of the surrounding environment onto the bodywork. And finally as illustrated in Figure 17.1, if customization through AR is useful for cars purchased by the average customer, it is even more useful for utility vehicles due to their extreme customizable interior furnishings.

### 17.2.5 Driving Assistance

In driving conditions, perception of the surrounding environment is a key element of the driving task. AR can be used to improve this perception in various ways. First, AR can superimpose, with respect to the driver’s field of view, his or her current trajectory in order to support him or her during a maneuver. AR can also suggest a trajectory to help the driver during his or her navigation decisions (see Figure 17.2). In this latter case, AR can reduce the cognitive effort of the driver, since, contrarily
FIGURE 17.1  AR for sales assistance in an automotive showroom: while constrained VSLAM aligns accurately the CAD model of the vehicle with the video stream (top left), AR is used to customize the interior furnishings (bottom left) of the vehicle. The whole processing is done in real time on a mobile tablet (bottom right). (Courtesy of www.diotasoft.com.)

FIGURE 17.2  Driving-assistance in AR. While the vehicle is accurately localized with a constrained VSLAM using a camera, standard GPS, and a coarse 3D city mode (top left), AR is used to display the path to follow (top right), safety information such as dangerous intersections (top right) and crosswalks (bottom left), or touristic information such as hotel (bottom right). (Courtesy of CEATECH.)
to current aided navigation systems, the driver would not have to mentally transpose schematic representation of a road network with his/her current perception of the environment. AR can also be used to focus the driver’s attention onto potential dangers, such as at crosswalks and intersections (see Figure 17.2), and road users.

AR can also improve the driver’s perception of the roadway environment when observation conditions are degraded due to weather or other reasons. For example, with AR, the edges of the road can be highlighted in foggy conditions; and vehicles occluded by a building (Barnum et al. 2009) or by another vehicle (Gomes et al. 2013) can be displayed to the driver using AR. Finally, the development of an autonomous vehicle will probably favor the development of infotainment applications, such as touristic guide, and AR will be especially useful in providing this information.

17.2.6 User Manual and Maintenance Support

AR also provides an interesting alternative to the paper-based user and technical manual that is provided with a vehicle. Indeed, in such manuals, even if the instructions are provided through schematics or pictures, the transcription of these instructions onto the real vehicle is not necessarily intuitive. By providing these instructions directly onto the real vehicle (Friedrich 2002; Gay-Bellile et al. 2012), AR avoids confusions (see Figure 17.3). Moreover, if the user does not look at the area of the vehicle needing his or her attention, AR can guide him or her to the correct location. And finally, if the manual is not sufficient, a remote expert can provide dedicated instruction by annotating, online, the live video stream of an AR device (Reitmayr et al. 2007).

17.3 Technological Challenges for a Large-Scale Deployment

While studies have demonstrated the potential benefits of AR in the automotive field, most of the proposed solutions remained at the level of proof of concept. To advance another step further, it is necessary to develop prototypes that demonstrate the ability of AR technology to reach the level of requirements as follows:

- **Service quality**: The virtual elements of an AR system should provide clear and unambiguous information, and accurate and stable registration of these elements onto the real environment; otherwise the system would lead to errors or would require a significant cognitive effort for the end user to compensate for these inaccuracies.

- **Service continuity**: The AR system should provide a high availability ratio since it engages the productivity or the brand image of the car manufacturer. The robustness to variations of its environmental conditions of use is a main concern.

- **Ease of use and ergonomics**: The AR system should be intuitive enough for its use by the targeted end user, who, realistically, is not expected to be an expert in AR technology. Moreover, the AR system should be compatible
Applications of Augmented Reality for the Automotive Industry

with the task it is required to support. For example, for assembly-assistance or driving-assistance applications, the requirements of the task require a hands-free device.

• **Ease of deployment:** The cost for establishing and operating the system should be compatible with the benefits provided by its use. For a large-scale deployment, the equipment and data required for its exploitation must be available at low cost.

To reach these objectives, the main challenge for developing AR for the automobile industry is localization technology. Indeed, as we will underline in the following sections, most of the current solutions rely on a tracking system that cannot be deployed at large scale due to their lack of accuracy, lack of robustness, or their prohibitive cost.
Of course, localization is not the only challenge to implementing AR in the automotive industry. Providing a hands-free AR device remains a key challenge for the solution’s ergonomics, as illustrated by the research activity on semitransparent glasses and windshields. However, depending on the targeted application, the use of a deported screen (such as the screen of a tablet or the screen of Google Glasses) or of a video projector (spatial AR) can provide an imperfect but nevertheless acceptable alternative.

Consequently, in the following sections, we will focus our attention to the localization issue and introduce the recent advances we have made on this subject. Also, we will distinguish tracking solutions with respect to their applicative context: AR on the vehicle and AR from the vehicle. The former covers all the applications for which the car, or one of its components, is the target of the augmentation, while the latter covers mainly the driving-assistance applications.

### 17.4 TRACKING A VEHICLE OR ONE OF ITS COMPONENT

The automotive industry remains a challenging context for current tracking technologies since it requires an accurate localization of potentially complex objects by a nonexpert user. In the following, we will first introduce the current state-of-the-art solutions and their relative limitations with respect to the requirements outlined in Section 17.3. Then, we will introduce our tracking solution and discuss its benefits in the automotive context.

#### 17.4.1 State of the Art

Among the automotive industry, various tracking solutions have been attempted. A first solution consists of exploiting the 6DoF mechanical measurement arms that are already in use in the automotive industry (Kahn and Kuijper 2012). As long as the arm is required for the targeted task, this solution provides extremely accurate and robust localization. However, in most of the applicative contexts, such an approach does not provide an efficient solution. Indeed, the volume covered by such an arm is limited by nature and its use induces a limitation in terms of accessibility (collision between an arm segment and an element of the environment). Moreover, its cost represents a prohibitive drawback for its deployment.

A second approach relies on optical tracking solutions that use 2D (Calvet et al. 2012) or 3D markers (Pintaric and Kaufmann 2007). These solutions consist of rigidly attaching a set of markers onto the target object (inside-out approach) or onto the observer (outside-in approach). Within the inside-out approach, the pose of the object with respect to the camera is directly estimated from an image of the marker in the camera’s video stream. Since at least one marker must be visible to provide localization, the volume covered by this approach is limited by the number of markers. However, even attaching a single marker onto the target object can be unacceptable in some contexts. For example, selling-support applications require that the appearance of the car remains unchanged. In such contexts, the outside-in approach provides an alternative solution. With this solution, the marker is no longer attached to the target object, but onto the observer. A constellation of static cameras is used to
localize the observer with respect to the object. For that, an off-line calibration process has to be performed beforehand to determine the pose of the target object with respect to the camera constellation. Similar to the inside-out approach, the outside-in approach requires the marker to be visible by at least one camera of the constellation. Therefore, the volume covered by such an approach is limited. In both cases, the necessity of instrumenting the environment with markers or cameras implies a deployment process whose complexity is inappropriate with most of the applicative contexts. For the outside-in approach, the cost of the camera constellation can also be a further limitation.

To prevent these deployment problems, optical solutions that rely exclusively on natural features of the target object have been developed. They are referred to as model-based tracking and rely on a 3D model that describes the natural features of the target object (assumption of known object). These 3D features can be geometrical, such as edges of a 3D mesh model (Drummond and Cipolla 2002; Wuest et al. 2007), or photo-geometric, such as a collection of 3D texture patches (Rothganger et al. 2003; Gordon and Lowe 2006). The pose $P$ of the object can be estimated from a camera image by matching a subset $\{Q_i\}_{i=1}^n$ of these 3D model features with their corresponding 2D features $\{q_{i,j}\}_{i=1}^n$ in the $j$th camera image. The object pose is then defined as the one that minimizes the reprojection error, that is, the distance between the 2D projection of the 3D feature according to the object pose and the camera’s intrinsic parameters and the position of their associated 2D features in the image:

$$\arg\min_P \sum_{i=1}^n d^2(q_{i,j}, \pi(PQ_i))$$

where
- $d^2$ is a 2D Euclidian distance
- $\pi$ is the camera projection function

While model-based tracking solutions are easy to deploy and result in accurate localization, their lack of robustness is not compatible with the quality and continuity of service required.

Indeed, model-based tracking is subject to two major drawbacks. First, the matching of the 3D features with 2D observations can only be achieved under restrictive conditions of use. On the one hand, the information encoded in geometric features is not discriminant enough to easily distinguish between two observations. To succeed, a usual solution consists in introducing a small motion assumption. However, this assumption reduces the condition of use of the system. On the other hand, photometric features are discriminant enough to achieve the matching under fast motion. However, because the appearance of an object varies with the lighting conditions, a 3D model based on photometric features is only valid under specific lighting conditions. This constraint reduces the use of this solution to an environment where the lighting conditions are controllable.
Second, the pose of the camera is not accurately estimated when the object is small in the image, subject to a large occlusion, or out of the field of view, since those configurations do not provide enough geometric constraints: the number of 2D features matched with the 3D model is small and/or these 2D features are located in a small area of the image.

Among the optical tracking solutions, a last approach provides a greater robustness to lighting condition variations and large viewpoint variations. This solution is usually referred as visual simultaneous localization and mapping (VSLAM) or structure from motion (SfM), depending on the particular scientific community (Mouragnon et al. 2006; Klein and Murray 2007).

While the camera pose is estimated in a similar way to a model-based tracking solution, VSLAM approaches do not use an a priori model of an object but reconstructs, online and in real time, a 3D model of the whole scene (assumption of unknown environment). To achieve this reconstruction, VSLAM uses the principle of multiview geometry (Hartley and Zisserman 2004) in order to assess the 3D position of scene features, such as 3D points, from the motion of their apparent 2D motion in the video stream. Since a long-enough 2D displacement is required to estimate the depths of the features, this reconstruction process is not achieved in each frame. The frames at which the reconstruction process is achieved are usually referred to as keyframes. To reach an optimal trajectory and scene reconstruction with respect to the multiview geometry constraints, both of them are optimized simultaneously with a nonlinear optimization process, referred to as bundle adjustment (BA). This optimization process minimizes the error $\varepsilon_R$ that corresponds to the sum of square differences between the 2D projection of each 3D point and its associated 2D observations in each keyframe (also referred to as reprojection error):

$$
\varepsilon_R(\theta) = \sum_{i=1}^{n} \sum_{j \in A_i} d^2(q_{i,j}, \pi(P_j, Q_i))
$$

where
- $\theta$ stands for the scene parameters optimized by the BA (i.e., the coordinates $\{Q_i\}_{i=1}^{n}$ of the reconstructed 3D features and the pose parameters $\{P_j\}_{j=1}^{1}$ at keyframes)
- $d^2$ is the squared Euclidean distance
- $\pi$ is the projection function of a point $Q_i$ with respect to a camera pose $P$
- $A_i$ is the set of keyframe indexes in which the point $Q_i$ is associated to an observation $q_{i,j}$

Since the camera localization is estimated from a 3D model that covers the whole observed scene, the VSLAM approach provides an excellent robustness to a large and fast camera motion and partial scene occlusion. Moreover, since the 3D model of the scene is built online, the photometric appearance of the model features is, by construction, consistent with the current lighting conditions. Unfortunately, the VSLAM approach provides poor localization accuracy. Indeed, in spite of the BA, the reconstruction process is usually subject to error accumulation due to its incremental nature. Moreover, VSLAM localization is expressed in an arbitrarily chosen coordinate
Applications of Augmented Reality for the Automotive Industry

frame, and an arbitrary scale factor when a single camera is used. However, this last drawback can be partially resolved by bootstrapping the 3D model of the scene with the help of a model-based tracking approach. While the model-based solution is used to estimate the pose of the first keyframe, an initial reconstruction of the scene is achieved by back-projecting the 2D image features onto the CAD model of the target object (Bleser et al. 2006). If this initial reconstruction process provides a coordinate frame and a scale factor to the VSLAM localization, their accuracy remains limited since the reconstruction is achieved from a single point of view.

In the following, we will introduce our optical tracking approach that goes one step further toward the combination of model-based tracking and VSLAM, and provides an accurate and robust localization while remaining easy to deploy.

17.4.2 Our Solution: VSLAM Constrained by a CAD Model

While the VSLAM approach uses the whole scene to estimate the 3D camera motion, it assumes that no a priori knowledge on this environment is available (assumption of unknown environment) and that model-based approaches use exclusively the elements of the scene for which an a priori model is available (assumption of known object). However, in most of the automotive applications, the scene is constituted of a known object (the car or one of its components for which a CAD model is available) surrounded by an unknown environment (assumption of partially known environment). In this context, it would be unfortunate to neglect the information provided by the unknown environment or by the CAD model. To the benefit of both of these situations (unknown environment, CAD model), we proposed to merge the VSLAM solution and the model-based solution in a new approach referred as constrained VSLAM. More specifically, since car components are usually textureless, we will consider a model-based approach based on purely a geometric model.

To develop a tracking solution that is accurate, robust, and easy to deploy with this new approach, we consider the following assumptions:

• A CAD model of the target object is available.
• The target object remains static with respect to its surrounding environment during the tracking process.

We propose that both of these assumptions are reasonable in most automotive applications. Indeed, CAD models are widely available in the automotive industry and it is unusual to move a car during a sale, for quality control, or a maintenance process.

17.4.2.1 Principle

Similar to Bleser (Bleser et al. 2006), a model-based approach based on a CAD model is used to bootstrap a VSLAM process. However, while Bleser ignores the CAD model during the remaining tracking process, we propose to use this a priori knowledge during the whole tracking process to preclude VSLAM from the problem of error accumulation.

Indeed, if only the multiview constraints are used to estimate the motion of the camera, the resulting camera trajectory and scene reconstruction will be subject to
error accumulation. The resulting drift can be observed, for example, at keyframes by projecting the CAD model onto the image with the camera pose estimated by the VSLAM process. The error on the camera pose will generate an offset between the sharp edges of the CAD model and their corresponding contours in the image. If the constraints provided by the CAD model were strictly respected, these offsets would be null.

To prevent drift, we introduce the constraints provided by the CAD model directly into the VSLAM process. Consequently, the optimal trajectory and environment reconstruction must minimize simultaneously the multiview constraints (i.e., the reprojection errors of the 3D features reconstructed online) but also the constraints provided by the CAD model (i.e., the offset between the projection of the 3D sharp edges of the CAD model and their corresponding contours in the image).

17.4.2.2 Implementation

We propose to use model-based constraints provided by the sharp edges of the 3D model (i.e., for a polygonal model, edges used by two triangles and whose dihedral angle is superior to a threshold). Similar to Drummond and Cipolla (2002), these sharp edges are sampled in a set of oriented points \( \{ \mathbf{L}_i \}_{i=1}^k \), usually referred as an edgelet, each point being parameterized by its 3D position \( \mathbf{M}_i \) and the 3D direction \( \mathbf{D}_i \) of the sharp edge from which it was extracted. An illustration of an edge-based model for the bodywork of a car is illustrated in Figure 17.4.

To introduce the edge-based constrain into a VSLAM process while keeping its real-time performance, we propose to modify exclusively the BA process. Indeed, since it can be achieved as a background process (Klein and Murray 2007), the performance of the tracking process will not be altered by our modifications. However, modifying the BA will affect the result provided by the tracking process since the BA refines the position of the 3D features used to estimate the camera pose.

In the BA, the edge-based constraint takes the form of an additional term \( h_{CAD} \) in the cost function. This term corresponds to the orthogonal distance between the
projection of the edgelet $P_jM_i$ and its corresponding contour $m_{i,j}$ in the image. This corresponding contour is usually defined as the nearest point of contour located along the normal of the edgelet direction with an orientation similar to that of the projected edgelet. Consequently, the deviation $h_{\text{CAD}}$ of the scene $\vartheta$ with respect to the CAD-model constraints is defined as follows:

$$h_{\text{CAD}}(\vartheta) = \sum_{i=1}^{n} \sum_{j \in B} |n_{i,j} \cdot (m_{i,j} - \pi(P_j,M_i))|$$

where

- $n_{i,j}$ is the normal of the projected direction $P_jD_i$
- $\pi$ is the camera projection function of a 3D edgelet $M_i$ with respect to a camera pose $P_j$
- $B$ is the set of keyframe indexes for which the edgelet $M_i$ is associated to a 2D contour point $m_{i,j}$

The optimal scene reconstruction $\vartheta_{\text{optim}}$ is therefore defined as the scene that minimizes simultaneously the multiview and CAD-model constraints:

$$\vartheta_{\text{optim}} = \min_{\vartheta} (\varepsilon_R(\vartheta) + h_{\text{CAD}}(\vartheta))$$

This biobjective cost function is optimized through a BA. More details about its implementation can be found in Gay-Bellile et al. (2012).

17.4.3 Discussion

The combination of model-based and VSLAM approaches provides two main benefits. On the one hand, the constraints provided by the CAD model provide accuracy to the VSLAM process. Indeed, it defines a reference frame but also prevents the drift of the scale factor and the error accumulation, which are the main drawbacks of the standard VSLAM process. Moreover, our solution provides a better robustness to coarse initialization than usual solutions that bootstrap a VSLAM process with a model-based solution (Bleser et al. 2006). In fact, the error that a coarse initialization induces on the online reconstruction will be progressively corrected since the model-based constraint is optimized for several keyframes simultaneously. Consequently, our constrained VSLAM process is easier to deploy and provides a more accurate online reconstruction.

On the other hand, the reconstruction of the unknown environment realized by the VSLAM process provides the robustness that was missing in the standard model-based approach. Therefore, the features used to estimate the camera poses are distributed over the whole image since the whole environment is used to estimate the pose of the camera. Even if the object is small in the image, occluded, or out of the field of view, its location with respect to the camera can still be estimated from the observation of its surrounding environment. Moreover, because the edgelet-to-contour association requires a small motion assumption, standard model-based solutions are not robust to fast motion. However, in our constrained VSLAM approach,
a first motion estimation is provided by the VSLAM process. Since the edgelet-to-contour matching process is achieved from the camera pose provided by this first estimation, the small motion assumption is always checked.

Also, since VSLAM and geometric model-based tracking are both robust to illumination variations, the resulting constrained VSLAM maintains this robustness. Consequently, our solution provides both accuracy and robustness, thus meeting the requirement of quality and service continuity. On the other hand, the ease of deployment requirements is also fulfilled by our constrained VSLAM solution since it uses only a standard camera and CAD model that are widely used in the automotive industry, and since it can be coarsely initialized.

Our constrained VSLAM solution was successfully applied on various parts of a vehicle, such as the bodywork, a cylinder head, and the automobile’s interior. Some results are shown in Figure 17.1 where our approach was used for sales assistance using AR technology. Notice that these scenarios would have been extremely challenging given the usual model-based or VSLAM approaches since the bodywork provides few sharp edges while its texture is unstable since it is mostly generated by specular reflections of the surrounding environment.

17.5 VEHICLE LOCALIZATION FOR AIDED NAVIGATION IN AN URBAN CONTEXT

While standard navigation aiding systems often settle for a coarse localization provided by a standard GPS system, the use of AR requires an accurate 6DoF localization at high frequency. Indeed, while a driver can mentally compensate an inaccurate localization in a usual exocentric top-down view of the road network, this is no longer the case when the road to follow is displayed in an egocentric view. In the latter case, navigational instructions must be perfectly aligned with the real world in order not to mislead the driver.

In the following material, we present an overview of the accurate localization solutions for large-scale localization in a city. Among these existing solutions, we then describe in detail the fusion of VSLAM with GPS data. Finally, we present our solution that uses the GIS models to improve the localization accuracy of the fusion of VSLAM with GPS data.

17.5.1 State of the Art

Accurate vehicle localization is a very active field of research among the autonomous vehicle community. However, most of the solutions require sensors (LIDAR, GPS RTK, etc.) that are too expensive for our context. Consequently, we will mostly focus on vision-based solutions that can be initiated with low-cost cameras.

A first family of vision-based approaches relies on the exploitation of a geo-referenced landmark database that has been previously built off-line. Among these solutions, we can usually distinguish two approaches. The first one uses a database of geo-referenced 2D pictures, such as Google StreetView. The vehicle localization is therefore defined as the location of database images whose appearance is the most similar to the appearance of the live video stream (Cummins and Newman 2011). However, those solutions, usually referred as topological VSLAM,
Applications of Augmented Reality for the Automotive Industry

do not provide an accurate 6DoF localization and, consequently, do not meet the quality of service criterion expressed earlier.

The second approach relies on a database of 3D visual features, such as 3D points associated with a description of their visual appearance. This approach corresponds to a photo-geometric 3D model that can be used online by model-based tracking solutions (Lothe et al. 2009) previously introduced in Section 17.4.1 to estimate the 6DoF of the camera pose at high frequency. However, this approach is subject to serious drawbacks. First, the databases are not widely distributed and their constructions are usually expensive since it tries to collect data from a dedicated vehicle that moves along all the streets of all the cities. Second, the localization process is usually not robust to large illumination variations with respect to illumination condition observed during the database construction. Therefore, the continuity of service cannot be guaranteed with this solution.

To avoid the drawbacks induced by landmark database, another family of vision-based approaches relies on the exploitation of VSLAM (cf. Section 17.4.1). However, as already mentioned, this approach is subject to error accumulation, and scale factor drift when a single camera is used. To prevent these problems, most of the solutions try to fuse the motion estimated by VSLAM with the location provided by GPS (Schleicher et al. 2009). While these solutions provide high frequency localization, the estimation remains inaccurate. Indeed, the accuracy of the in-plane parameters (cf. Figure 17.5) are usually limited to the GPS accuracy, while the out-of-plane parameters (cf. Figure 17.5) are not estimated or result in high uncertainty.

Consequently, none of the previously mentioned solutions provide at the same time the service quality/continuity and ease of deployment. However, in the following, we will demonstrate that the VSLAM approach can fulfill all these requirements when it is correctly constrained with additional data, such as that derived from a GPS and a geographic information system (GIS).

17.5.2 Constrained VSLAM for Large-Scale Vehicle Localization

To provide a solution that fulfills both the quality and ease of deployment requirements, we propose to combine the VSLAM approach with data provided by a GPS
and a GIS, both being low cost and widely used. Similar to the approach that we introduced for tracking vehicle components, our solution relies on a constrained VSLAM framework. The two solutions only differ by the nature of the constraints used.

In the following, we will first describe how the in-plane DoF of a VSLAM can be coarsely constrained with GPS data (Section 17.5.2.1). Then, we will demonstrate that a standard GIS can be used to refine both the in-plane (Section 17.5.2.2) and out-plane DoF (Section 17.5.2.3) of the localization.

### 17.5.2.1 Constraining VSLAM with GPS

The GPS constraint to consider is that the position of the camera positions provided by the VSLAM must follow the ones provided by the GPS. In fact, if we consider that the camera is rigidly embedded to the GPS and the distance between these two sensors is negligible, then this constraint can be observed by analyzing the gap between the GPS measurements and the camera positions. It results in the following GPS constraints:

\[
\begin{align*}
||t_j - (g_j)||^2 + \begin{pmatrix}t_{jx} - (g_{jx}) \end{pmatrix}^2 = \sum_{j=1}^{n} ||t_j - (g_j)||^2
\end{align*}
\]

where

\[
(t_{jx}, t_{jy}) \text{ is the jth camera’s position and } (g_{jx}, g_{jy}) \text{ its associated GPS data}
\]

1 is the number of the camera optimized in the BA

Including this GPS constraint in the BA gives

\[
\varepsilon_A(\hat{\theta}) = \varepsilon_R(\hat{\theta}) + w \times h_{GPS}(\hat{\theta})
\]

A weight \( w \) must be added to take into account the uncertainty of the GPS data and its variation over time (which is not the case for the CAD model constraints since it is assumed error free). The presence of aberrant data makes the weight estimation a challenging problem. To avoid this problem, Lhuillier (2012) proposes a new formulation of the constraint that allows more robustness against inaccurate data. The principle is to progressively introduce the additional constraints (GPS, 3D model) while not degrading significantly the multiview geometry (i.e., the reprojection error) beyond a certain threshold:

\[
\hat{\theta} = \begin{cases} 
\text{arg min } h(\theta) \\
\hat{\theta} \\
\varepsilon_R(\hat{\theta}) < \varepsilon_t
\end{cases}
\]

where

\( \hat{\theta} \) is a vector containing all the parameters optimized in the local BA (the \( n \) 3D points and the \( l \) camera poses)

\( \varepsilon_R(\hat{\theta}) \) is the reprojection error previously introduced in Section 17.4.1
The inequality constraint $\varepsilon_R(\theta) < e_t$ can be formalized through a cost function that contains, in addition to the constraint term $h(\theta)$, a regularization term prohibiting a degradation of the reprojection error beyond the predefined threshold $e_t$. This regularization term, computed from the standard reprojection error, has a negligible value when the mentioned condition is respected and tends toward infinity when the considered condition is close to being broken. Consequently, the resulting cost function is given by

$$\varepsilon_I(\theta) = \frac{w}{e_t - \varepsilon_R(\theta)} + h(\theta)$$

where $w$ is a constant that attributes a negligible influence to the regularization term when the degradation of the reprojection error is low.

To realize a BA with the inequality constraint, two steps are required. The first consists in performing a classic BA where the standard cost function $\varepsilon_R(\theta)$ is minimized. This step allows us to determine the maximal degradation of the reprojection error $e_t$. Next, the additional data are introduced through the constrained BA with inequality constrain by minimizing $\varepsilon_I(\theta)$.

Fusing GPS data with VSLAM through the constraint BA described earlier results in a 6DoF geolocalization. However, even if aberrant GPS data are filtered due to the inequality constraints, the accuracy of the resulting camera positions are broadly equivalent to the GPS uncertainties. Furthermore, adding GPS data in the BA process constrains only the camera position and implicitly the yaw angle (or cap) of the camera orientation, we call these parameters the in-plane camera parameters (cf. Figure 17.5). The other degrees of freedom (altitude, pitch, and roll angle) that we call out-plane parameters (cf. Figure 17.5) are not constrained and may drift over time. Consequently, constraining a VSLAM with a GPS is not sufficient to reach the required accuracy. In the following, constraints provided by a GIS will be added to solve this remaining problem.

### 17.5.2.2 Improving the In-Plane Accuracy with GIS Constraint

Nowadays, GISs are constituted of several layers, such as a road map, a digital elevation model (DEM) of the ground, or a 3D buildings model. This is the last layer that we will use as an additional constraint to improve the in-plane accuracy of our VSLAM process.

An intuitive approach would consist of adding the model-based constraints introduced in Section 17.4.2 in the BA constrained to GPS data (cf. Section 17.5.2). However, there are two major difficulties with this intuitive approach. The first is inherent to the quality of the buildings models that are less accurate and detailed than CAD models of small objects. Even if there exist high quality models of cities, they are currently expensive and not available everywhere. The most widespread models are obtained by aerial photogrammetry thus limiting the accuracy obtained, which may cause errors (up to 2 m). Furthermore, the resulting models are very simple (no details: doors, windows, etc.) and textureless. For these reasons, exploiting 3D features extracted from a city model in a constrained BA is not feasible. We propose in Section 17.5.2.2.1 another constraint that is well suited to the simplicity of buildings models.
The second difficulty is that the GPS and buildings constraint both act (explicitly or implicitly) on the camera position and have different *optimal* solutions due to the important GPS uncertainty in dense urban areas. Thus, merging these two constraints in a unique constraint BA can lead to convergence problems. Therefore, we propose a solution that uses the buildings models to remove the bias that affect the GPS data before including them in the BA process.

### 17.5.2.2.1 Buildings Constraint

The principle of the constraint provided by the 3D buildings models is based on the following hypothesis: a perfect VSLAM reconstruction (i.e., without any drift) must lead to a 3D point whose cloud will be almost aligned with the 3D buildings models of the observed scene. Consequently, it is possible to evaluate the compliance or the noncompliance of the buildings constraint by measuring the difference between the positions of the reconstructed 3D points and their corresponding facades in the 3D buildings model. However, all the 3D points reconstructed by the VSLAM algorithm do not necessarily belong to a facade. In fact, some 3D points may represent elements belonging to the parked cars, trees, and road signs. This set of points is not concerned by the buildings constraint. Consequently, the first step to establish this constraint is to identify the set $M$ of 3D points that belongs to buildings fronts and associate each one to its correspondent facade. The constrained term associated to the buildings model must measure the distance between each point $Q_i \in M$ and its corresponding facade $\nabla h_i$. For that, each point $Q_i \in M$ is expressed in the coordinate frames of its facade:

$$
\begin{pmatrix}
Q_{hi} \\
1
\end{pmatrix} = \mathbf{T}_{hi} \begin{pmatrix}
Q_i \\
1
\end{pmatrix}
$$

where $\mathbf{T}_{hi}$ is the $(4 \times 4)$ transfer matrix from the world coordinates frame to the facade $\nabla h_i$ coordinates frame. In this coordinate frame the distance between the 3D points and the buildings model is simply given by the $z$ coordinate of $Q_{hi}$. Therefore, the compliance of a point cloud with respect to the building constraint can be estimated as follows:

$$
h_{\text{bdg}}(\theta) = \sum_{i \in M} \left\| Q_{hi} \right\|^2$$

To deal with many wrong 3D point/buildings model associations, a robust estimator works better than the L2 norm. More details on the buildings constraints are given in Tamaazousti et al. (2011).

### 17.5.2.2.2 Improving GPS Accuracy

It is a common assumption that the uncertainties of the GPS data can approximately be modeled by a bias and small Gaussian noise. The magnitude of this bias depends on many parameters, as, for example, the constellation of visible satellites, but can usually be approximated as locally constant. Determining and correcting locally the bias of the GPS will improve the accuracy of the VSLAM and GPS fusion.
Recent studies propose to correct the bias of standard GPS by using georeferenced information provided by GIS. In Kichun et al. (2013) road markings are used: white lines delimiting the road are detected in the images and matched with the numerical model of the road to estimate the lateral bias of the GPS. Crosswalks are also exploited with the same principle to calculate the bias in the direction of the vehicle displacement. However, white lines are not present on all roads and crosswalks are not frequent enough to estimate regularly the GPS bias. In addition, these road markings can be regularly occluded by other cars. Finally, this type of information (white lines and crosswalks) is currently not common in GIS, which makes this solution difficult to deploy.

To overcome these problems, we propose to estimate the bias of the GPS from the reconstruction obtained by the fusion of VSLAM and GPS data by exploiting the buildings that are more visible in urban areas and for which 3D models are widely available in GIS. The bias of the GPS is not directly observable, unlike the VSLAM reconstruction error generated by this bias that results in a misalignment between the reconstructed 3D point cloud and the buildings models. Our solution is based on the following hypothesis: error affecting locally (i.e., the n last camera position and the 3D points they observe) the VSLAM reconstruction after fusion with GPS data corresponds to a rigid transformation in the ground plane, that is, with 3DoF (position and yaw angle). This assumption appears to be a sufficient approximation for a coarse correction.

We add a correction module to the constrained VSLAM process that estimate the GPS bias at each keyframe as seen in Figure 17.6. This module takes as input the buildings model and the VSLAM reconstruction that is not perfectly aligned with the buildings model due to the GPS bias. The first step of this module is to identify which points of the 3D point cloud come from buildings facades or from the rest of the environment such as parked cars or trees. Once this segmentation is performed, the 3DoF rigid transformation is estimated by minimizing the distances between the 3D points associated with the buildings model and their associated facade. This transformation is then locally applied to the GPS data associated to the n last camera position to correct their bias.

**FIGURE 17.6** Overview of our proposed solution for accurate localization in urban environment. It combines VSLAM with GPS data and GIS models (buildings and DEM).
17.5.2.3 Improving the Out-Plane Accuracy with a SIG Constraint

To improve the accuracy of the out-plane DoF, we propose to constraint the VSLAM trajectory with the DEM layer of a SIG. The DEM is a 3D geometric model that provides a simple representation of the altitude variations of the roads network. In addition to its availability in urban environment and rural areas, this model has the advantage that it informs us about the out-plane degrees of freedom of the camera. In fact, since the camera is rigidly embedded in the vehicle, its altitude, roll, and pitch angles can be considered constant relative to the road. In the following, since we use very simple DEM that doesn’t provide the road inclination, we use only an altitude constraint and exploit the fact that the camera height $h$ relative to the ground is constant (i.e., we assume that the variations of the height related to the shocks are negligible). This implies that the trajectory of the camera must belong to a surface parallel to the road and located at a distance $h$ from the ground. It seems necessary to identify the road in the DEM where each camera pose is located. This association is based on a proximity criterion: a camera pose is associated with the neared road $\Delta k^j$ in the DEM in terms of orthogonal distance. In the constrained BA, the constrained term associated with the DEM corresponds to the difference between the camera altitudes $(t^k_j)_2$ expressed in its road plane coordinate frames and the desired height $h$:

$$h_{DEM}(\theta) = \sum_{j=1}^{L} \left\| (t^k_2 - h) \right\|^2$$

where $(t^k_j)_2 = L^k_j (t^k_j)_1$ with $L^k_j$ the $(4 \times 4)$ transfer matrix from the world coordinates frame to the road $\Delta k^j$ coordinates frame associated to the $j$th camera. More details on the DEM constraints are given in Larnaout et al. (2013a).

17.5.2.4 Overview of Our Complete Solution for Vehicle Localization

To summarize the proposed solution, it is a VSLAM approach with a constrained BA that includes buildings, DEM, and GPS constraints. The resulting cost function of the constrained BA is given by

$$\epsilon_t(\theta) = \frac{w}{e_i - e_R(\theta)} + h_{GPS}(\theta) + h_{DEM}(\theta) + h_{bdg}(\theta)$$

To overcome the convergence problem, GPS data are coarsely corrected beforehand through an additional module that acts as a differential GPS where geo-referenced antennas are replaced by a buildings model. Figure 17.6 presents an overview of the proposed solution.

We performed many experiments to evaluate our vehicle geolocalization solution. For our experiments, sequences of several kilometers of the roadway were acquired in France. To obtain these sequences, the vehicle was equipped with a standard GPS 1 Hz and an RGB camera providing 30 frames per second. The GIS models used are provided by the French National Geographic Institute with an uncertainty that do not exceed 2 m. We present results for one of these sequences in Figure 17.7.
The localization accuracy is clearly improved with the proposed solution. This is enhanced by the good alignment between the resulting 3D point cloud and the buildings model where considering the fusion of VSLAM with the raw GPS, data gaps can be observed between the resulting 3D points and the buildings model. Using a complete GIS model (buildings and DEM) to improve the fusion of VSLAM and GPS data yields accurate 6DoF localization that fulfill the quality criteria required by aided navigation application as illustrated in Figure 17.2.

17.5.3 Discussion

Using GIS models can improve the localization accuracy resulting from the fusion of VSLAM with GPS data. This fulfills both the quality and ease of deployment requirements of navigation-aided applications since GIS is now widely available and some GIS databases are free (e.g., OpenStreetMap 3D). Furthermore, since our approach relies on a VSLAM approach, it provides a continuity of service that solutions based on visual featured databases cannot guarantee. On the other hand, the localization quality provided by our solution is still less accurate than approaches based on a database. However, the accuracy provided by our solution remains sufficient for most
navigation-aided applications. The localization accuracy reached with the proposed solution can still be improved by using other low-cost sensors such as the odometer or inertial sensor. Indeed, the information on the vehicle trajectory provided by these sensors can be easily introduced in our generic constrained VSLAM framework by adding a new constraints term in the BA.

It is also possible to exploit the reconstruction provided by the constrained VSLAM as an initial database of 3D visual features that can be transferred to a server and then refined off-line to improve its accuracy (Larnaout et al. 2013b). Thus, subsequent users that circulate in this area would be able to locate themselves via model-based solutions. Our approach would be used to localize the vehicle and simultaneously update and extend the database only when a user enters an area that is not mapped in the database or mapped with an incompatible lighting condition. Our approach results in a collaborative solution to the database creation issue while maintaining the continuity of service in area devoid of database.

17.6 CONCLUSION

Applications of AR in the automotive industry are numerous. Many AR techniques for the automobile industry have been studied but few actually deployed. Even if recent advances in tracking technologies, such as the constrained VSLAM, allow engineers to remove the main technological locks, some challenges will still remain.

The first challenge relates to the ergonomy of the solution. For example, most of the applications require a hands-free device. While solutions are already available, such as spatial AR or semitransparent glasses, their ergonomy (e.g., width of the field of view, dynamic focus distance, luminosity, and contrast) should be improved to reach a high end user acceptance. But ergonomic issues are not limited to hardware. The displayed information should also be designed to facilitate the work of the end user without disturbing him or her or introducing potential dangers. For example, the iconography used to display information on a windshield must be designed to reduce the risk of hiding pedestrian or vehicles from the driver’s view.

The second challenge concerns the integration of AR in the product life management process. For example, the goal is to lower the cost of content creation of an AR application and facilitate its updates; therefore, the product data management (PDM) should also integrate the needs of AR applications. Further, the development of a 3D documentation could benefit from the 3D models and animations that were created during the conception stage of the vehicle. Consequently, the development of new norms and standards concerning 3D models, animations, interactions, and documentation will probably be necessary. In summary, this chapter presented an overview of the use of AR in the automotive industry; we expect more integration of AR in the product life cycle of automobiles in the future.

ACKNOWLEDGMENTS

We would like to thank Diotasoft for supporting our research with applications in sales support in concession and vehicle maintenance through AR. We also
thank Valeo and Renault for their financial supports that allow development of AR-based aided navigation solutions.

We are grateful to our colleagues at Vision and Content Engineering Laboratory and Michel Dhome from the Institute Pascal for many useful comments and insights that helped us to develop and refine the work.

REFERENCES


