A mobile markerless Augmented Reality system for the automotive field

Vincent Gay-Bellile*
CEA, LIST

Steve Bourgeois
CEA, LIST

Mohamed Tamaazousti
CEA, LIST

Sylvie Naudet Collette
CEA, LIST

Sebastian Knodel
Diotasoft

ABSTRACT

Recently, we introduced the concept of SLAM constrained with a CAD model. In this paper, we demonstrate that such concept can be extended to a generic and flexible tracking framework. The resulting solution is validated through various Augmented Reality applications dedicated to the automotive field: assisted maintenance on car engine, customization of a car bodywork and user guide for dashboard. In spite of this variety of applications and contexts, the same framework is used on a consumer tablet to provide an accurate and real-time tracking system. Videos are available as supplemental materials[1].

Index Terms: Sensors [Vision-based registration and tracking]: ;— [System architecture]: Performance issues [real-time approaches], Wearable and mobile computing—; MR/AR applications [Industrial and military MR/AR applications]: —

1 INTRODUCTION

Augmented reality (AR) has assuredly a great potential in the automotive field since there is many application where adding visual information has a real interest. Process efficiency of sales and after-sales in the automotive field can be greatly improved by using augmented reality technologies. Customizing cars by customers, explaining the operation of the vehicle at the physical scale allows an immediate understanding and ownership by user faster. Similarly, it is easier to show a maintenance operation than to explain it.

In spite of a great number of potential applications Augmented Reality systems in automotive field are not widely spread. The main reason is that they required a very accurate and stable (*i.e.* absence of jitter) localization to obtain a realistic rendering. This is very important for acceptability of Augmented Reality systems. Until then, existing tracking solutions do not reach these both requirements. Model-based tracking solutions imply that the object of interest is always visible and represents a large part of the images during the whole sequence for stable tracking. On the other hand, SLAM solutions are stable but suffer from error accumulation and scale ambiguity that introduce a less and less accurate localization over time.

Recently, we have proposed in [11] a constrained bundle adjustment framework for SLAM algorithm. The principle is to include the constraints provided by a 3D model of one object of interest in the bundle adjustment process to deal with disadvantage of slam algorithm for AR purposes. It results in a very accurate and stable localization. This previous result leaded to a collaborative study between Renault, Diotasoft and CEA LIST. The objective of this study is to evaluate the ability of this tracking framework to provide an efficient Augmented Reality platform for the automotive field. Three augmented scenarios are considered for this study: interactive customization of the car body, user guide for the dashboard and assisted maintenance of the car engine. To deals with such variety of context, we present in this paper an extension of the original framework with a constraint based on a point-based model. The flexibility of the constrained bundle adjustment is also enhanced through the combination of multiple constraints in this single framework. The scenarios described above allow evaluating the performances of the resulting tracking process for the different parts of the car (dashboard, car body, engine) and illustrate the benefits of AR for the automotive industry through real applications. This paper presents the results of this study.

Plan. Section 2 briefly describes the related works of Augmented Reality in the automotive field. The tracking system we used and its improvements for this context of study are introduced in Section 3. Then, we present our AR platform relying on a consumer tablet designed for improving the AR experience. We disclose the results of the tracking evaluation for the three scenarios in Section 4. Finally, we give our conclusions and discuss further works, in Section 5.

2 RELATED WORK

Automotive is probably the domain that is the most frequently used to illustrate the potentiality of Augmented Reality. The first applications were based on markers and used for virtual car visualization, see [1, 3]. Then, improvement in template tracking solutions *e.g.* [13] allowed the deployment of virtual advertising on paper especially for automobile[2]. Car toys are also frequently used to demonstrate the ability of model-based tracking system, see *e.g.* [15].

Nevertheless, in spite of the fact that the BMW video-concept[3] of aided maintenance is among the most famous video of Augmented Reality, AR system for real size car are not widely spread. This situation relies on the fact that a car is not favorable for appearance-based and geometry-based tracking solutions. To our knowledge, only Metaio demonstrated in [8] the ability of its 3D visual tracking system to handle a real car. Nevertheless, since the tracking process relies exclusively on the visual appearance of 3D feature points, this solution was only demonstrated on the most textured parts of the car, *i.e.* the engine.

Among the car components, the bodywork is the most challenging for tracking solutions. Indeed, on one hand, due to its volume, it is not fully observable from a single viewpoint. Combining this with an unevenly distribution of the sharp edges make geometry-based tracking such as [2, 5, 15] unstable. On the other hand, the lack of texture over the car body and the presence of specularities and reflections are not favorable for appearance-based tracking such as [5, 9, 14]. Existing solutions for the textureless parts of the car are based on expansive additional sensor, such as a high-precision robotic measurement arm or an ART tracking system. This is why most of the Augmented Reality demonstration in the last automo-

*CEA, LIST, Vision and Content Engineering Laboratory, Point Courrier 94, Gif-sur-Yvette, F-91191 France. e-mail: vincent.gay-bellile@cea.fr

[1] https://vimeo.com/45906225

[2] http://www.youtube.com/watch?v=pFS6EHzBGVc

[3] http://www.bmw.com/com/en/owners/service/augmented_reality_introduction_1.html

tive shows (*e.g.* Mahindra XUV500 at Auto Expo 2012[4]) still relied on a motionless camera.

In the next section, we describe a single markerless vision-based system that is able to track the different car components, included the bodywork, for Augmented Reality applications.

3 The proposed Markerless Tracking Solution

In this section, we present our tracking framework based on a constrained SLAM. We firstly describe in Section 3.1 the principle of the constrained SLAM that unify Keyframe-SLAM and model-based tracking solutions through a constrained bundle adjustment framework. Then, we introduce the constrained SLAM with an edge-based model in Section 3.2 and with a point-based model in Section 3.3. Finally, the flexibility of our framework is highlight in Section 3.4 through the combination of edge-based and point-based constraints in a SLAM process.

3.1 Principle of constrained SLAM

SLAM and model-based tracking solutions are the two main approaches to estimate in real-time the displacement of the camera. The former is based on a combination of two processes. The first one is the reconstruction and the refinement of a 3D map of features (*e.g.* 3D points, 3D segments, *etc*) from their 2D observations in multiple views. The second process estimate the camera poses by matching 2D features of the current view with this 3D map. Slam algorithms allow estimating the camera motion in an unknown environment and since 2D features of the whole image are used for the tracking, the resulting localization is not subject to jitter and is robust to large motions. Nevertheless, since the 3D map is reconstructed on-line, SLAM process is subject to error accumulation and the reference frames as well as the scale factor are unknown. In the rest of the paper, we concentrate on SLAM solutions that reconstruct 3D points on-line, *e.g.* [4, 6].

On the other hand, model-based solutions rely on a prebuilt map of 3D features. The pose of the camera is estimated on-line by matching 2D features extracted from the images with the 3D features of the map. Since those 3D features are error-free, the resulting camera localization is not subject to error accumulation, and the reference frame is the one associated to the 3D map. Nevertheless, since the camera motion is only estimated from the 2D features associated to the object model, the tracking is sensitive to jitter and failure when the object is small in the image or occluded.

We recently proposed in [11] a unified solution by constraining the bundle adjustment of the SLAM process with the constraints provided by the 3D map inherent to model-based trackers. Since these model-based features are error-free, their positions are not optimized in the SLAM process (*i.e.* they are 0-DoF features in the bundle adjustment). Nevertheless, their reprojection errors are taken into account in the bundle adjustment. They constrain directly the camera motion and thus indirectly the features reconstructed on-line by the SLAM process.

To underline the benefits of our constrained SLAM framework, two situations have to be considered:

- The object is visible in the image: the 0-DoF features provide the reference frame, the scale factor and prevent error accumulation. Therefore, since the camera localization is improved, the features reconstructed by the SLAM are more accurate.

- The object is occluded or small in the image: the 3D features reconstructed on-line by the SLAM process maintain an accurate tracking as long as the object is static with respect to its environment.

[4] `http://www.youtube.com/watch?v=iA-2ElU9Qs8`

In the following, two kind of model-based features are considered: edges and points. Then, we will show that these model-based features can be simultaneously integrated in the SLAM process through a single constrained bundle adjustment.

3.2 Constrained SLAM with an Edge-Based Model

In [12], the combination of a SLAM process with an edge-based model has been introduced for texturless 3D object tracking. The initial map of 3D segments is obtained by sampling off-line the sharp edges of a CAD model. An example of edge-based model is illustrated in Figure 2. During the SLAM process, these 3D segments are projected in the image and matched to the nearest contour with similar orientation. The orthogonal distance between the 3D segments projection and their associated contours (edge-to-model distance) are used as additional residual errors in a constrained bundle adjustment. It results a two terms cost function: the first one correspond to the reprojection errors associated to the points reconstructed on-line by the SLAM process and the second term correspond to edge-to-contour distances. The weighting of these two terms is handled with the strategy introduced in [11]. This solution may be disturbed by highly textured object since the probability of matching errors increase due to the contours generated by the texture.

3.3 Constrained SLAM with a Point-Based Model

When the object of interest is highly textured, using a point-based model seems to be a more adapted alternative. A point-based model is a 3D points cloud of the object of interest build off-line by a SfM reconstruction process (*e.g.* [10]) or directly extracted from the texture of a CAD model. An example of point-based model is illustrated in Figure 1.The 3D points are completed with one or more appearance descriptors. These prebuilt points are similar to the points reconstructed by the SLAM process except their position is kept constant during the tracking process (0-DoF points). Therefore, it results a constrained bundle adjustment that minimizes a two terms cost function: the first one correspond to the reprojection error of the 3-DoF points reconstructed on-line by the SLAM process whereas the second term correspond to the reprojection error of the 0-DoF points provided by the point-based model. The weighting of the two terms is also handled with the strategy introduced in [11].

3.4 Constrained SLAM with an Edge&Point-Based Model

The constrained SLAM framework is flexible thus combining different source of constraints is very simple. For example, it might be very useful for not too highly textured object to use both constraints provided by edge-based and point-based models. If both models are expressed in the same coordinate frame, the combination results in a constrained bundle adjustment that minimizes a three terms cost function. The first one correspond to the reprojection error of the 3-DoF points reconstructed on-line by the SLAM process, the second one correspond to the reprojection error of the 0-Dof points of the point-based model and the third one correspond to the edge-to-contour distances. The weighting of these three terms is also handled with the strategy introduced in [11].

4 Experimental Results and Applications on Different AR Scenarios

In this section, we demonstrate that our tracking solution describe in Section 3 offers the sufficient flexibility, accuracy and performance to handle various automotive AR applications on a mobile platform. First, the mobile platform used for the experiments is described in Section 4.1. Then, three applications are considered each of them focus on a specific part of the car: engine (Section 4.2), car bodywork (Section 4.3) and car dash (Section 4.4). Since

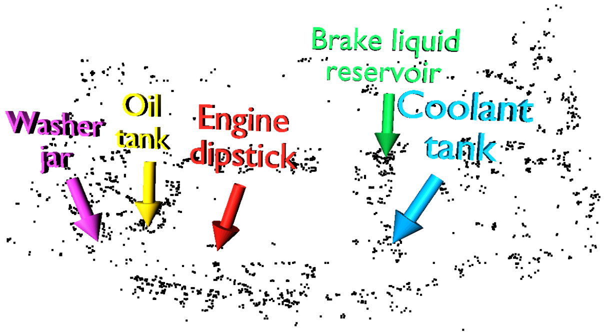

Figure 1: The point-based model of the engine with registered virtual informations used for the augmentation.

Figure 2: The edge-based model of the car bodywork.

no ground truth was available, the evaluation is reduced to a visual assessment.

4.1 Mobile platform setup

Our augmented reality system uses a Selltic[5] tablet, which was originally developed as a technological solution for the field of consultative selling. The tablet's casing was custom designed to embody all hardware components and to provide a better usability for the user. The Selltic tablet uses a Samsung Series 7 Slate tablet PC (Intel Core i5-2467M Processor, Intel HD Graphics 3000 and Dimensions (11.66" (W) x 7.24" (H) x 0.51" (H))) and a wide-angle uEye 2210SE camera (Focal length 3.5 mm, 640×480 pixels) providing images up to 45 fps. The video camera is integrated in the tablet's casing, as presented in Figure 3. The camera's video signal is transmitted to the tablet, where both tracking and rendering are realized. Note that the tracking can also be performed by using the tablet's integrated camera. We decided to use an external camera to preserve a good image quality and thus improve the Augmented Reality experience of the consumers. The image processing time of our tracking algorithm on this tablet is 25ms on average which is close to the maximal image acquisition frequency.

4.2 Augmented Reality for engine maintenance and repair

The first Augmented Reality application is the assistance of an operator during maintenance and repair operations. Virtual elements are added to guide the operator during the maintenance procedure. Such scenario implies that the localization process to be robust to large changes of viewpoint, fast motion and occlusions (introduced by the hand of the operators for example).

The object of interest is an engine of a Renault Clio which is placed in an outdoor context that implies challenging illumination conditions. Since the engine is highly textured, we use a point-cloud model to constrain the bundle adjustment. This model, illustrated in Figure 1, is built off-line with a multi-view reconstruction process and is composed of approximately 2000 points. The system yields an accurate, stable and robust localization and is thus perfectly designed for this Augmented Reality application. This accuracy can be appreciated in Figure 4, where different parts of the engine (dipstick, bake liquid reservoir, coolant tank, ...) are pointed out virtually by the system.

Positions of the engine dipstick, brake liquid reservoir, coolant tank, *etc*, have been virtually indicated.

[5] *Selltic*^{TM} is a brand distributed by Diotasoft which provides a set of technological solutions intended for supporting sales in storehouses and beyond.`www.selltic.eu`.

4.3 Virtual car bodywork customization through Augmented Reality

The second application is the interactive customization and virtual tuning of the car bodywork through Augmented Reality. The goal is to enable the costumer in car dealership to see the different designs of a car's model *i.e.* colors, materials, *etc* while moving around it. Such scenario implies that the localization process to be robust to the lack of texture on the car bodywork. Moreover the whole bodywork is not seen continuously, especially if the user approaches the car, which makes the tracking challenging.

Since the car's bodywork is textureless, the combination of SLAM with an edge-based model is selected. To achieve real-time performances, the CAD model of the Renault Clio used in our experiments is simplified to 50000 triangles. 2000 3D segments are then extracted from this model as illustrated in Figure 2. The car is placed in an indoor parking.

In our experimentation, the personalization of the car is done by changing, very realistically, the color of the bodywork and by adding white strips and black plastic elements. Thus the classical Renault Clio car is virtually transformed with the Clio Gordini design as shown in Figure 5. Note that the personalization application can be extended to individual custom tunning. Indeed, our tracking system is enough accurate to add some virtual elements like a spoiler, or to change the color of the windows, windshields and the bumpers of the car.

4.4 Virtual customization of the car dash through Augmented Reality

The last applications focus on the dashboard of the car. Augmented reality is used to customize the dashboard and to provide an interactive user guide to enhance the car properties by indicating the different features of the dashboard. It is a challenging application since a high tracking accuracy is required while the lack of recoil promotes fast and pure rotational movements of the user. The lack of light within the vehicle is also not favorable for the tracking.

Since the dashboard is not too highly textured and a CAD model is available, the combination of SLAM with both edge-based and point-based models is selected. The 3D model of a Renault Clio used in our experiments is composed of 50000 triangles. We extract from this model 2000 3D segments. The pre-build point cloud model is for its part composed of approximately 1200 3D points. The tracking is accurate during the whole sequence in spite of large scale variations and fast displacements as illustrated in Figure 6. The design of the car dash has been changed while the functionality of some elements are virtually indicated.

4.5 Discussion

In summary, our constrained SLAM framework achieves high accuracy and stability and can deal with different illumination con-

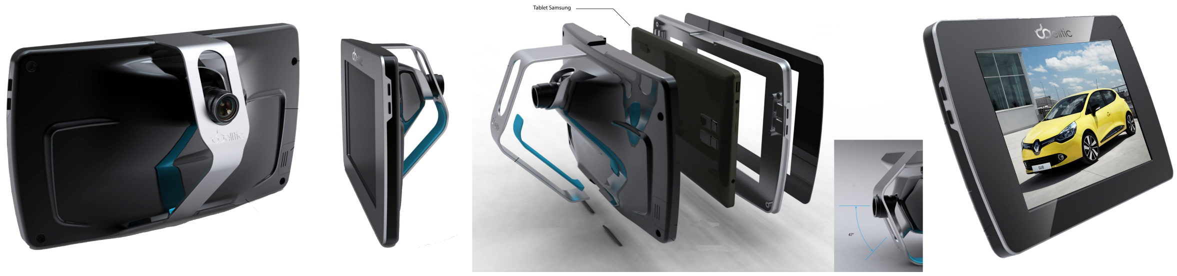

Figure 3: The Selltic tablet, containing a Samsung slate tablet PC and a uEye camera.

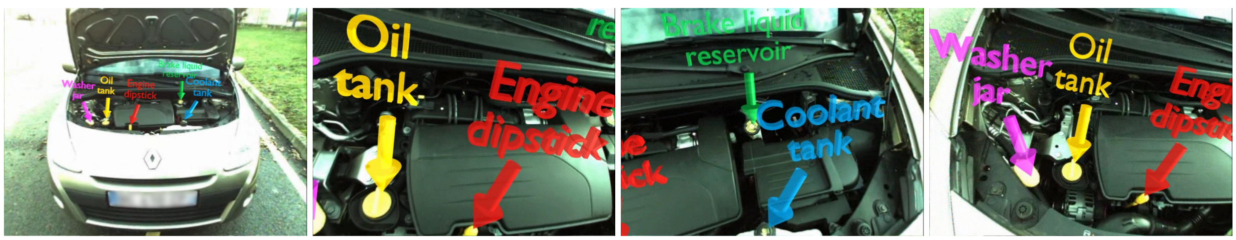

Figure 4: Augmented Reality on a car engine.

Figure 5: Augmented Reality on a Renault Clio car. The car with its original color (Left) and the same car virtually personalized. Note that only its bodywork has been changed, the wheels, the windows and the windshield are real.

Figure 6: Virtual customization of the car dash through Augmented Reality. The original car dash (Left) is customized. Virtual information are also added to enhance some car properties.

ditions during the tracking stage, partial occlusions and rapid motions. Our system underwent several extensive test series under real conditions and proved to be useful for different scenarios such as engine maintenance and repair, virtual car bodywork customization, *etc*.

To initialize our tracking algorithm (coordinate frames and scale), we use the point-based model combining with classical relocalization techniques such as *e.g.* [7] for the car dash and the engine. However, this approach is not robust enough for texturless objects such as car bodyworks. We rather learned the appearance of the car's environment in an off-line step and is used on-line to initialize the constrained SLAM until the car is not moved.

5 Conclusion

We have demonstrated in this paper that our tracking framework is perfectly designed for Augmented Reality in the field of automotive. It allows selecting or combining different kind of models. The experimental results prove that such flexibility allows handling a large diversity of scenes, such as engine, dashboard and car bodywork, with high stability, accuracy and robustness. Real-time performances have been achieved on a consumer tablet which allows envisaging a large deployment of Augmented-Reality applications in the automotive field.

Further work will concentrate on the initialization of the presented tracking system for textureless car components such as the bodywork. Currently, we deals with the initialization step by using the car's environment. The counterpart is that the environment has to be relearned at each displacement of the car. We study an initialization procedure for textureless objects based on segments extraction, description and matching. We also plan to extend our constrained SLAM framework to the on-line reconstruction of 3D segments.

6 Acknowledgements

We wish to thank Renault SAS and its Information Systems Department for the funding concerning the automotive applications of Augmented Reality presented in this paper, and more particularly Delphine Stoll and Frederic Nivoix for their support.

References

[1] Arvika. `www.arvika.de`.

[2] Tom Drummond and Roberto Cipolla. Real-time visual tracking of complex structures. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(7):932–946, 2002.

[3] Denis Kalkofen, Erick Mendez, and Dieter Schmalstieg. Interactive focus and context visualization for augmented reality. In *Proceedings of the 2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality*, ISMAR, pages 1–10, Washington, DC, USA, 2007. IEEE Computer Society.

[4] Georg Klein and David Murray. Parallel tracking and mapping for small AR workspaces. In *ISMAR*, 2007.

[5] Vincent Lepetit and Pascal Fua. Monocular model-based 3d tracking of rigid objects: A survey. In *FTCGV*, 2005.

[6] E. Mouragnon, M. Lhuillier, M. Dhome, F. Dekeyser, and P. Sayd. Real time localization and 3d reconstruction. In *CVPR*, 2006.

[7] David Nister and Henrik Stewenius. Scalable recognition with a vocabulary tree. In *CVPR*, pages 2161–2168, 2006.

[8] Juri Platonov, Hauke Heibel, Peter Meier, and Bert Grollmann. A mobile markerless ar system for maintenance and repair. In *ISMAR*, 2006.

[9] Iryna Skrypnyk and David G. Lowe. Scene modelling, recognition and tracking with invariant image features. In *ISMAR*, pages 110–119, 2004.

[10] Noah Snavely, Steven M. Seitz, and Richard Szeliski. Photo tourism: Exploring photo collections in 3d. In *SIGGRAPH Conference Proceedings*, pages 835–846, New York, NY, USA, 2006. ACM Press.

[11] Mohamed Tamaazousti, Vincent Gay-Bellile, Sylvie Naudet-Collette, Steve Bourgeois, and Michel Dhome. Nonlinear refinement of structure from motion reconstruction by taking advantage of a partial knowledge of the environment. In *CVPR*, pages 3073–3080, 2011.

[12] Mohamed Tamaazousti, Vincent Gay-Bellile, Sylvie Naudet-Collette, Steve Bourgeois, and Michel Dhome. Real-time accurate localization in a partially known environment: Application to augmented reality on 3d objects. In *International workshop on AR/MR registration, tracking and benchmarking*, 2011.

[13] Simon Taylor, Edward Rosten, and Tom Drummond. Robust feature matching in 2.3μs. In *IEEE CVPR Workshop on Feature Detectors and Descriptors: The State Of The Art and Beyond*, 2009.

[14] Luca Vacchetti and Vincent Lepetit. Stable real-time 3d tracking using online and offline information. *PAMI.*, 2004.

[15] Harald Wuest, Folker Wientapper, and Didier Stricker. Adaptable model-based tracking using analysis-by-synthesis techniques. In *Proceedings of the 12th international conference on Computer analysis of images and patterns*, CAIP, pages 20–27, 2007.