Augmented Reality Based Traffic Sign Recognition for Improved Driving Safety

Lotfi Abdi[1], Aref Meddeb[2], and Faten Ben Abdallah[2]

[1] National Engineering School of Tunis
University of Tunis-El Manar, Tunis, Tunisia
lotfiabdi@hotmail.com
[2] National Engineering School Of Sousse
University of Sousse, Tunis, Tunisia
Aref.Meddeb@infcom.rnu.tn,
faten.benabdallah@enit.rnu.tn

Abstract. In recent years, automotive active safety systems have become increasingly common in road vehicles since they provide an opportunity to significantly reduce traffic fatalities by active vehicle control. Augmented Reality (AR) applications can enhance intelligent transportation systems by superimposing surrounding traffic information on the users view and keep drivers and pedestrians view on roads. However, due to the existence of a complex environment such as weather conditions, illuminations and geometric distortions, Traffic Sign Recognition(TSR) systems has always been considered as a challenging task. The aim of this paper is to evaluate the effectiveness of AR cues in improving driving safety by deploying an on-board camera-based driver alert system against approaching traffic signs such as stop, speed limit, unique, danger signs, etc. A new approach is presented for marker-less AR-TSR system that superimposes augmented virtual objects onto a real scene under all types of driving situations including unfavorable weather conditions. Our method is composed of both online and offline stages. An intrinsic camera parameter change depending on the zoom values is calibrated. A Haar-like feature with Adaboost has been used to train a Haar detector in the offline stage. Extrinsic camera parameters are then estimated based on homography method in the online stage. With the complete set of camera parameters, virtual objects can be coherently inserted into the video sequence captured by the camera so that synthetic traffic signs may be added to increase safety.

Keywords: Augmented Reality, 3D, SURF, Camera Calibration, Traffic Signs, OpenCV.

1 Introduction

Improving traffic safety is one of the most important goals of Intelligent Transportation Systems (ITS). In such systems, a video camera is installed in the interior of the vehicle and the environment is scanned to provide useful information for the driver (such as obstacles, traffic lights, speed limit, etc.). In fact,

M. Kassab et al. (Eds.): Nets4Cars/Nets4Trains/Nets4Aircraft 2015, LNCS 9066, pp. 94–102, 2015.
DOI: 10.1007/978-3-319-17765-6_9

an increasing effort has been dedicated to developing driver assistance systems based on computer vision in the past years, aiming to reduce the number and fatality of traffic accidents and focusing on several aspects such as pedestrian detection [4],[6], lane detection/lane departure warning systems [11],[19], or traffic sign detection/recognition [13],[12]. Augmented Reality (AR) is a technique that combines live views in real-time with virtual computer generated images, creating an augmented experience of reality.

In cars, AR is becoming an interesting means to enhance active safety in the driving task. Guiding a drivers attention to an imminent danger somewhere around the car is a potential application. In current status of driving technology, it is an important issue that let drivers obtain driving information easily. There are many advanced electronic devices used for driving safety assistance such as automatic navigation system and lane departure warning system. In-vehicle driver assistance technologies, such as AR cuing, may help direct driver attention to roadway hazards [7],[8],[14], improve target detection [20], and reduce collision risk [10].

AR combines natural and artificial stimuli by projecting computer graphics on a transparent plane [1]. The graphical augmentation can highlight important roadway objects or regions or provide informative annotations. From this point of view, object detection and recognition from in-vehicle camera images have been widely studied, in particular for pedestrians [5], traffic signs [9], and other targets. There are two main problems to solve before the widespread usage of AR. First, we have the problem of rendering and merging virtual objects along with the high quality video stream. The second problem can be stated as finding a transformation of virtual scene into the one perceived by the human scene [17]. It requires that virtual objects are placed in the correct 3D position and orientation; and comply with the human eyes scale factor. For every frame of video stream, virtual objects have to be placed in a correct position and orientation and rendered again. This is a complex and time consuming process[15].

The use of AR in a vehicular context is becoming widely used and various projects are exploring this technology as a way to increase road safety and the overall driving experience. Traffic signs or road signs are an important part of road environment as they provide visual messages, not only for drivers, but also for all road users. Detection algorithms based on edge detection are more resilient to changes in illumination. The interaction of backgrounds and natural lighting can affect the usability of AR graphics (e.g., text legibility), a phenomenon that has been measured in studies such as [2] and [3].

Limited research has been conducted on the effectiveness of AR cuing for traffic signs detection. A handful of manufacturers offer a video-based AR auxiliary display to aid drivers in maneuvering the vehicle and identifying potential hazards while backing up. As the automotive industry moves toward the smart car, it makes driving safer, more pleasant, and more convenient. While we are already seeing some successful video-based AR auxiliary displays (e.g., center-mounted backup aid systems), the application opportunities of optical see-through AR as presented on a drivers windshield are yet to be fully tapped; nor are the related visual, perceptual, and attention challenges fully understood.

In this paper we present a marker-less tracking method which relies upon the use of features already available in the scene, without adding specific patches or markers. The system is responsible for detection and identification of traffic signs, speed limits, and any additional information that might be useful for drivers. Significant amount of work have addressed traffic sign recognition and four main factors have been identified in TSR: 1)segment The image and extract a Region Of Interest (ROI); 2)verify the traffic signs in the ROI based on a binary classification; 3)recognize multi-class traffic signs obtained from detection; and 4)track the traffic signs in the video sequences and add augmented information to the objects properly.

The reminder of this paper is organized as follows: in Section 2 we describe our approach referred to as Augmented Reality Traffic Signs Recongition (AR-TSR). In section 3 we provide experimental results and we finally conclude by providing some future work direction.

2 AR-TSR Approach

Due to the limitations of traditional marker AR system, we make full use of Marker-less information of video frames to make our system more natural and effective. We also need to take into account the visual system of the driver.

The approach proposed in this paper aims to contribute to AR applications on which virtual images can be merged with real images in real-time, without using ÃŕÂňducial markers, and considering the angle of view of the observer and the position of the object of interest. The first phase of the system deals with the detection of traffic Signs using a scanning window with Haar cascade detector for each image of the input stream (target images), which eliminates most of the "not-objects". The second phase makes the system more robust by verifying the detected traffic signs based on Speeded Up Robust Features (SURF) detector and by adding augmented information to the objects properly. Our method is composed of two stages: online and offline. In in both stages we assume that the intrinsic and distortion parameters of the camera are known and do not change. These two stages are detailed in the following Sections.

2.1 Offline Stage

The offline phase may be summarized in the following two steps:

Camera Calibration. Camera calibration is the first and the most important step in any computer vision application. Camera calibration is a necessary step in 3D computer vision in order to extract metric information from 2D images. An offline camera calibration is performed by use OpenCV camera calibration routines and a planar chessboard pattern to estimate the internal parameters of the camera and its distortion coefficients. Intrinsic parameters define the optical characteristics and the internal geometry of the camera.

Training of Haar-like Features. Haar-like features were originally proposed in the framework of object detection in face recognition. We train an AdaBoost cascade using Haar-like features offline. A boosting algorithm is used here to train a classifier with Haar-like features of positive and negative samples. The Adaboost Algorithm trains iteratively a strong classifier which is the sum of several weak classifiers. Each stage is trained so that the false positives of the previous stage are labeled as negatives and added to the training set. The positive images are those containing the object (e.g. traffic signs), and negatives are those which do not contain the object.

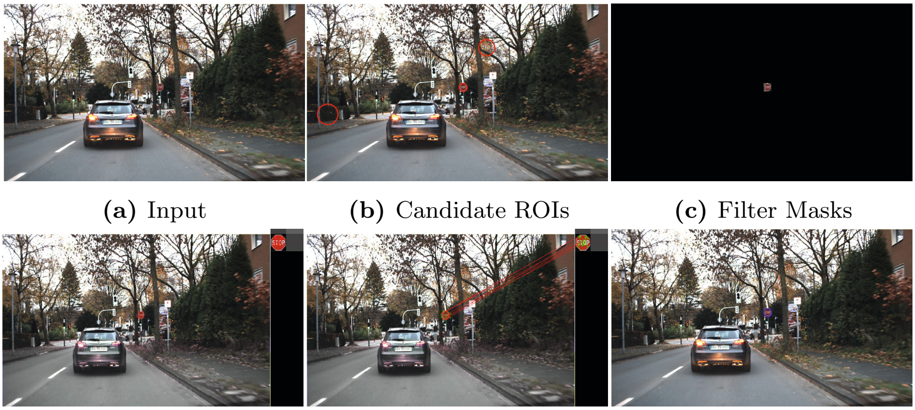
2.2 On-line Stage

Traffic Sign Detection. In the detection step, the distinctive features of traffic signs shall be considered. Since traffic signs are normalized in specific colors and shapes, it is convenient to use those features to decide the candidate signs. The goal of our detection stage is to identify image regions that may contain a traffic sign. To ensure a high system performance, we focus on fast detection methods. We have based our research on one of the most outstanding approaches in object detection for real-time applications of the last decade i.e., the Viola-Jones face detector framework [18].

The initial candidate detection phase of a traffic sign recognition system has much computational costs because candidates in a large range of scales have to be searched in the complete image. The Viola-Jones detector works by sliding a detection window across an image. At each position, the classifier decides if there is a desired object inside the window. During the detection phase, the system scans each window of the input image and extracts Haar features of that particular window, which is then used to compare with the cascade classifier. Finally, only a few of these sub-windows accepted by all stages of the detector are regarded as objects. The detection process takes an image as input and provides the regions that contain the ROI at the output, as illustrated in fig 1. Firstly, we roughly find out all candidate ROIs in a 24x24 sliding window. Secondly, the candidate ROIs are resized to 50x50 windows and further verified. In order to reduce the search space and improve detection performance, we apply the SURF to extract key-points from output of the first detector stage ROIs.

During the online tracking stage, after extracting and describing the SURF key-points, the system matches the current image key-points with the set of key-points from the reconstructed model acquired in the offline phase. Then a candidate image is matched by individually comparing each feature of the candidate with the special database and then the ROI is determined to be a true positive. The false positive rate is reduced significantly when the system is verified with part based SURF detector. The recognition module performs finer validations over the ROIs and generates the final detection results.

Pose Estimation and Augmentation. In marker-less AR, the problem of finding the camera pose requires significantly complex and sophisticated algorithms, e.g. disparity mapping, feature detection, and object classification. The



(a) Input **(b)** Candidate ROIs **(c)** Filter Masks

(d) Keypoint Extraction **(e)** keypoint Matching **(f)** True Positive ROIs

Fig. 1. A method for feature extraction of traffic sign detection

camera calibration allows the combination of virtual world and real world objects in a single display. In case of images or video, the relative position of an element on a screen can be calculated from camera parameters and relative position information of the camera with respect to the element.

After optimizing the matching process by RANSAC algorithm, a homography is applied to an ROI on an image and we apply the transformation of a planar object moving in the scene relative to a virtual camera. Firstly, the system takes four known points from the image of a scene and sets four separate tracking windows around the points. Based on these points, we calculate a camera calibration matrix. The mathematical model used is the projection transformation is expressed by (2), where λ is the homogeneous scale factors unknown a-priori, P is 3×4 projection matrix, $x = (x, y)$ represents the homogeneous coordinates of image features, $X = (X, Y, Z)$ represents the homogeneous coordinates of feature points in world coordinates, and $K \in R^{3\times3}$ is the matrix with the camera intrinsic parameters, also known as camera matrix. The joint rotation-translation matrix $[R|t]$ is the matrix of extrinsic parameters, $R = [r_x r_y r_z]$ is the 3×3 rotation matrix, and $T = [t]$ is the translation of the camera.

$$x = PX = K[R|t]X \tag{1}$$

$$\begin{aligned} P &= \overbrace{K}^{\text{Intrinsic Matrix}} * \overbrace{[R|t]}^{\text{Extrinsic Matrix}} \\ &= \overbrace{\underbrace{\begin{pmatrix} 1 & 0 & x_0 \\ 0 & 1 & y_0 \\ 0 & 0 & 1 \end{pmatrix}}_{\text{2D Translation}} * \underbrace{\begin{pmatrix} f_x & 0 & 0 \\ 0 & f_y & 0 \\ 0 & 0 & 1 \end{pmatrix}}_{\text{2D Scaling}} * \underbrace{\begin{pmatrix} 1 & s/f & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}}_{\text{2D Shear}}}^{\text{Intrinsic Matrix}} * \overbrace{\underbrace{(I|t)}_{\text{3D Translation}} * \underbrace{\left(\begin{array}{c|c} R & 0 \\ \hline 0 & 1 \end{array}\right)}_{\text{3D Rotation}}}^{\text{Extrinsic Matrix}} \end{aligned} \tag{2}$$

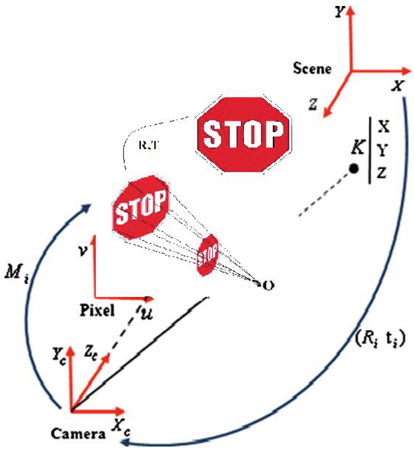


Fig. 2. Projection of Virtual Objects

In the final stage, the projection of virtual objects is accomplished once the pose is known. Having calculated the camera interior orientation and the camera exterior orientation for a video frame, the 3D can be drawn at the right position with the proper scale orientation and perspective in the scene of the real world, as shown in fig 2. With the complete set of camera parameters, virtual objects can be coherently inserted into the video sequence captured by the camera, so that synthetic traffic signs may be added to increase safety.

3 Experiment Results

To evaluate the performance of our AR-TSR method, we implement it using the hardware environment of Intel (R) Core (TM) i5 (2.5 Hz) and the software environment of Windows 7, Visual Studio 2010 using OpenGL and OpenCV Library. In this paper, we focus on the detection and recognition of speed limit, unique signs, and danger signs. The German Traffic Sign Recognition Benchmark (GTSDB) dataset [16] is used for verifying the effectiveness of our method.

3.1 Detection Performance

The first experiment, we separately evaluate the detection and classification modules. To evaluate the detector performance, we train a cascade of detectors using the evolutionary method with ordinal dissociated dipoles. The database used to train the detectors was collected from GTSRB dataset, the Belgian Traffic Signs Dataset (BelgiumTS), and our own images. In order to evaluate the systems robustness, we tested the accuracy of our algorithm when tracking the ROIs in the captured frames in various lighting and weather conditions, as shown in fig 3.

Fig. 3. Detection of traffic signs in adverse conditions

Fig. 4. Frames illustrating the insertion of virtual 3D object sign

We notice that the chances of missing true positives is comparatively less when compared with other systems. On the other hand, the false positive rate is reduced significantly when the system is tested with a SURF detector.

3.2 Recognition Performance

We evaluate the AR tracking by superimposing 3D graphics on target images. To provide driving-safety information using the proposed AR-TSR, various sensors and devices were attached to the experimental test vehicle, as shown in fig 4.

Our experiments demonstrate that the system can accurately superimpose virtual textures or 3D object to a user-selected planar part of a natural scene in real-time, under general motion conditions, without the need of markers or other artificial beacons.

4 Conclusions

In-vehicle contextual AR has the potential to provide novel visual feedback to drivers for an enhanced and more exciting driving experience. We have employed this approach to improve the accuracy of a traffic sign detector to assist the driver in various driving situations, increase the driving comfort, and reduce traffic accidents. We have demonstrated that AR can be very effective to enhance TSR. AR can be used to improve driving safety and minimize driving workload. The information provided is represented in such a way that it is easily understood, requiring little cognitive load on the driver. Further research needs to be conducted into the amount of information and the types of representations that can further minimize both driver's cognitive and visual loads, respectively.

Acknowledgments. This work was performed under the MOBIDOC device, part of the Support to Research and Innovation System project (PASRI), funded by the European Union (EU) and administered by the National Agency for Promotion of Scientific Research (ANPR).

References

1. Azuma, R., Baillot, Y., Behringer, R., Feiner, S., Julier, S., MacIntyre, B.: Recent advances in augmented reality. IEEE Computer Graphics and Applications 21(6), 34–47 (2001)
2. Gabbard, J.L., Edward Swan III, J., Hix, D.: The effects of text drawing styles, background textures, and natural lighting on text legibility in outdoor augmented reality. Presence: Teleoperators and Virtual Environments 15(1), 16–32 (2006)
3. Gabbard, J.L., Swan, J.E., Hix, D., Schulman, R.S., Lucas, J., Gupta, D.: An empirical user-based study of text drawing styles and outdoor background textures for augmented reality. In: Proceedings of IEEE Virtual Reality 2005, pp. 11–18. IEEE (2005)
4. Geronimo, D., Lopez, A.M., Sappa, A.D., Graf, T.: Survey of pedestrian detection for advanced driver assistance systems. IEEE Transactions on Pattern Analysis and Machine Intelligence 32(7), 1239–1258 (2010)
5. Guo, H., Zhao, F., Wang, W., Jiang, X.: Analyzing drivers attitude towards hud system using a stated preference survey. Advances in Mechanical Engineering 2014 (2014)

6. Guo, L., Ge, P.S., Zhang, M.H., Li, L.H., Zhao, Y.B.: Pedestrian detection for intelligent transportation systems combining adaboost algorithm and support vector machine. Expert Systems with Applications 39(4), 4274–4286 (2012)
7. Ho, C., Reed, N., Spence, C.: Multisensory in-car warning signals for collision avoidance. Human Factors: The Journal of the Human Factors and Ergonomics Society 49(6), 1107–1114 (2007)
8. Ho, C., Spence, C.: Assessing the effectiveness of various auditory cues in capturing a driver's visual attention. Journal of Experimental Psychology: Applied 11(3), 157 (2005)
9. Hussain, K., Kaptan, V.: Modeling and simulation with augmented reality. RAIRO-Operations Research 38(02), 89–103 (2004)
10. Kramer, A.F., Cassavaugh, N., Horrey, W.J., Becic, E., Mayhugh, J.L.: Influence of age and proximity warning devices on collision avoidance in simulated driving. Human Factors: The Journal of the Human Factors and Ergonomics Society 49(5), 935–949 (2007)
11. McCall, J.C., Trivedi, M.M.: Video-based lane estimation and tracking for driver assistance: survey, system, and evaluation. IEEE Transactions on Intelligent Transportation Systems 7(1), 20–37 (2006)
12. Mogelmose, A., Trivedi, M.M., Moeslund, T.B.: Vision-based traffic sign detection and analysis for intelligent driver assistance systems: Perspectives and survey. IEEE Transactions on Intelligent Transportation Systems 13(4), 1484–1497 (2012)
13. Park, J.-G., Kim, K.-J.: Design of a visual perception model with edge-adaptive gabor filter and support vector machine for traffic sign detection. Expert Systems with Applications 40(9), 3679–3687 (2013)
14. Scott, J.J., Gray, R.: A comparison of tactile, visual, and auditory warnings for rear-end collision prevention in simulated driving. Human Factors: The Journal of the Human Factors and Ergonomics Society 50(2), 264–275 (2008)
15. Sobel, D., Jędrasiak, K., Daniec, K., Wrona, J., Jurgaś, P., Nawrat, A.M.: Camera calibration for tracked vehicles augmented reality applications. In: Innovative Control Systems for Tracked Vehicle Platforms, pp. 147–162. Springer (2014)
16. Stallkamp, J., Schlipsing, M., Salmen, J., Igel, C.: The german traffic sign recognition benchmark: A multi-class classification competition. In: The 2011 International Joint Conference on Neural Networks (IJCNN), pp. 1453–1460. IEEE (2011)
17. Topór-Kamiñski, T., Krupanek, B., Homa, J.: Delays models of measurement and control data transmission network. In: Nawrat, A., Simek, K., Świerniak, A. (eds.) Advanced Technologies for Intelligent Systems. SCI, vol. 440, pp. 257–278. Springer, Heidelberg (2013)
18. Viola, P., Jones, M.J.: Robust real-time face detection. International Journal of Computer Vision 57(2), 137–154 (2004)
19. Wang, J.G., Lin, C.J., Chen, S.M.: Applying fuzzy method to vision-based lane detection and departure warning system. Expert Systems with Applications 37(1), 113–126 (2010)
20. Yeh, M., Wickens, C.D.: Display signaling in augmented reality: Effects of cue reliability and image realism on attention allocation and trust calibration. Human Factors: The Journal of the Human Factors and Ergonomics Society 43(3), 355–365 (2001)