SIViP (2018) 12:75–82
https://doi.org/10.1007/s11760-017-1132-5

ORIGINAL PAPER

In-vehicle augmented reality TSR to improve driving safety and enhance the driver's experience

Lotfi Abdi[1] · **Aref Meddeb**[2]

Received: 30 September 2016 / Revised: 15 May 2017 / Accepted: 6 June 2017 / Published online: 13 June 2017

Abstract In-vehicle contextual augmented reality (AR) has the potential to provide novel visual feedbacks to drivers for an enhanced driving experience. In this paper, we propose a new AR traffic sign recognition system (AR-TSR) to improve driving safety and enhance the driver's experience based on the Haar cascade and the Bag-of-Visual-Words approach, using spatial information to improve accuracy and an overview of studies related to the driver's perception and the effectiveness of the AR in improving driving safety. In the first step, the region of interest (ROI) is extracted using a scanning window with a Haar cascade detector and an AdaBoost classifier to reduce the computational region in the hypothesis generation step. Second, we proposed a new computationally efficient method to model global spatial distribution of visual words by taking into consideration the spatial relationships of its visual words. Finally, a multiclass sign classifier takes the positive ROIs and assigns a 3D traffic sign for each one using a linear SVM. Experimental results show that the suggested method could reach comparable performance of the state-of-the-art approaches with less computational complexity and shorter training time, and the AR-TSR more strongly impacts the allocation of visual attention during the decision-making phase.

✉ Lotfi Abdi
lotfiabdi@hotmail.com

Aref Meddeb
Aref.Meddeb@infcom.rnu.tn

[1] Networked Objects Control and Communication Systems Laboratory, National Engineering School of Tunis, University of Tunis El Manar, Tunis, Tunisia

[2] Networked Objects Control and Communication Systems Laboratory, National Engineering School Of Sousse, University of Sousse, Sousse, Tunisia

Keywords Augmented reality · Traffic sign recognition · SVM · Haar cascade · Bag-of-Visual-Words · Driving safety · Intelligent transportation systems

1 Introduction

An important social and economic problem in present days is traffic safety; according to several recent statistical estimates, road accidents have been among the top ten leading causes of death and have attributed to approximately 1.3 million deaths annually (WHO Report, 2012). Most traffic accidents have been caused by drivers oversight of important objects such as pedestrians, traffic signs, traffic signals, and so on. Research shows that human errors including driver inattention or cognitive overload lead to misjudgments and delays in environment recognition and constitute a major factor of road accidents. Even though some developments in passive safety technologies, such as seatbelts, airbags, crumple zones, etc., have partially reduced damages and improved safety during accidents, further progress in these technologies is limited due to their inherited limitations [15].

In-vehicle contextual augmented reality (AR) has the potential to provide novel visual feedback of other automate functionalities to drivers for an enhanced driving experience like traffic signs recognition, lane deviation warnings, safety distance indication and forward collision warnings. The AR-HUD technologies aim to optimize the visual attention of the driver by increasing the salience of high-value elements and to enhance the intelligent transportation systems by superimposing surrounding traffic information on the users' view and by keeping the drivers' view on roads. However, due to the existence of a complex environment such as weather conditions, illuminations and geometric distortions, the AR-HUD traffic sign recognition (TSR) systems have always

been considered as a challenging task. Although traffic signs are designed to be clearly visible, they can be missed due to driver distraction or sign masking.

There are several challenges involved in developing a complete TSR system that includes traffic sign detection and classification. These include occlusion of signs due to different background, weather condition, viewpoints and sign deformations. In order to achieve fast and robust TSR, designing a computing efficient and highly discriminative feature is essential. Besides, classification of traffic signs is a complicated matter, since sign types are similar. Recently, the Bag-of-Visual-Words (BoVW) has been frequently used in the classification of image data. There is a significant amount of work which present interesting advances for creating better dictionaries [12,22].

In the traditional BoVW model, spatial information between keypoints is ignored during visual words construction when using simple clustering algorithms such as *k*-mean. However, one major limitation of the standard BoVW model is that it ignores spatial information of visual words in image presentation and comparison. Researchers have demonstrated that the object recognition performance can be improved by including spatial information, which is important for similarity measurement between images [4,11,24]. Therefore, combining the frequency of occurrence and spatial information of visual words is a promising direction for improving classification accuracy.

In this paper, we present two key contributions. Firstly, in order to improve driving safety and enhance the driver's experience, we propose a new AR-TSR system that displays visual cues on the drivers' view while keeping drivers view on roads. We provide a prototype implementation of a visual AR system that significantly improves driving experience. Secondly, a novel approach for visual words construction is presented, which takes the spatial information of keypoints into account in order to enhance the quality of visual words generated from extracted keypoints. We demonstrate the complementarities of the additional relative spatial information provided by our approach to improve accuracy while maintaining short retrieval time.

2 Related work

Vehicular safety has been actively explored in the recent years. In fact, even before the appearance of motorized vehicles a lot of devices had been developed and placed in vehicles [16]. The design of TSR has been a challenge problem for many years and hence becoming an important and active research topic in the area of intelligent transport systems. Traffic sign localization and classification form a base for advanced methods used for accurate TSR and autonomous vehicle driving so that traffic accidents can be prevented and safety of traffic participants can be increased.

The most common approach, quite sensibly, consists of two main stages: detection and recognition. The considered baseline algorithms represent some of the most popular detection approaches such as the Viola–Jones detector, based on Haar-like feature [9], and the linear classifier relying on the histogram of oriented gradients (HOG) descriptors. Some recent methods, such as [10], have used the HOG features for road sign feature extraction, using complementary features to reduce the computation complexity of TSD, and then using the SVM to implement the traffic sign classification.

Moreover, the convolutional neural network (CNN) has been adopted in object recognition for its high accuracy. In [14], they applied convolutional networks (ConvNets) to the task of traffic sign classification. The ConvNets were biologically inspired multistage architectures that automatically learned hierarchies of invariant features. The CNNs consist of a multistage processing of an input image to extract hierarchical and high-level feature representations. In [5], a real-time system for traffic signs was put forward, which used a sliding window method combining various DNNs trained on differently preprocessed data into a multicolumn DNN (MCDNN).

The above-described approach ignores structural information of features, which is important for similarity measurement between images. It is therefore necessary to classify the characteristics of the given information and find a way to represent the information according to these characteristics. Several methods were recently proposed to incorporate spatial information to improve the BoVW model such as the spatial pyramid matching method [13], spatiotemporal interest point [6] and the distance between joint histograms to measure the similarity between a target and its candidate patches [20]. Considering the processing time and classification accuracy as a whole, we have developed a novel technique to incorporate spatial information of visual words to improve accuracy while maintaining short retrieval time.

3 Augmented reality traffic signs recognition

The vision algorithms for driver assistance systems usually need to fulfill strong real-time constraints. Hence, we draw a particular focus on real-time capability of the algorithms evaluated here. Our detector is inspired by a detector presented by Viola and Jones [21]. In the first step, the ROI is extracted using a scanning window with a Haar cascade detector and an AdaBoost classifier to reduce the computational region in the hypothesis generation step. Next in the verification phase, to confirm whether each ROI is traffic sign or not, a second stage is needed to eliminate some false positives. In this stage, with the feature extraction of traffic signs based on the speeded up robust features (SURF), the code-

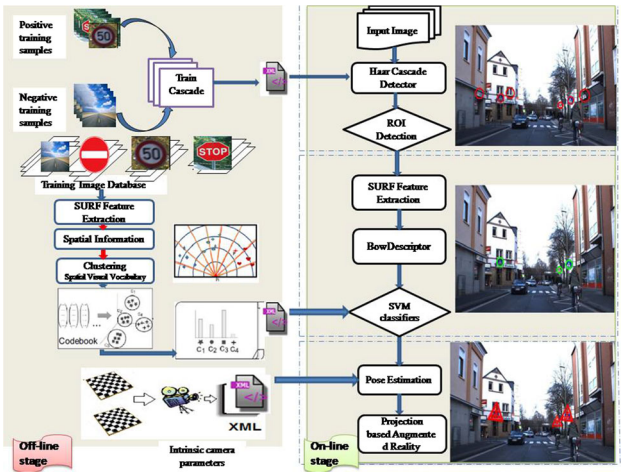

Fig. 1 Overview of the AR-TSR application

book is generated by these feature clustering and the images are described by histograms using the BoVW for verification. To ensure rotation invariance, we proposed a new computationally efficient method to model global spatial distribution of visual words and improved the standard BoVW representation, by taking into consideration the spatial relationships of its visual words. Finally, a multiclass sign classifier takes the positive ROIs and assigns a 3D traffic sign for each one using a linear SVM.

Figure 1 shows the overall procedure of the marker-less AR system, which is split up in two distinct stages. In the above two stages, we assume that the intrinsic and distortion parameters of the camera are known and do not change; these two stages are detailed in [1].

3.1 Generation candidate detection bounding boxes

The initial detection phase of a TSR system has much computational costs because ROIs in a large range of scales have to be searched in the complete image. In order to reduce the search space, the adopted solution is to combine a cascade with fewer stages with other methods that eliminate the false positives. During the detection phase, the system scans each window of the input image and extracts the Haar-like features of that particular window, which is then used to compare to 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 an input and gives at the output the regions that contain the ROI. The false alarm rate of the Haar cascade detector, without a hypothesis verification, is higher, but it eliminates most of the non-object regions.

The Haar-like features were originally proposed in the framework of object detection in the face detection approach. An AdaBoost cascade using Haar-like features is trained offline and a boosting algorithm is used to train a classifier with the Haar-like features of positive and negative samples. The AdaBoost algorithm trains iteratively a strong classifier, which is the sum of several weak classifiers. The object is classified positively only if it is positively classified in each cascade stage. The final classifier works in real time. In fact, from an integral image, in a classifier produced by AdaBoost, voting is done as a summation of weighted classifiers. On average, only a small subset of classifiers vote positively because of cascading.

The real-time capability of the approach is mainly enabled by two properties: Most sliding windows are only evaluated by the first stages which contain few classifiers/features [8]. To reduce the false alarm rate, the detected traffic sign output of the detector stage is processed with a part based on a verification module.

3.2 Verification system based on BoVW

In the traditional BoVW model, spatial information between keypoints is ignored during visual words construction when using simple clustering algorithms such as k-mean. However, this modeling approach does not take into consideration the spatial relationships of these words, which is important for similarity measurement between images. To solve this challenging task, recent approaches try to capture information about the relative spatial location of visual words. This paper presents a new approach to integrate the spatial information to BoVW model, with explicit local and global structure models.

To address this issue, we introduce a novel way to incorporate both distance and angle information in the BoVW representation. This method exploits spatial orientations and distances of all pairs of similar descriptors in the image. In the BoVWs model, a visual vocabulary $\text{Voc} = v_i, i = \{1, \ldots, k\}$, then it is built by clustering these features into a certain number of K visual words. A given descriptor d_k is then mapped to a visual word v using euclidean distance in Eq. (1) as follows:

$$v(d_k) = \text{argminDist}(v, d_k) \tag{1}$$

where $v \in \text{Voc}$, d_k is the kth descriptor in the ROI, $\text{Dist}(v, d_k)$ is the distance between the descriptor and the visual word based on the euclidean distance. For this reason, we consider the weighted sum of ROIs to implicitly represent spatial information which is important for similarity measurement between images.

In the training stage, the SURF features are extracted from all the training samples, using a dense grid. Since we are interested in the sign contents, only the descriptors that do not fall outside the sign contour are taken into account. Our system exploits the SURF features, which have shown a high robust-

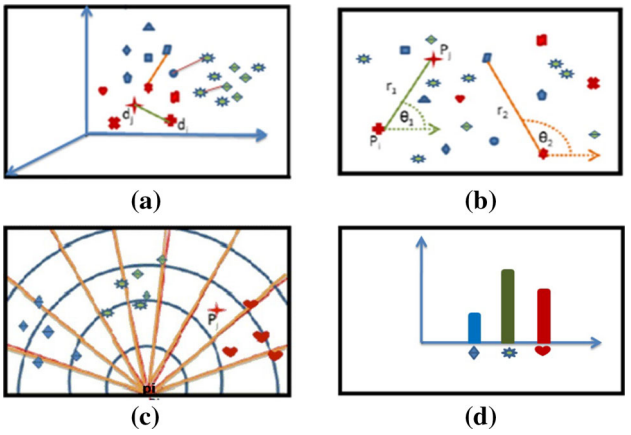

Fig. 2 Spatial histogram of similar pairwise using distance and orientation: **a** spatial distance of similar pairwise, **b** spatial orientations of similar pairwise, **c** pairwise similarity distance orientations information of similar pairwise, **d** pairwise spatial histograms

ness to varied recording conditions. After the SURF features are extracted for all the training samples, the number of feature points of each image is not entirely consistent, which will bring great difficulties to subsequent operations. Assignment of a visual feature to the vocabulary depends on the similarity metric. We propose a method that incorporates spatial information at feature level. We measure the spatial relationships between visual words using distance and orientation.

For each visual word, the average position and the standard deviation is computed based on all the occurrences of the visual word in the image. We consider the interaction between visual words by encoding their spatial distances, orientations and alignments. Figure 2 shows an example to better understand our approach. To encode spatial information, we use the distance (2a) and orientation (2b) information between pairs of patches in the image space.

More formally, we consider the set S_k of all the pairwise, where at least one patch in the pair belongs to the visual word w_k. A given pair $(P_i, P_j) \in S_k$ is characterized both by a pair of descriptors (d_i, d_j) and a pair of positions in the image space denoted (p_i, p_j) is illustrated in Fig. 2. Note that both d_i and p_i are vectors with $d_i \in R^D$ and $p_i \in R$.

Then, for each pair of points of the feature, we compute the angle θ formed with the horizontal axis using Eq. (2) above:

$$\theta = \begin{cases} \arccos \dfrac{\overrightarrow{P_i P_j}.\vec{u}}{\left\| \overrightarrow{P_i \ P_j} \right\|}, & \text{if } \overrightarrow{P_i P_j}.\vec{v} > 0 \\ \pi\text{-}\arccos \dfrac{\overrightarrow{P_i P_j}.\vec{u}}{\left\| \overrightarrow{P_i \ P_j} \right\|} & \text{Otherwise} \end{cases} \tag{2}$$

where $\overrightarrow{P_i P_j}$ is the vector formed by two points P_i, P_j, and u, v are orthogonal unit vectors defining the image plane. After clustering, the spatial information is implicitly included in the visual vocabulary. A pairwise spatial histogram (2d) of similar patches is then defined considering a discretization of the image space into M bins denoted $b_m, m = \{1, \ldots, M\}$ with an angle $\theta \in [0, \pi[$ split into M_θ angle bins and the radius $r \in [0, R]$ split into M_r radial bins so that $M = M_\theta \cdot M_r$.

For those purpose, a novel structural relationship between patches are defined for evaluating superpixels similarity. In this paper, the simple linear iterative clustering (SLIC) superpixels [2] are used as an adaptive analysis window for extracting spatial features. The SLIC method is chosen, because it produces high-quality superpixels and is simple to implement. We show that the choice of interest point detector is crucial in the Bag-of-Visual-Words approach, and the distance measured between pixels and the superpixel centers is the key issue of the SLIC algorithm. In the proposed method, spatial information derived from superpixels is utilized to improve the performance of classification. It generates superpixels by grouping pixels with a local k-means clustering method, where the distance is measured as the Euclidean distance integrated with the data and spatial distances.

Particularly, simple spatial relations between visual words are considered the spatial locations of the words and the spatial relationship between the words were added to describe images in the BoW model. This histogram encodes spatial information [distance and orientation (2d)] of pairwise similar patches, where at least one of the patches belongs to V_k. To have a global representation, we replace each bin of the BoVW frequency histogram with the spatial histogram associated to w_i. By this way, we keep the frequency information intact and add the spatial information.

3.3 Pose estimation and augmentation

The key to realize a AR 3D registration is to obtain a camera projection matrix, which represents the relationship between the 2D points from the image and the 3D points from the model. The geometric relationship between 3D world lines and their projections on the camera image are built to estimate the relative 6-DOF camera pose consists of rotation parameters and translation parameters [3]. From the planar homography, we can easily compute the camera position and rotation, which provides the motion estimates. The used mathematical model is the projection transformation, which is expressed by Eq. (3) where λ is the homogeneous scale factors unknown a priori, where P is a 3×4 projection matrix, $x = (x, y)$ are the homogeneous coordinates of the image features, $X = (X, Y, Z)$ are the homogeneous coordinates of the feature points in the world coordinates, $K \in R^{3\times3}$ 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.

$$x = \lambda P X = K[R|t]X \tag{3}$$

The projection matrix P is the key to creating a realistic augmented scene using the intrinsic parameters of the camera, the dimensions of the video frame and the distances of the near and far clipping planes from the projection center. In our method, we assume that the intrinsic parameters are known in advance and do not change, and this is reasonable in most cases.

$$P = \overbrace{K}^{\text{Intrinsic matrix}} * \overbrace{[R|t]}^{\text{Extrinsic matrix}} = \overbrace{\underbrace{\begin{pmatrix} 1 & 0 & x_0 \\ 0 & 1 & y_0 \\ 0 & 0 & 1 \end{pmatrix}}_{\text{2D translation}} * \underbrace{\begin{pmatrix} f_x & 0 & 0 \\ 0 & f_y & 0 \\ 0 & 0 & 1 \end{pmatrix}}_{\text{2D scaling}} * \underbrace{\begin{pmatrix} 1 & s/f & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}}_{\text{2D shear}}}^{\text{Intrinsic matrix}} * \overbrace{\underbrace{(I|t)}_{\text{3D translation}} * \underbrace{\left(\begin{array}{c|c} R & 0 \\ \hline 0 & 1 \end{array}\right)}_{\text{3D rotation}}}^{\text{Extrinsic matrix}} \tag{4}$$

Once K is known, the extrinsic parameters for each image are readily computed. From Eq. (3), we have:

$$\begin{array}{l} r1 = \lambda + K^{-1}h_1 \\ r2 = \lambda + K^{-1}h_2 \\ r3 = r1 * r2 \\ t = \lambda + K^{-1}h_3 \end{array} \quad where \begin{cases} h_1 = [h_{11}\ h_{21}\ h_{31}]^{\mathrm{T}} \\ h_2 = [h_{12}\ h_{22}\ h_{32}]^{\mathrm{T}} \\ h_3 = [h_{13}\ h_{23}\ h_{33}]^{\mathrm{T}} \\ r_1 = [r_{11}\ r_{21}\ r_{31}]^{\mathrm{T}} \\ r_2 = [r_{12}\ r_{22}\ r_{32}]^{\mathrm{T}} \\ r_3 = [r_{13}\ r_{23}\ r_{33}]^{\mathrm{T}} \end{cases} \quad \begin{array}{l} Where\ H = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \\ and\ \lambda = \frac{1}{\|K^{-1}+h_1\|} \end{array} \tag{5}$$

In order to integrate virtual objects into the real-world seamlessly, the AR system must be able to recognize and track its desired environment. In this final stage, the projection of virtual objects will be easily accomplished once the pose is known. Having calculated the camera's interior and exterior orientations 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. 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.

The projection-based AR corresponds to the use of projection technology to augment and enhance 3D objects and spaces in the real world by projecting images onto their visible surfaces. Once there are enough successful matches, a RANSAC method is applied to calculate the homography matrix between the image of the frame and the image of the object. Then, we are able to estimate the 3D pose and draw a virtual 3D object on the top of the real object. The camera calibration allows combining virtual and real-world objects in a single display.

To correctly model the perspective projection of the camera, we must mimic the intrinsic camera parameters in the virtual environment. When we have the camera calibrated in a frame, we can synchronize the real camera with a virtual camera and project the virtual objects onto the real image using OpenGL. Technically, this can be described with a projection matrix that maps 3D points onto a 2D plane. After the world has been aligned with the camera using the view transformation, the conversion from an intrinsic matrix to the model view and projection matrices requires a conversion from the world coordinates to the normalized view volume coordinates used by OpenGL. The perspective projection matrix is expressed by Eq. (6), where width, height, far, near represent the positions of the clipping planes.

$$\begin{bmatrix} x_{\text{clip}} \\ y_{\text{clip}} \\ z_{\text{clip}} \\ w_{\text{clip}} \end{bmatrix} = \begin{bmatrix} \frac{2*c_x}{\text{width}} & 0 & 1-\frac{2*x_0}{\text{width}} & 0 \\ 0 & \frac{2*c_y}{\text{height}} & -1+\frac{2*y_0}{\text{height}} & 0 \\ 0 & 0 & \frac{\text{near}+\text{far}}{\text{near}-\text{far}} & -2*\frac{\text{near}*\text{far}}{\text{near}-\text{far}} \\ 0 & 0 & -1 & 0 \end{bmatrix} \begin{bmatrix} X_{\text{camera}} \\ Y_{\text{camera}} \\ Z_{\text{camera}} \\ 1 \end{bmatrix} \tag{6}$$

As it is indicated, most of current marker-less tracking approaches require a 3D model of the environment for matching 2D features to those lying on the model. In addition to the complexity of building a model, such a strategy will result in performance problems when the model is very complex or the environment is dynamic. In contrast, our approach does not need to perform 3D engineering of the environ-

Table 1 Recall and precision results for traffic sign detection

Traffic signs	Number of signs	TP	Recall (%)	Precision (%)
Speed limit	127	126	99.21	98.43
Danger signs	79	75	97.46	98.97
Unique signs	50	49	98	99.04
Mandatory signs	135	134	99.14	99.25
Derestriction signs	120	117	99.02	98.96
Other prohibitory signs	115	115	99.15	99.08

ment. Also, we use a simple virtual 3D model, with a known size, to define a reference coordinate system. This stage is composed of a feature tracker that finds point matches, and a homography-based method can be applied to find the rotation and translation of the camera. Finally, the registration matrix is calculated using the above homography, and the virtual objects are rendered on the real scenes using OpenGL.

4 Experiment results

To evaluate the performance of the proposed algorithm, we implement the proposed AR-TSR method using the hardware environment of Core i7 640LM 2.13 GHz and the software environment of Windows 7, Visual Studio 2010 using OpenGL and OpenCV Library 2.4.8. We implement the suggested method in C++ and test the real-time performance on the German Traffic Sign Recognition Benchmark (GTSRB) and German Traffic Sign Detection Benchmark (GTSDB) datasets [17]. These classes of traffic signs have been divided into six subsets speed limit sign subset, danger sign subset, mandatory sign subset, unique sign subset, derestriction sign subset and other prohibitory sign subset.

4.1 Performance of the proposed method

4.1.1 Detection performance

The database used to train the detectors has been collected from the GTSRB dataset, the Belgian Traffic Signs Dataset (BelgiumTS) [19], and our own images. Our training dataset consists of 4500 interest traffic signs and 6000 non-traffic signs. The sizes of traffic sign examples are in range from 15×15 to 250×250 pixels. The achieved detection performances are summarized in Table 1 versus the number of test images.

The experimental results of Table 1 demonstrate an excellent performance of our system. The results show that the proposed algorithm attains an average precision rate of 98.95% and an average recall rate of 98.66%. As previously mentioned, the detection system robustness is demonstrated through its tolerance of changes in lighting and in plane rotations. In order to evaluate the system robustness, we have 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

The missing chances of true positives are comparatively less when compared with other systems. The false alarm rate is reduced greatly when the system is tested with a part-based BoVW verification. It has been proved by experiments that our algorithm is not only highly efficient, but also more accurate than previous algorithm during detection.

4.1.2 Classification performance

In the classification stage, we determine whether a detected image region contains a particular traffic sign or whether it has to be rejected as a false positive. In order to evaluate the occlusion robustness of the suggested classification method, the content of the detected ROI is identified using the tree classifiers. This classifier is tested on static, low-resolution sign images. A comprehensive performance evaluationon GTSRB dataset is carried out, where Table 2 shows the classification rates of the linear SVM.

Table 2 Confusion matrices of traffic sign classification

	Speed limit	Danger	Unique	Mandatory	Derestriction	Prohibitory
Speed limit	0.998	0	0.001	0	0	0.001
Danger	0.001	0.993	0.004	0	0.002	0
Unique	0	0	1	0	0	0
Mandatory	0	0	0	0.999	0	0.001
Derestriction	0	0.001	0	0.001	0.998	0
Prohibitory	0.001	0	0	0	0	0.999

Table 3 Performance comparison with other TSR methods

Method	[5] (%)	[18] (%)	[14] (%)	[23] (%)	[7] (%)	Our (%)
Speed limit	99.47	97.63	98.61	95.95	98.82	99.13
Danger signs	99.07	98.67	98.03	92.08	96.85	98.97
Unique signs	99.22	100.00	98.63	98.73	100.00	99.51
Mandatory signs	99.89	99.72	97.18	99.27	96.86	99.45
Derestriction signs	99.72	98.89	94.44	87.50	97.93	99.32
Other prohibitory signs	99.93	99.93	99.87	99.13	98.27	99.47

A key idea of our method is to project the 3D object sign using the corresponding sparse dictionary and then to classify the projected vector with the SVM. Furthermore, we evaluate the classification task on the detected signs returned by the previous detection module. As shown in Table 2, the overall classification accuracy is 99.31%. Note that only 3 (out of 1500) speed limit signs, while only 6 (out of 890) danger signs, are falsely classified. If the recognition is complete, a multiclass sign classifier takes the positive ROI and assigns a 3D traffic sign to each one. Experiments demonstrate that our approach succeeds in adding relative spatial information into the BoVW model by encoding both the global and local relative distributions of visual words over an image.

4.2 Comparisons with other state-of-the-art methods

In order to verify the discrimination performance and computation efficiency of the proposed feature for TSD, the experiments on the public available dataset of traffic signs are implemented. Because the training and testing samples in the GTSRB dataset are split according to a fixed rule, an absolute performance comparison with other reported approaches is possible. We report these results in Table 3, where the results of the winning system from the IJCNN challenge and some reported results in the IJCNN 2011 are provided as references.

According to the results for the GTSRB dataset, shown in Table 3, this work achieves a 99.31% recognition accuracy, which is a comparable performance of 0.24% less then the work by [5], and a performance of 0.17% higher than the work by [18] and 1.51% than the work by [14]. The accuracy of recognizing unique signs reach 99.31%, which is comparable with the best achieved one. The danger signs which have triangular shape have given the worst results compared with other traffic sign categories. Compared with other methods, this paper presents an overview of studies related to drivers' perception and cognition when this information is displayed on the windshield HUD, as it can be a solution to reduce the duration and frequency of drivers looking away from the traffic scene, which is very important in safe driving assistance systems.

4.3 Augmented reality tracking

In this section, the results obtained during real-time tests, performed with a fully equipped vehicle, are presented. We have started the evaluation of the AR tracking by superimposing 3D graphics on target images. To provide driving safety information using the proposed AR-TSR, various sensors and devices have been attached to the experimental test vehicle. The system has been empirically tested under different lighting conditions, in sunny or cloudy days, in the rain and at night (Fig. 4).

The experimental results have shown that the proposed method has significantly reduced the computational cost and also stabilized the camera pose estimation process. A virtual object is attached to a real object for the augmentation purpose, and the camera pose are used to superimpose virtual objects onto the real environment. Therefore, the AR-HUD is an important step in the direction of holistic human machine interaction concepts in vehicles for a more comfortable, more economic and safer driving experience. The experiments have confirmed that the system can accurately superimpose virtual textures or 3D objects to a user-selected planar part of a natural scene in real time, under general motion

Fig. 4 Insertion of virtual 3D object sign in cloudy days, nighttime, sunny days and snow days

conditions, without the need of markers or other artificial beacons.

5 Conclusions

To improve driving safety and minimize the driving workload, the provided information should be represented in such a way that it is more easily understood and imposing less cognitive load onto the driver. A new AR-HUD approach to create real-time interactive traffic animations is introduced, in terms of rules for placement and visibility, types of traffic signs and migration of these to an in-vehicle display. The AR-TSR supplements the exterior view of the traffic conditions in front of the vehicle with virtual information for the driver. We have chosen to combine the Haar cascade detector and hypothesis verification using BoVW with the relative spatial information between visual words, which has proved to be a good compromise between the resource efficiency and overall performance. Experimental results show that the suggested method could reach comparable performance of the state-of-the-art approaches with less computational complexity and shorter training time.

References

1. Abdi, L., Meddeb, A., Abdallah, F.B.: Augmented reality based traffic sign recognition for improved driving safety. In: International Workshop on Communication Technologies for Vehicles, pp. 94–102. Springer, Berlin (2015)
2. Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Süsstrunk, S.: Slic Superpixels. Tech. rep. (2010)
3. Amamra, A., Aouf, N., Stuart, D., Richardson, M.: A recursive robust filtering approach for 3d registration. Signal Image Video Process. **10**(5), 835–842 (2016)
4. Chen, C., Zhang, B., Su, H., Li, W., Wang, L.: Land-use scene classification using multi-scale completed local binary patterns. Signal Image Video Process. **10**(4), 745–752 (2016)
5. Cireşan, D., Meier, U., Masci, J., Schmidhuber, J.: Multi-column deep neural network for traffic sign classification. Neural Netw. **32**, 333–338 (2012)
6. Dawn, D.D., Shaikh, S.H.: A comprehensive survey of human action recognition with spatio-temporal interest point (stip) detector. Vis. Comput. **32**:1–18 (2015)
7. Haloi, M.: A Novel PLSA Based Traffic Signs Classification System. arXiv:1503.06643 (2015)
8. Houben, S., Stallkamp, J., Salmen, J., Schlipsing, M., Igel, C.: Detection of traffic signs in real-world images: the German traffic sign detection benchmark. In: The 2013 International Joint Conference on Neural Networks (IJCNN), pp. 1–8. IEEE (2013)
9. Jiang, S., Ning, J., Cai, C., Li, Y.: Robust struck tracker via color Haar-like feature and selective updating. Signal Image Video Process. 1–8 (2017)
10. Kun, Z., Wenpeng, W., Guangmin, S.: An effective recognition method for road information based on mobile terminal. Math. Probl. Eng. **2014**:1–8(2014)
11. Li, Y., Xu, J., Zhang, Y., Zhang, C., Yin, H., Lu, H.: Image classification using spatial difference descriptor under spatial pyramid matching framework. In: International Conference on Multimedia Modeling, pp. 527–539. Springer, Berlin (2016)
12. Nguyen, K.D., Le, D.D., Duong, D.A.: Efficient traffic sign detection using bag of visual words and multi-scales sift. In: International Conference on Neural Information Processing, pp. 433–441. Springer, Berlin (2013)
13. Ren, Y.: A comparative study of irregular pyramid matching in bag-of-bags of words model for image retrieval. Signal Image Video Process. **10**(3), 471–478 (2016)
14. Sermanet, P., LeCun, Y.: Traffic sign recognition with multi-scale convolutional networks. In: The 2011 International Joint Conference on Neural Networks (IJCNN), pp. 2809–2813. IEEE (2011)
15. Shahid, M., Nawaz, T., Habib, H.A.: Eye-gaze and augmented reality framework for driver assistance. Life Sci. J. **10**(3):1–8 (2013)
16. Silvéria, M.K.: Virtual Windshields: Merging Reality and Digital Content to Improve the Driving Experience. arXiv:1405.0910 (2014)
17. 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)
18. Stallkamp, J., Schlipsing, M., Salmen, J., Igel, C.: Man vs. computer: benchmarking machine learning algorithms for traffic sign recognition. Neural Netw. **32**, 323–332 (2012)
19. Timofte, R.: Kul Belgium Traffic Signs and Classification Benchmark Datasets. http://btsd.ethz.ch/shareddata
20. Uzyıldırım, F.E., Özuysal, M.: Instance detection by keypoint matching beyond the nearest neighbor. Signal Image Video Process. **10**(8), 1527–1534 (2016)
21. Viola, P., Jones, M.J.: Robust real-time face detection. Int. J. Comput. Vis. **57**(2), 137–154 (2004)
22. Virupakshappa, K., Han, Y., Oruklu, E.: Traffic sign recognition based on prevailing bag of visual words representation on feature descriptors. In: 2015 IEEE International Conference on Electro/Information Technology (EIT), pp. 489–493. IEEE (2015)
23. Zaklouta, F., Stanciulescu, B., Hamdoun, O.: Traffic sign classification using KD trees and random forests. In: The 2011 International Joint Conference on Neural Networks (IJCNN), pp. 2151–2155. IEEE (2011)
24. Zhu, Q., Zhong, Y., Zhao, B., Xia, G.S., Zhang, L.: Bag-of-visual-words scene classifier with local and global features for high spatial resolution remote sensing imagery. IEEE Geosci. Remote Sens. Lett. **13**(6), 747–751 (2016)