Efficient Detection and Tracking of Road Signs Based on Vehicle Motion and Stereo Vision

Chang-Won Choi, Sung-In Choi, and Soon-Yong Park

School of Computer Science and Engineering
Kyungpook National University, Daegu, Republic of Korea
choi408@vision.knu.ac.kr, ellim5th@naver.com, sypark@knu.ac.kr

Abstract. The road signs provide important information about road and traffic to drivers for safety driving. These signs include not only common traffic signs but also the information about unexpected obstacles and road constructions. Accurate detection and identification of road signs is one of the research topics in vehicle vision area. In this paper we propose a stereo vision technique to automatically detect and track road signs in a video sequence which is acquired from a stereo vision camera mounted on a vehicle. First, color information is used to initially detect the candidates of road signs. Second, the Support Vector Machine (SVM) is used to select true signs from the candidates. Once a road sign is detected in a video frame, it is tacked from the next frame until disappeared. The 2-D position of the detected sign on the next frame is predicted by the motion of the vehicle. Here, the vehicle motion means the 3-D Euclidean motion acquired by using a stereo matching method. Finally, the predicted 2-D position of the sign is corrected by the template matching of a scaled sign template in the near regions of the predicted position. Experimental results show that the proposed method can detect and track road signs successfully. Error comparisons with two different detection and tracking methods are shown.

Keywords: stereo, traffic sign, detection, tracking, motion.

1 Introduction

With the emergence of the vehicle vision many smart technologies such as forward and backward obstacle detection, navigation systems, and unmanned vehicle driving have many research attention. Despite the development of automobile technologies, still the driver's negligence causes many accidents on the road. The road signs are very simple but give very important information about the road condition and dangerous situation to the drivers and pedestrians to avoid accidents. If the technology can be used to automatically detect and recognize road signs we can avoid most of these accidents caused by the negligence of the drivers. In this research we introduce an automatic road sign detection and tracking technique to provide safety information to vehicle driver.

Many researches have been studied to detect and recognize road signs in various road and environment conditions. Some of them have used color information to detect

The original version of this chapter was revised: The copyright line was incorrect. This has been corrected. The Erratum to this chapter is available at DOI: 10.1007/978-3-319-02895-8_64

J. Blanc-Talon et al. (Eds.): ACIVS 2013, LNCS 8192, pp. 608–619, 2013.

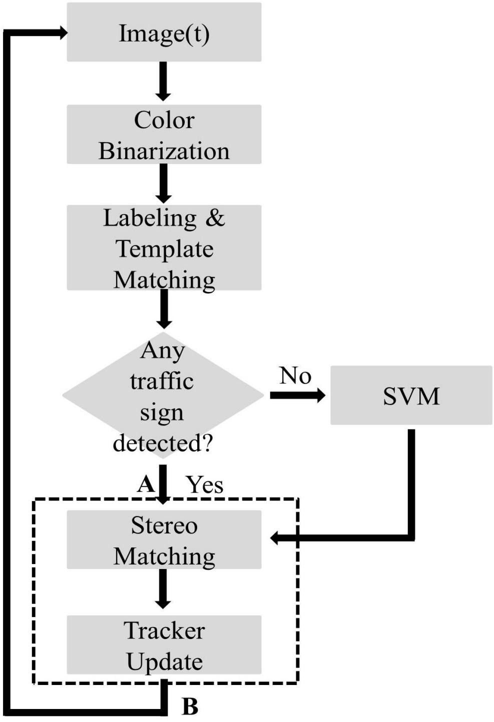


Fig. 1. A block diagram of the proposed algorithm

the road signs or machine learning techniques to recognize the signs [1][2][3]. However, when some obstacles occlude the road signs or color information is not clear with the weather changes, it is difficult to detect the road signs correctly. Some methods have used the edges and color features to detect the traffic signs based on neural networks [4]. In order to track the road signs, they assumed that the speed of the vehicle and the physical sizes of the road signs are known. Then the road signs in the next frame are detected using this information. Some investigations have been introduced to make use of the transformation matrix between the camera coordinate system and the object coordinate system to find the 2D-3D point correspondences for detecting traffic signs [5][6]. In recent researches, various methods have been used together for detecting and recognizing the road signs more accurately and robustly. In [7], color information, cross-correlation function, and PCA (Principal Component Analysis) are used to detect the road signs and a binary tree is used to recognize them.

In this paper, we propose a road sign detection and tracking technique using the 3D information from a stereo vision camera mounted on a moving vehicle. Color information and the SVM (Support Vector Machine) are used to detect road signs. The 3-D position of a detected sign is provided by the depth map calculated by the stereo vision camera. Then the 3-D position information is projected to the image space of the next from to track the sign. By applying this sign tracking algorithm based on the information found in the previous frame, the proposed method is more robust to the environmental changes. As the first step of the proposed sign detection and tracking technique, we classify color objects in a video frame using a look-up table in HSI color space. Then we run labeling and canny edge extraction algorithms to find the candidates of traffic sign objects. After founding candidates, we perform a template matching to filler out some erroneous traffic signs from these candidates and detect true signs sing SVM [8]. Once a road sign is detected, its 3-D position is obtained from the depth map of the current frame as mentioned above. Then the 3-D position is projected to the 2-D image plane of the next frame to predict the road sign. The predicted road sign is refined by a template matching of the sign. When doing the template matching, the template of the sign in the previous frame is scaled according to the actual 3-D size of the sign. This scaled template increases the tracking performance. The tracking algorithm runs iteratively until the sign disappear from the sequence.

2 Road Signs Detection and Tracking

Road sign detection from an image sequence is the first step of the proposed technique which is followed by road sign tracking. A stereo camera is mounted on a moving vehicle to acquire the image sequences. Fig. 1 shows the overall processes of the proposed tracking technique. First, in every video frame, RGB color model is converted to HSI color model and the hue component is converted to a binary image by a threshold value [8]. Second, initial sign candidates are obtained by labeling and matching with sign templates of triangle, circle and rectangle, and etc. Third, using the learning data of SVM, we decide correct signs from the candidates. At the same time, the 3-D positions of detected signs are computed from the depth map of the current frame, which is obtained by the SGBM (Semi Global Block Matching) stereo matching method. Fourth, the 3-D positions of the signs are projected to the image plane of the next frame using the PPM (Perspective Projection Matrix) of the stereo camera. The sin tracking algorithm in the next frame will be described later in another section.

2.1 Image Threshold Using Color Information

To detect road signs in a video frame, we use the color information of the road signs. In general, the color of road sign composed of red, blue, yellow, and white. To extract the candidates of sign, first we convert every image frame to HIS color image. In order to detect road objects such as temporary construction signs and general traffic signs, the hue image is converted to a binary image using Equation 1 which divides color value as achromatic and chromatic values. The reason we use Equation 1 is that most road signs is chromatic. In Equation 1, the range of R,G and B is 0 to 255. If f is greater than 1, it is regarded as the chromatic and the remainder is filtered out considering it as achromatic. The total range of hue value is 0° ~ 360°, and we use hue value of 330° ~ 360° and 0° ~ 40°.

$$f = \frac{|R-G| + |G-B| + |B-R|}{60} \tag{1}$$

2.2 Labeling and Template Matching

A group of sign candidates is decided in the binarized video frame through the labeling process. When there is a binary image, the labeling process is to give identification to each object group. Therefore, different groups of objects have different labels. Also, If the two road signs is connected, we must divide these. We can distinguish the two road signs using canny edge extraction algorithms. To decide whether a given labeled object is a road sign or not, we match them with the standard shape of road signs, circles, triangles, inverted triangles and etc. After the image of candidate is scaled same with the template image size, the template matching method measures brightness difference between the template and candidates. Table 1 shows the number and type of the template matching stages.

Table 1. Number and type of traffic sign templates

Form of templates	Number of templates
Triangle	12
Inverted triangle	12
Circle	6
Rectangle	6
Rhombus	4

2.3 Traffic Sign Detection Using SVM

In this stage we decide whether a labeled object is a traffic sign or not through previous learning data. The SVM is a method to classify observations with the two categories basically. When a sign object is tested, it is recognized as positive and negative. In this paper, sign images are actually small and environment images are set as training data. Then the traffic signs are set positive, the environment elements are set negative. When the data is trained, feature elements are values of brightness, data size is 80×80, the number of positive images is 500, and the number of negative images is 2000.

2.4 Stereo Matching

Stereo matching is performed to determine the 3-D motion of the vehicle and the 3-D coordinates (x,y,z) of a detected traffic sign. After performing the calibration of our stereo camera by Zhang's calibration method [9], we obtain the Perspective Projection Matrix (PPM) of the stereo cameras. PPM consists of camera focal length, image scale and origin of the difference between image plane and pixel plane. By setting the two dimensional coordinates(x,y) of sign objects in the left and right image planes as the input, we compute the three-dimensional coordinates of the objects. In this paper, we extract the sign objects from the left image and set the ROI(Region Of Interest) based on the information. We generate a 3-D depth map from the stereo image by using SGBM[10], which is a common matching method. The motion between consecutive video frames is computed based on the depth map. And the 3-D coordinates of the matched signs are obtained by using the depth map also.

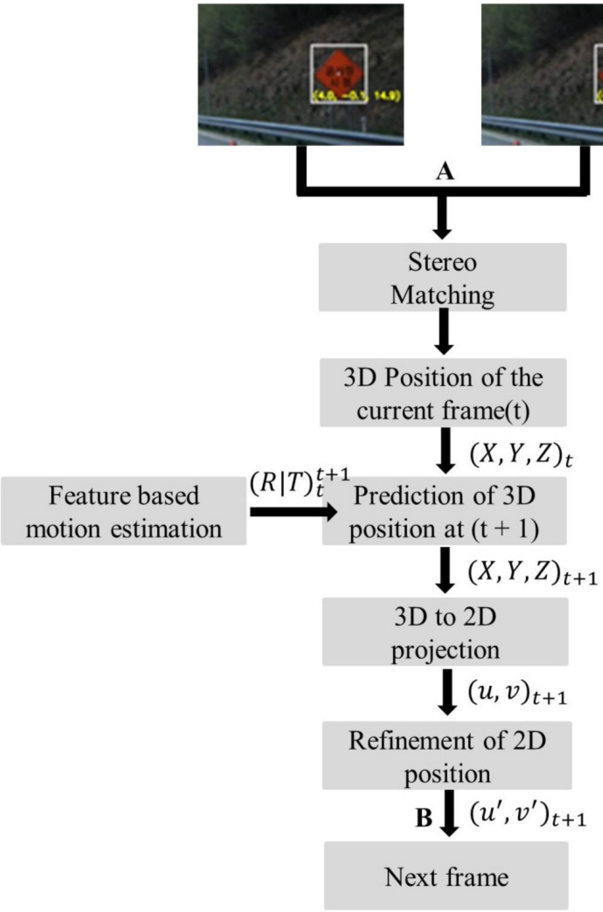

A
Stereo Matching
3D Position of the current frame(t)
$(X,Y,Z)_t$
Feature based motion estimation
$(R|T)_t^{t+1}$
Prediction of 3D position at (t + 1)
$(X,Y,Z)_{t+1}$
3D to 2D projection
$(u,v)_{t+1}$
Refinement of 2D position
B
$(u',v')_{t+1}$
Next frame

Fig. 2. The tracking algorithm, based on motion and stereo vision

2.5 Traffic Sign Tracking

There are two major reasons that sign tracking is needed. First, in the current frame, if we cannot find the signs which is found in the previous frame due to the changes in the image color or the sign is blocked by an obstacle we can use the tracking method to find it. Second, if the current signs are the same as the signs in the previous frame, we can improve the execution speed of the detection step by skipping the SVM operation that takes time. Fig. 2 shows the overall process of the proposed tracking method.

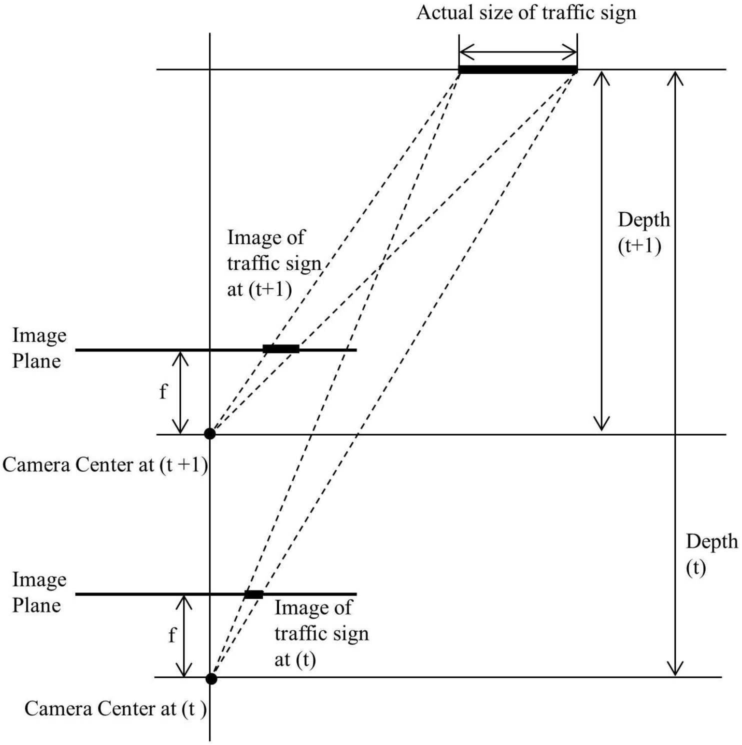


Fig. 3. The size of traffic signs at time (t) and (t+1)

Suppose the time of the current frame is *t* and the next frame is *t+1*. The camera motion information is calculated in the form of transformation matrix between the frame *t* and *t+1*[11]. In [11], after three-dimensional information of road sign is calculated based on the depth map, Random Sample Consensus (RANSAC) algorithm is performed. Then, the initial motion vector is calculated. Later, the motion vector is refined by using a non-linear model. For more information, see [11]. And at frame *t*, we estimate the three-dimensional coordinates of the signs at frame *t+1* using the three-dimensional coordinates of the signs and the transformation matrix at frame *t*. We calculate the two-dimensional pixel coordinates in the *t+1* frame from operation between three-dimensional value that obtained by transformation matrix operation and inverse PPM of the camera information.

2.6 Tracking Refinement

The sign position obtained by the motion of the vehicle is not accurate due to the motion estimation error. Thus, based on object information in the pixel coordinates of predicted signs at *t+1* frame, we calculate accurate pixel coordinates of the sign by template matching. Template matching is done after setting the ROI in the *t+1* frame using the normalized correlation coefficient map. In order to perform correct template matching, we need exact image size of road signs. If we don't know the image size of the signs, template matching gives erroneous results. To know the exact image size, we calculate the size using geometric relations between *t* and *t+1* frames. Fig. 3 shows how to calculate the size of the road signs. Here, f is focal length. First, we have to calculate the actual size of the signs. The actual size of the signs is calculated using Equation 2. In Equation 2, D is depth between the camera and a detected sign, W is the width of the sign in the world coordinate, and w is the width of the sign in the pixel coordinate. Also in Equation 3, H is the height of the sign in the world coordinate and h is the height of the sign in the pixel coordinate. Once we calculate the actual size of the sign in *t* frame, we can calculate the size of the sign in *t+1* frame. Calculation of the size is applied to width and height. Finally, we perform template matching in the *t+1* frame with the calculated size of the signs.

Actual tracking process is managed by a structure object in the C++ programming language. One structure object stands for one road sign. We manage the structure object during our tracking algorithm. When the tracking algorithm finds a road sign, frame number, image size of the sign, pixel position, ROI, and type of the sign are stored in the structure object. And, the information of the same sign tracked in the next frame is updated also. The sign information in the structure object is maintained until the sign disappears.

$$W = \frac{D \times w}{f} \tag{2}$$

$$H = \frac{D \times h}{f} \tag{3}$$

3 Experiment

To obtain experimental video sequences, we use a Bumblebee XB3 stereo camera with 800×600 resolutions at 30 frames rate. Fig. 4 shows the algorithm flow of a road scene. Fig. 5 shows sign detection results with and without tracking method. The left image is t frame, and the right image is t+1 frame. With tracking method, more signs are detected in the consecutive frames. Quantitative performance comparisons of the proposed method are done with two other methods. In the color segmentation with SVM method, tracking algorithm is not performed. In other words, only the sign detection algorithm is performed. The third method is tracking only method. When applying only template matching, the vehicle motion information and the predicted image size of the sign are not used. Cumulative number of missing signs is shown in Fig. 6. In this figure, x-axis represents frame number and y-axis represents the cumulative number of missing signs. Using color and SVM, average processing speed is six frames per second and about 45% of road signs in the test video frames are not detected. In case of using the third method, average performance speed is about eight frames per second and the missing rate of the sign is about 30%.

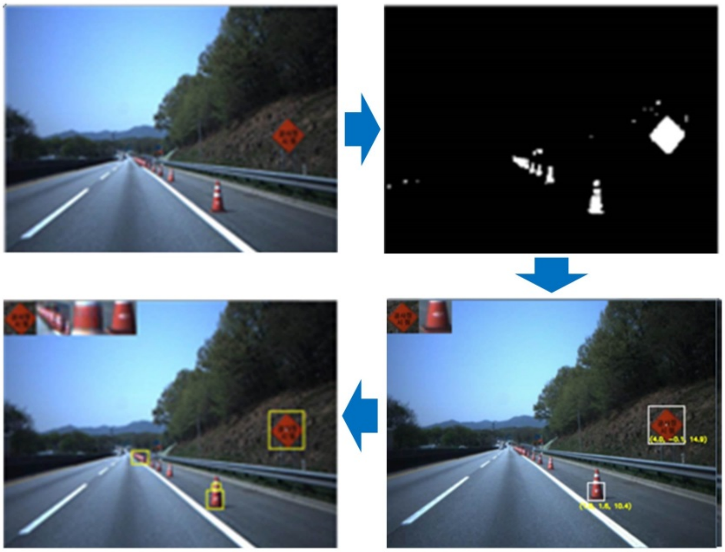
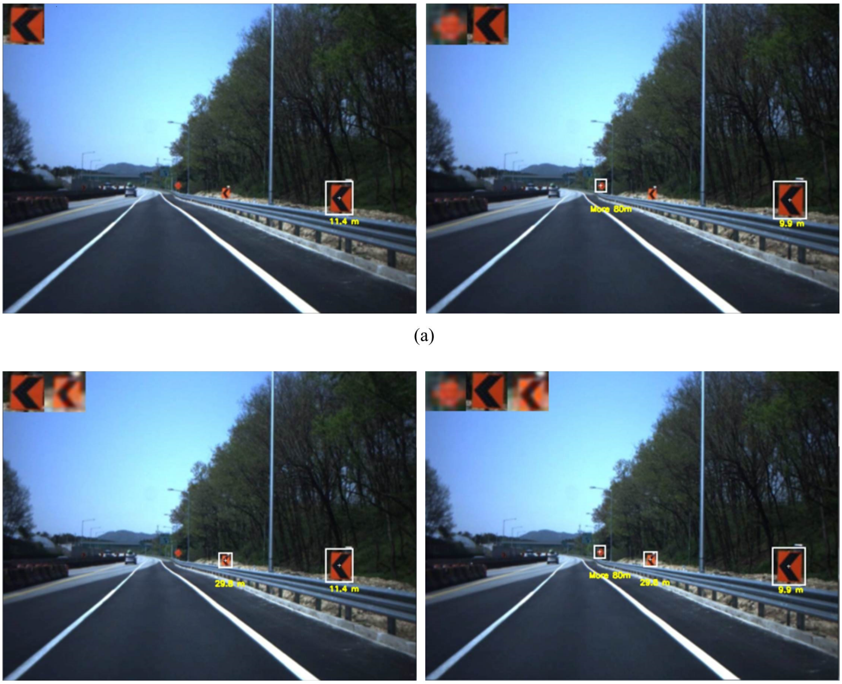


Fig. 4. The Image step of function



(a)

(b)

Fig. 5. Result of the tracking algorithm, (a) without tracking algorithm, (b) with tracking algorithm

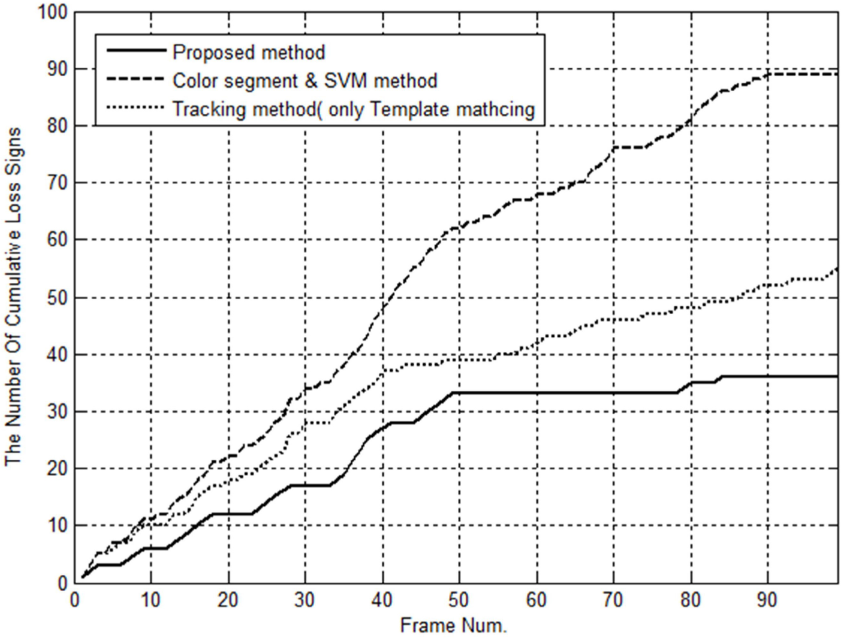


Fig. 6. Number of cumulative missing signs

Table 2. Experiment results table of Color segment & SVM

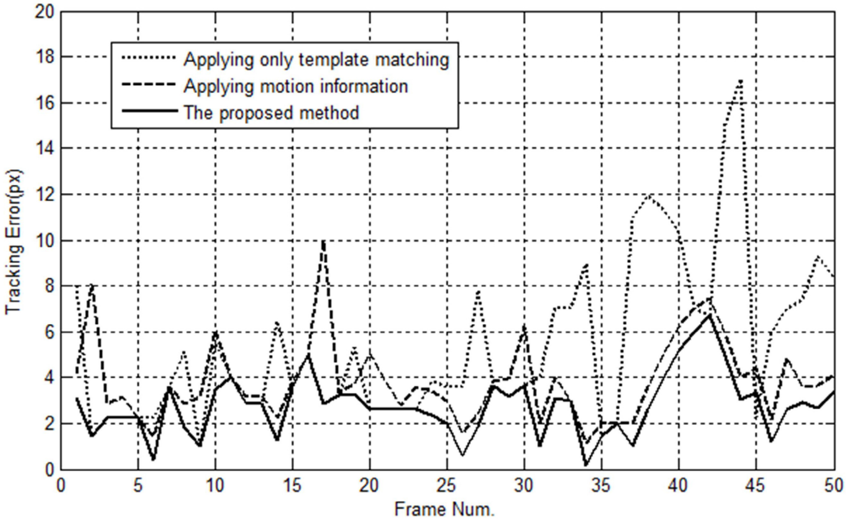
	Data set 1	Date set 2	Data Set 3	Data Set 4
Number of full signs	192	57	35	108
Number of detected signs	109	30	18	58
Success rate	55.1%	52.6%	51.4%	53.7%

Table 3. Experiment results table of Tracking method (only template matching)

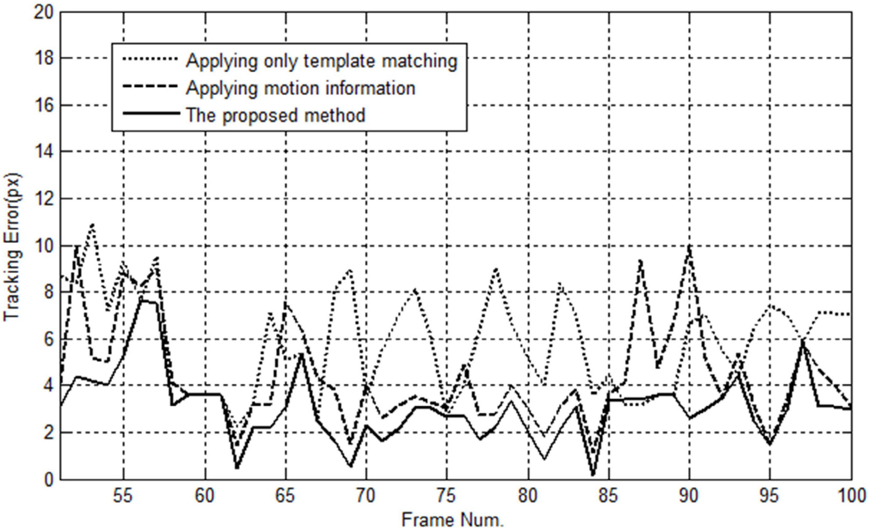
	Data set 1	Date set 2	Data Set 3	Data Set 4
Number of full signs	192	57	35	108
Number of detected signs	138	40	25	74
Success rate	71.8%	70.1%	71.4%	68.5%

Table 4. Experiment results table of the proposed method

	Data set 1	Date set 2	Data Set 3	Data Set 4
Number of full signs	192	57	35	108
Number of detected signs	162	48	31	90
Success rate	84.4%	84.2%	88.6%	83.3%



(a)



(b)

Fig. 7. Distance of the signs center coordinate between ground truth and results of the different methods

Contrast to above mentioned methods, the proposed algorithm runs eight frames per second in average and the missing rate of the signs is about 15%. The graph shown in Fig. 6 shows that higher detection rate of the sign is achieved by using the proposed methods. However, if we use only the template matching, the system will not be able to adapt to the changes of the sign size in each frame while the vehicle is moving. The camera motion information can be used to track the signs more accurately. In the evaluation we check the detection of the signs in each frame. If there is a sign closer than 80m from the vehicle, but it is not detected by the algorithm, we consider it as a missing sign. Main reasons of missed sign are bad color information of the sign and occlusion by obstacles. Table 2, 3 and 4 gives the number of total signs available in the data set, number of successful detections and its percentage using the same three methods respectively. Each data set is different road video with different kind of signs in daytime.

Fig. 7 shows the error of tracking algorithm measured by the distance between the tracked location and the ground truth of the sign center. Prior to the experiment, we find the center coordinate of the signs on each frame manually and make the ground truth. In the figure x-axis represents the frame number and y-axis represents the distance between ground truths and results in pixels. Three different lines represent the three methods and two graphs shows the results for two different data sets.

Motion information method calculates the center coordinate of the signs by multiplying the three-dimensional coordinate of the signs on the previous frame with the camera motion matrix. This method does not use the template matching. Results in Fig. 7 shows that when the signs are closer, the error will be higher when we use only the template matching because it is not considered the changes of the sign size on the frame. The average error is 6.073 pixels for the method which uses only the template matching, 4.410 pixels for the method which uses only the motion information and 3.126 pixels for the proposed method.

4 Conclusion

With the experimental results we find that our proposed method has advantages in road sign detection and processing speed. Once a road sign is detected in a video frame, its 3-D position is calculated by using stereo matching. And the 2-D position of the sign in the next frame is calculated by the motion of the vehicle computed by the stereo vision method. Since we can measure the size of the sign using the 3-D information of the sign, accurate 2-D size of the sign is calculated. This gives better template matching and accurate sign detection. Although we apply a new tracking method for road sign detection, still there is a problem due to color variation among the video sequences. Currently there are about 15% signs are not detected due to color segmentation error, SVM recognition error and so on. In the future, we will continue to increase the detection rate.

Acknowledgment. This research was supported by the MSIP(Ministry of Science, ICT & Future Planning), Korea, under the C-ITRC(Convergence Information Technology Research Center) support program (NIPA-2013-H0401-13-1005) supervised by the NIPA(National IT Industry Promotion Agency) and the Industrial Strategic technology development program, Development of Driver-View based in-Vehicle AR Display System Technology Development(10040927), funded by the Ministry of Knowledge Economy (MKE, Korea).

References

1. Ruta, A., Li, Y., Liu, X.: Detection, Tracking and Recognition of Traffic Signs from Video Input. In: 11th International IEEE Conference on Intelligent Transportation Systems, ITSC 2008, October 12-15, pp. 55–60 (2008)
2. de la Escalera, A., Armingol, J.M., Pastor, J.M., Rodriguez, F.J.: Visual sign information extraction and identification by deformable models for intelligent vehicles. IEEE Transactions on Intelligent Transportation Systems 5(2), 57–68 (2004)
3. de la Escalera, A., Moreno, L.E., Salichs, M.A., Armingol, J.M.: Road traffic sign detection and classification. IEEE Transactions on Industrial Electronics 44(6), 848–859 (1997)
4. Fang, C.-Y., Chen, S.-W., Fuh, C.-S.: Road-sign detection and tracking. IEEE Transactions on Vehicular Technology 52(5), 1329–1341 (2003)
5. Ruta, A., Li, Y., Uxbridge, M., Porikli, F., Watanabe, S., Kage, H., Sumi, K., Amagasaki, J.: A New Approach for In-Vehicle Camera Traffic Sign Detection and Recognition. In: Proc. IAPR Conference on Machine Vision Applications, Japan, (2009)
6. Timofte, R., Prisacariu, V., Van Gool, L., Reid, I.: Combining TrafficSign Detection with 3D Tracking Towards Better Driver Assistance. Emerging Topics in Computer Vision and Its Applications (2011)
7. Uchida, T., Hanaizumi, H.: An automated method for understanding road traffic signs in a video scene captured by a mobile camera. In: 2012 IEEE International Conference on Industrial Technology (ICIT), March 19-21, pp. 108–111 (2012)
8. Maldonado-Bascon, S., Lafuente-Arroyo, S., Gil-Jimenez, P., Gomez-Moreno, H., Lopez-Ferreras, F.: Road-Sign Detection and Recognition Based on Support Vector Machines. IEEE Transactions on Intelligent Transportation Systems 8(2), 264–278 (2007)
9. Zhang, Z.: Camera calibration with one-dimensional objects. In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (eds.) ECCV 2002, Part IV. LNCS, vol. 2353, pp. 161–174. Springer, Heidelberg (2002)
10. Dröppelmann, S., et al.: Stereo Vision using the OpenCV library (2010)
11. Choi, S.-I., Zhang, L., Park, S.-Y.: Stereo Vision Based Motion Adjustment of 2D Laser Scan Matching. In: Image and Vision Computing New Zealand, IVCNZ 2011 (November 2011)