IEICE TRANS. INF. & SYST., VOL.E91–D, NO.12 DECEMBER 2008

LETTER

Traffic Light Detection Using Rotated Principal Component Analysis for Video-Based Car Navigation System*

Sung-Kwan JOO†, Yongkwon KIM†, Seong Ik CHO††, Kyoungho CHOI†††, *Nonmembers*, *and* Kisung LEE†a), *Member*

SUMMARY This letter presents a novel approach for traffic light detection in a video frame captured by an in-vehicle camera. The algorithm consists of rotated principal component analysis (RPCA), modified amplitude thresholding with respect to the histograms of the PC planes and final filtering with a neural network. The proposed algorithm achieves an average detection rate of 96% and is very robust to variations in the image quality.

key words: *car navigation system, traffic light, crossroad detection, principal component analysis*

1. Introduction

Recently, video-based car navigation systems (CNS) have come to be seen as representing the direction of future CNS technology [1]. While conventional CNS provides guidance information with a 2D map or 3D graphics [2], as shown in Fig. 1, the video-based system provides guidance information superimposed on real-time video captured by an in-vehicle camera.

The detection of traffic lights in a video frame is one of the important issues that still need to be solved in order to locate crossroads in video-based CNS. Few studies have investigated traffic light detection. Hwang et al. used hue-saturation-intensity (HSI) color-based segmentation and center detection using Gaussian mask, followed by a verification step using an existence-weight map [3]. In addition to the color features, the shapes of the lamp (circle) and the traffic light unit (rectangle) were also explored for more precise detection by Lindner et al. [4]. The major drawback of these methods is their high sensitivity to variations in the acquired image quality, which strongly depends on the weather, acquisition time, camera performance, etc. In addition, if the vehicle is located more than 100 m from the traffic light unit, the lamp occupies a maximum of only about five pixels in the video frame. In this case, it is difficult for the existing algorithms to detect the traffic lights.

We proposed a novel approach using rotated principal component analysis (RPCA) features which are robust to variations in the quality of the acquired image. The purpose of traffic light detection in this study is to determine the location of the crossroad in a video frame, so that CNS can provide reliable graphical guidance to drivers. Thus in Sect. 3, we also present the video-based CNS equipped with the proposed algorithm. The performance of the proposed method was compared with that of the existing, color-based segmentation method to demonstrate the superiority of the algorithm.

Fig. 1 An example of the video-based car navigation systems (CNS).

2. Traffic Light Detection

2.1 Feature Plane Generation

PCA has been widely used in various image processing applications, due to its ability to rearrange data into orthogonal basis vectors. In the proposed algorithm, a captured video frame is decomposed into R, G, B, and gray level images. After those four values are lined up at each pixel location, PCA is used to extract the PC planes. Then, the first principal component $p(x, y)$ is removed from the channel data by means of the following relation:

$$Y_n(x, y) = X_n(x, y) - \overline{E} \cdot p(x, y) \quad (1)$$

where $Y_n(x, y), n = 1, \ldots, N$ is the image pixel at channel n after the first PC data have been removed, $X_n(x, y)$ the previously aligned channel data, $\overline{E}$ the eigenvector of PCA and $p(x, y)$ the first PC at each pixel location. Thus $\overline{E} \cdot p(x, y)$ is a reconstructed image using only the first PC plane. $Y_n(x, y)$ is acquired by subtracting the reconstructed image from the original image. This technique, known as RPCA, has been successfully utilized in meteorology to emphasize meaningful features by eliminating dominant but undesirable elements. For example, Barlow et al. used RPCA to extract the primary modes of sea surface temperature by removing seasonal variation which was considered to be a dominant and intrinsic feature and was also conspicuous in the first

Manuscript received March 31, 2008.
Manuscript revised July 28, 2008.
†The authors are with Korea University, Seoul, Korea.
††The author is with ETRI, Yuseong-ku, Daejon-shi, Korea.
†††The author is with Mokpo National University, Mokpo-shi, Korea.
*This work was supported by a Korea University Grant.
a) E-mail: kisung@korea.ac.kr
DOI: 10.1093/ietisy/e91–d.12.2884

PC vectors [5].

Figure 2 (a) displays the first PC plane, $p(x, y)$, of a sample video frame. In order to speed up the performance and prevent possible noise factors, the region-of-interest (ROI) is determined first by removing the road area. We used the lane detection algorithm proposed in [6] to do this work. Because the camera is mounted on the same spot, we do not need to detect the road in all frames. After collecting the results of the algorithm in the first hundreds of frames, the size of ROI window was determined by calculating statistical mean of the boundary positions of the detected areas. The determined ROI window was applied to the entire video clip. Since the PCA gives optimal compaction of the information, the first principal component $p(x, y)$ contains the most dominant features or the largest variations from the image data (R, G, B and gray level images) [7]. As shown in Fig. 2 (a), even though $p(x, y)$ contains such important information to our eyes, in our application it further increases the difficulty in finding the target by imposing complexity of the image. Meanwhile, the PC image (Fig. 2 (b)) of $Y_n(x, y)$ after eliminating $p(x, y)$ from the original image $X_n(x, y)$ is much less complex than Fig. 2 (a) and even emphasizes the target objects with simpler background. Our observations indicated that the area for traffic lights appears most obviously in the first or second PC planes of $Y_n(x, y)$. In addition, we focus on red and green traffic lights for this study. These lights look clear in the first and/or second PC images of $Y_n(x, y)$. We, however, use red lights as an example to describe the proposed algorithm.

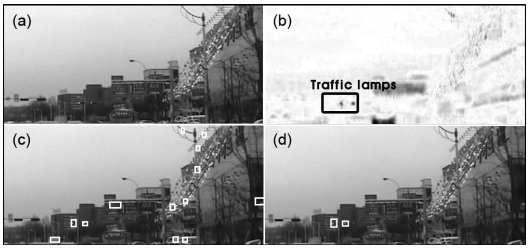

Fig. 2 Region of interest (ROI) images in a video frame during detection process for red lights: (a) PC plane, $p(x, y)$, of the original image $X_n(x, y)$, (b) PC plane of $Y_n(x, y)$, (c) after modified amplitude thresholding, and (d) after artificial neural network (ANN)-based target selection.

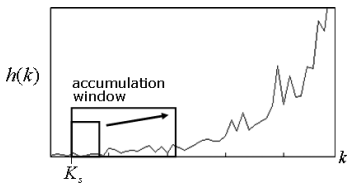

Fig. 3 Histogram of a rotated PC plane (i.e., PC plane of $Y_n(x, y)$). The graph is truncated at the right-hand side in order to zoom in on the accumulation window area. The arrow in the search window represents the direction of accumulation for Eq. (3).

2.2 Object Segmentation

The histogram of the PC plane that was acquired in the previous step is used to segment the possible candidates of the traffic lights by the modified version of the amplitude thresholding method [7].

Figure 3 shows the histogram $h(k)$ of $p(x, y)$ in Fig. 2 (b) where k is the value of $p(x, y)$. In Fig. 3, the higher values of k belong to background area of Fig. 2 (b) while the traffic lights belong to the bins with lower values (left-hand side of the plot). The starting point, K_s, of the accumulation window in the histogram, is determined by:

$$K_s = \arg\min_k (h(k) = H_s) \tag{2}$$

where H_s is the threshold value for deciding the starting point of k. Since a traffic light usually occupies five to fifteen pixels in the video frame, we set $H_s = 5$ for the experiment.

Then the function for the accumulated histogram is defined as:

$$A(u) = \sum_{k=K_s}^{u} h(k). \tag{3}$$

where u is a histogram bin representing the end of accumulation window in Fig. 3. The threshold value V_{th} of the histogram bin is determined by finding u which minimizes the term inside the parenthesis of Eq. (4).

$$V_{th} = \arg\min_u (|A(u) - C|) \tag{4}$$

where C is a constant related to the occupied area of the target objects (i.e. traffic lights). Even though the area of the traffic lights in the PC plane is small (5∼15 pixels) and depends on the distance between the traffic lights and the vehicle, the value C is determined by taking into account the area occupied by the objects with similar color, as well as by the traffic lights. To include as many candidate objects as possible, we set the constant C to 300 for this study.

Once V_{th} is determined, the masks for the candidate objects in the video frame are segmented as follows:

$$b(x, y) = \begin{cases} 1, & \text{if} \quad h(p(x, y)) \geq V_{th} \\ 0, & \text{if} \quad h(p(x, y)) < V_{th} \end{cases} \tag{5}$$

After morphological dilation to refine the segmented masks, the size and shape of $b(x, y)$ were also investigated to select the highly likely candidates.

As shown in Fig. 2 (c), the segmentation algorithm detects not only the regions containing the traffic lights, but also the other objects considered as possible traffic light candidates.

2.3 Target Region Selection

The candidates segmented in the segmentation step indicated that the traffic lights have a different shape and colored texture from those of other objects in the street such as signs and buildings. Figure 4 illustrates these differences of the segmented candidates.

Artificial neural networks (ANNs) [8] offer great potential for quickly and accurately evaluating various pattern

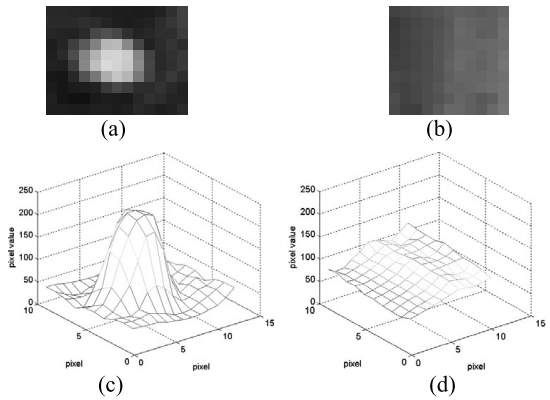

Fig. 4 Texture of objects in the video frame: (a) traffic signal, (b) an example of other candidates (a part of sign board in a building), (c) surface plot of (a), and (d) surface plot of (b).

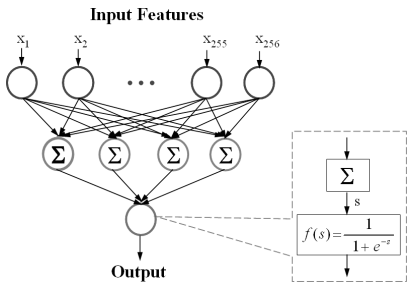

Fig. 5 Three-layer neural network structure to classify traffic lights.

recognition problems. In this research, ANN is adopted to filter out the wrong candidates and finalize the traffic light areas.

R, G, B, and gray planes of each segmented object are normalized to 8×8 pixels to enable the feature vectors to be composed with 256 elements. After various trials with different neural network topologies, a three-layer, neural network structure with 256 inputs, 1 output and 4 neurons in one hidden layer was selected. Figure 5 shows a schematic ANN structure used to classify the traffic lights.

The three-layer structure with the back-propagation algorithm is used to train the ANN. The weight is progressively updated until the maximum root mean square error becomes less than 0.01 during the neural network training. A total of 4000 sets of video frames captured by an in-vehicle camera were used for training the green and red traffic light units and the error percentage for the test sets was 98.7%. Figure 2 (d) shows the traffic lights correctly detected by ANN application.

3. Experimental Results

We tested the proposed approach using multiple sets of video clips which were acquired by a test vehicle [9] developed by the Electronics and Telecommunications Research Institute (ETRI), in Daejon City, Korea. The condition of the acquired video is shown in Table 1. Each video clip was recorded on different days and under varying weather conditions which lead to variation in the clips' intensity. The test data set includes a downtown area, which is very crowded with many buildings and cars, and a residential and suburban area with few vehicles on the road.

Table 1 Descriptions of the tested video clips.

Clip	Intensity	Recoding time	Environment
1	High	3 PM.	Suburb
2	Medium Low	10 AM.	Downtown
3	Medium	1 PM.	Residence
4	Low	9 AM.	Suburb
5	Low	6 PM	Residence

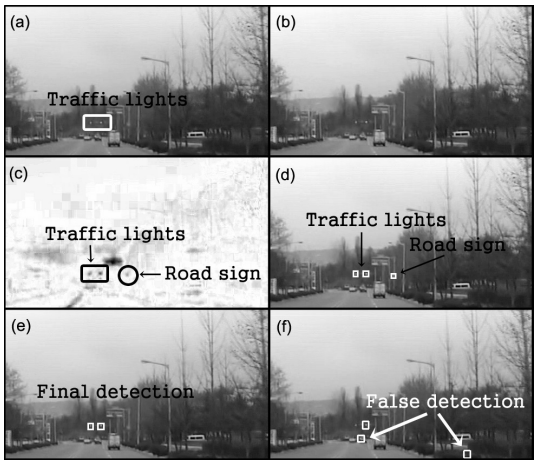

Fig. 6 Results of the proposed approach in comparison with the color-based approach: (a) original image, (b) the first PC plane, $p(x, y)$, removed in the RPCA process of Eq. (1), (c) the PC plane of $Y_n(x, y)$ in Eq. (1), (d) the detected traffic lights and other candidates after object segmentation (Sect. 2.2), (e) final results after the target selection with ANN (Sect. 2.3), and (f) results of the color-based approach.

The video sequences are 640×480 pixels/frame and 24 frames/sec. The algorithm was tested using a Pentium 2.0 GHz PC with 1 GB RAM. A total of 9,473 frames were used to test the detection rates within the distance range of 30–140 m.

Figure 6 shows the results of the proposed method. The first PC plane, $p(x, y)$, of the four channels' data (i.e., R, G, B and gray image planes) is displayed in Fig. 6 (b). According to Eq. (1), Fig. 6 (b) was multiplied by the eigenvector $\overline{E}$ and subtracted from the original 4 channel images. Then, the PC planes were calculated again from $Y_n(x, y)$ for the second iteration. In the second PC plane shown in Fig. 6 (c), the red areas are clearly distinguished from the other components in the video frame. Figure 6 (d) represents the detected traffic lights and other candidates after the object segmentation described in Sect. 2.2 and Fig. 6 (e) shows the final results for classifying traffic lights with ANN. While the proposed method accurately detected the target objects, the color-based approach [3] missed the targets, as shown in Fig. 6 (f).

Figure 7 illustrates the detection rates of the two approaches. The proposed method outperformed the existing colour-based method over all distance ranges. Hwang

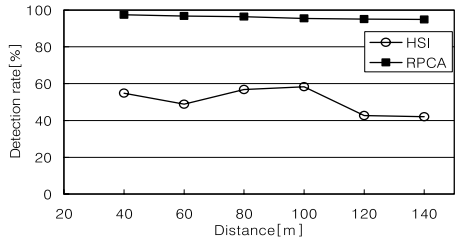

Fig. 7 Detection rates in terms of the distance between the test vehicle and the traffic light units.

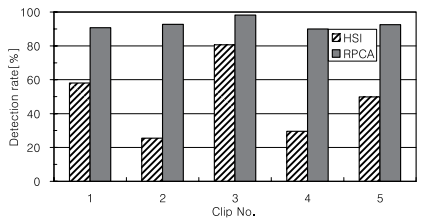

Fig. 8 Detection rates of the test video clips described in Table 1.

Fig. 9 Snapshot of the video-based car navigation system (CNS) which includes the proposed traffic light detection module. The blue box in the center represents the detected traffic lights.

et al. reported that the HSI-based method afforded an average detection rate of more than 90% from a distance of 30 to 90 m [3]. In our application, however, which uses a low-quality, cost-effective camera, the detection rate of the HSI-based method decreased dramatically to 52%, while that of the proposed method was as high as 96%.

Figure 8 also demonstrates the superiority of our method over the existing method based on the video clips in Table 1 that were acquired under different external conditions. The proposed method produced high and consistent detection rates while the HSI-based method was very sensitive to the variations in the image quality.

Figure 9 shows a snapshot of the demonstration system for the video-based CNS developed in ETRI Korea. The figure shows that the proposed algorithm works well and helps the system to guide the driver using graphical icons.

4. Conclusion

We have proposed a robust, traffic light detection method for video-based CNS. The algorithm consists of RPCA, amplitude thresholding on the histograms of the PC planes, and neural network-based refinement.

The proposed method afforded a detection rate of 96%, compared to the existing color-based method which recognizes only 52% of the traffic lights under the same condition. The results also demonstrated the robustness of our method when exposed to variations in the external conditions of the vehicle, acquired image quality, and camera capability. Our method was successful in locating the crossroad for graphical guidance in the video-based CNS developed by ETRI.

References

[1] W. Narzt, G. Pomberger, A. Ferscha, D. Kolb, R. Müller, and J. Wieghardt, "A new visualization concept for navigation systems," LNCS, vol.3196, pp.440–451, 2004.
[2] http://www.jp.sonystyle.com/Product/Car/Nv-xyz/index.html, Sony XYZ 3D Navigation systems.
[3] T.H. Hwang, I.H. Joo, and S.I. Cho, "Detection of traffic lights for vision-based car navigation system," LNCS, vol.4319, pp.682–691, 2006.
[4] F. Lindner, U. Kressel, and S. Kaelberer, "Robust recognition of traffic signals," Proc. IEEE Intelligent Vehicles Symposium, pp.49–53, 2004.
[5] M. Barlow, S. Nigam, and E.H. Berbery, "ENSO, pacific decadal variability, and U.S. summertime precipitation, drought, and stream flow," American Meteorological Society, vol.14, no.9, pp.2105–2128, 2001.
[6] S.G. Jeong, C.S. Kim, D.Y. Lee, S.K. Ha, D.H. Lee, M.H. Lee, and H. Hashimoto, "Real-time lane detection for autonomous vehicle," IEEE ISIE 2001, vol.3, pp.1466–1471, 2001.
[7] A.K. Jain, Fundamentals of Digital Image Processing, Prentice Hall, USA, 1988.
[8] C.M. Bishop, Neural Networks for Pattern Recognition, Oxford University Press, 1995.
[9] S.Y. Lee, K.H. Choi, I.H. Joo, S.I. Cho, and J.H. Park, "Design and implementation of 4S-Van: A mobile mapping system," ETRI Journal, vol.28, no.3, pp.265–274, June 2006.