A Far Sign Recognition by Applying Super-Resolution to Extracted Regions from Successive Frames

Hitoshi Yamauchi[1], Atsuhiro Kojima[2] and Takao Miyamoto[2]
[1]Okayama Prefectural University
[2] Osaka Prefecture University
Japan

1. Introduction

Many researches are proposed for realizing Intelligent Transport Systems (ITS). As a driving assistant system, a part of ITS, many techniques are proposed that extracting and recognizing road signs placed at the front of a car. At present, road-to-vehicle communication systems using dedicated short range communications (DSRC), currently used for electronic toll collection system (ETC) as one of applications, are in experimental phase. By using this communication, road sign information are able to be notified to drivers. Then, informing way of road signs may change from by watching into by such communication systems. Changing informing way by equipping infrastructure may be achieved quickly in urban area. However, long period is needed for completing infrastructure equipment spreading into rural area. Moreover, it is hardly expected that road signs placed at roadside will be abolished and removed after completing infrastructure equipment, while cars are controlled by human being.

For realizing road sign recognition and alerting systems, two techniques are needed. One is a technique for extracting regions including road signs from image sequences. The other is for identifying the type of those signs. In previous techniques, road signs are extracted using genetic algorithms (Uchiyama et al., 1998), active networks (Yabuki et al., 1998), and so on. Since these techniques require too much time to process, it is not always applicable to in-vehicle systems in which real-time processing is necessary. To overcome this problem, fast detection techniques are proposed using specific colors of road signs (Matsuura et al., 2002; Ohara et al., 2002; Zin & Hama, 2005).

As recognition techniques, template matching based technique (Gavrila, 1998) is proposed. However, template matching requires large amount of resize and rotation processing, then processing time will be long. As another approach, feature extraction based on frequency analysis, wavelet transform for example, are difficult to fast processing. To achieve fast processing, a contour vector matching technique is proposed (Yamauchi & Takahashi, 2003). It extracts outlines of the extracted road sign as vectors, and matches the vectors with

reference signs' vectors. By using outline vectors, the technique shows robustness against variation of extraction size and slight rotations.
Target images for these extraction and recognition techniques are still images in a scene. Then, these techniques are applied to frame by frame independently. This means that recognition results may be failed at the frame image which has illumination variances or occlusions. These failed results may cause drivers to be thrown into confusion. Against still images, successive frames in a scene have some correlations each other. By using these correlations, more effective processing can be done. Moreover, this can also remove occasional occlusion.
By the way, a technique of generating pseudo high resolution image by analyzing multiple still images of a same target is proposed (Park et al., 2003). This technique is called as super-resolution. This is used for capturing precise images of objects for digital museum or medical imaging. These targets may require high resolution images over the camera capacity.
In case of road sign detections, a targeted sign will appear in successive frames. And then, the size of the sign region will become larges and high resolution in latter frames. By treating extracted sign images as low resolution images in super-resolution technique, a pseudo high resolution image can be generated. Though an extraction image from a single frame does not have enough quality, early recognition can be done by combining number of successive frames.
In this chapter, a robust road sign recognition technique by applying super-resolution to extracted regions from successive frames is proposed. This technique consists of three steps: 1) searching and tracking road sign, 2) applying super-resolution technique to extracted sign regions, and 3) identifying type of the sign.
In the following, proposed technique is explained in Section 2. Next, the experimental results are shown in Section 3, and the considerations from the experimental results are in Section 4. Finally, the conclusion is in Section 5.

2. Recognition by Applying Super-Resolution to Extracted Regions from Successive Frames

There are many techniques for extracting and recognizing road signs from a frame image. Their target image is single frame, namely, a still image. This means that it is difficult to recognize because of low resolution when the distance to target sign is long yet. Thus, recognition process cannot be applied until the distance becomes short and the resolution is enough for recognition.
In case of road signs extracted from images taken by an in-vehicle camera, the extracted sign is placed continuously on the successive frames. Thus, plural images of a sign can be obtained by applying tracking process. These plural images may generate a pseudo high resolution image by using super-resolution processing, and the generated high resolution image may lead into high accuracy recognition.
The proposed technique is organized from 4 steps, 1) initial extraction of a road sign, 2) tracking the road sign, 3) super-resolution process, and 4) recognizing the super-resolution image. Details of each process are explained in the following.

2.1 Initial Extraction of a Sign Region

To track a sign, it is needed to extract an initial region. In this proposition, extraction of a faraway road sign is needed because early recognition is required. In this initial extraction, high resolution image is not yet necessary because latter super-resolution process may improve the resolution. At this moment, it is enough to judge by only sign shape and colors. For example, in case of restriction signs in Japan, their shapes must be circular and they have red ring at the outer edge (refer to Fig. 1 (a)). For the other example, in case of caution signs, their shapes must be rectangular and their inner area is almost filled by yellow (refer to Fig. 1 (b)). An extraction technique of such kind is presented in (Matsuura et al., 2002).

(a) An example of restriction signs (speed limitation to 50 km/h).

(b) An example of caution signs (existence of left curb ahead).

Fig. 1. Examples of road signs in Japan.

2.2 Tracking of the Sign Region

The proposition needs plural images that include a same road sign. These images are obtained by tracking from the initial extracted road sign region. For successive frames of a video image, many object tracking methods are proposed, CONDENSATION (Isard & Blake, 1998) is one of them. CONDENSATION selects plural candidates which are distributed just before tracking. Then, new candidates are generated from previous ones, obeying normal distribution with plausibility weights. Finally, a candidate which has highest fitness is selected from these new candidates as the tracking result at current frame.

To apply CONDENSATION, an evaluation function calculating fitness has to be designed properly. For road sign tracking, a function based on sign shape, color composition may work properly. For a red circular sign, for example, the parameters to be estimated by CONDENSATION can be selected as three-tuple, the center coordinates (x, y) and the radius r of the circle. In the following, let k as the number of parameter sets generated by CONDENSATION, and let (x_k, y_k, r_k) as parameter values of k -th candidate.

Then, the sign shape will be circular or slight ellipse, and red belt is placed at the outer edge of the circle. In addition, difference of the parameters between tracking results of next-by-next frame will be in small in most cases. Thus, these parameters are able to be employed for the evaluation function.

Let $p_R(x, y)$ as red component value of a pixel at (x, y), N as the number of division on circumference, W_{edge} as the weight for the edge strength. On this condition, the evaluation value E_k^{circle} is denoted as follows.

$$
\begin{aligned}
E_k^{circle} &= \sum_{i=0}^{N-1}\left\{W_{edge}\,p_R\left(x_i^{in}, y_i^{in}\right) - p_R\left(x_i^{out}, y_i^{out}\right)\right\},\\
x_i^{in} &= r_k \cos(2\pi i / N) + x_k,\\
y_i^{in} &= r_k \sin(2\pi i / N) + y_k,\\
x_i^{out} &= (r_k + 1)\cos(2\pi i / N) + x_k,\\
y_i^{out} &= (r_k + 1)\sin(2\pi i / N) + y_k
\end{aligned}
\qquad (1)
$$

By this equation, a parameter is fit to the outer edge of a red circle when E_k^{circle} is large value. The second evaluator E_k^{color} denoted as eq. 2 represents similarity of consistent color distribution between previous extraction result and current extraction. Where, let D_k and D^* as circular regions of k -th parameter of current frame and previous frame, respectively. In addition, let $|D|$ as the number of pixels in region D, and $P_Y(x,y)$, $P_B(x,y)$ as yellow, blue component values at coordinate (x,y), respectively.

$$
\begin{aligned}
E_k^{color} = &\sum_{(x,y)\in D_k}\frac{p_R(x,y)}{|D_k|} - \sum_{(x,y)\in D^*}\frac{p_R(x,y)}{|D^*|}\\
&+ \sum_{(x,y)\in D_k}\frac{p_Y(x,y)}{|D_k|} - \sum_{(x,y)\in D^*}\frac{p_Y(x,y)}{|D^*|}\\
&+ \sum_{(x,y)\in D_k}\frac{p_B(x,y)}{|D_k|} - \sum_{(x,y)\in D^*}\frac{p_B(x,y)}{|D^*|}
\end{aligned}
\qquad (2)
$$

The third evaluator $E_k^{distance}$ denoted as eq. 3 represents difference between current parameter and parameter (x^*, y^*, r^*) which is previous tracking results.

$$
E_k^{distance} = W_{pos}\sqrt{(x^* - x_k)^2 + (y^* - y_k)^2} + W_{rad}\sqrt{(r^* - r_k)^2} \qquad (3)
$$

Where, let W_{pos} and W_{rad} are weight constants for position variance and radius variance, respectively.

These three evaluation values are weighted properly, and totally evaluated. Then, the parameter of the highest evaluation is selected as a tracking result for each frame. For the other signs, such as blue circular, yellow rectangular signs, tracking process may be achieved by defining similar evaluators.

2.3 Application of Super-Resolution to Tracked Regions

A pixel image is achieved from a receptor of an image sensor by gathering rays through a lens. Then, the observed ray on each receptor becomes a mixed ray from acceptable area at the ratio specified by ray gathering characteristics. On the assumption that the target is a high resolution image, each pixel of the observed image is got by convolution of the high resolution image and a filter function which is represented as ray gathering characteristics. The super-resolution process is a technique to recover the original high resolution image from some observed low resolution images. Some super-resolution techniques are proposed, a local iterative calculation method is presented in (Irani & Peleg, 1991), which is one of the standard techniques of super-resolution. In the following, the outline of this method is described.

Let $f(x,y)$ as the true image to be calculated, and $c_i(u,v)$ as the observed low resolution images. And let $h_c(x,y)$ as a point spread function (PSF) which affects observed image. The super-resolution technique is done by iteration of 4 steps described as follows.

1. Let $f^{t=0}(x,y)$ as a first solution of an estimated high resolution image. In this paper, the latest observed image $c_{last}(x,y)$ is normalized to the regular size and adopted as solution at this moment.
2. Calculate simulated observed image $c_i^t(u,v)$ by applying PSF $h_c(x,y)$ to an estimated high resolution image $f^t(x,y)$ which is a last solution (refer to eq. 4).

$$c_i^t(u,v) = \sum_{(x,y)} f^t(x,y) h_c(x_u - x, y_v - y) \tag{4}$$

   Where, coordinates (x_u, y_v) on an estimated high resolution image is corresponding to coordinates (u,v) on a simulated observed image.
3. Calculate a sum of squared error between a simulated observed image $c_i^t(u,v)$ and a real observed image $c_i(u,v)$ (refer to eq. 5).

$$E^t = \sqrt{\sum_i \sum_{(u_i,v_i)} \left\{c_i(u_i,v_i) - c_i^t(u_i,v_i)\right\}^2} \tag{5}$$

   If E^t is under a threshold T_e, quit this iteration process.
4. Modify an estimated high resolution image $f^t(x,y)$ so that the error E^t may become small. This modification result $f^{t+1}(x,y)$ is a last estimated high resolution image. Then, return to step 2.
   In this step, a modification result $f^{t+1}(x,y)$ is calculated as eq. 6, 7, 8. Where, let coordinates $(u_{i,x}, v_{i,y})$ on a real observed image $c_i(u,v)$ as pixels observing coordinates (x,y) on the estimated high resolution image $f^t(x,y)$.

$$\Delta e_i^t(u,v) = c_i(u_{i,x}, v_{i,y}) - c_i^t(u_{i,x}, v_{i,y}) \tag{6}$$

$$\Delta E_i^t = \sum_{(u_{i,x},v_{i,y})} \Delta e_i^t(u_{i,x}, v_{i,y}) \frac{h_c(x - x_{u_i}, y - y_{v_i})^2}{C} \tag{7}$$

$$f^{t+1}(x,y) = f^{t}(x,y) + \sum_{i} \Delta E_i^t \tag{8}$$

In the process of super-resolution, each observed image must be registered in order to make correspondence between a pixel on a real observed image (u_x, v_y) and a pixel on a high resolution image (x, y). In this chapter, an ordinal template matching technique is used after resizing extracted images from observed images to a same size.
Note that extracted images by tracking are varying size related with the distance between the camera and the sign. This means that resolution of the extracted image grows. Thus, larger extraction size may mean that the accuracy is higher.
In applying super-resolution process to road sign recognition techniques, each extraction size is varyingly according to relative distance. This leads addition in eq. 8 may be varyingly. Thus, the accuracy varying may reflect the equation results.

2.4 Recognition of Super-Resolution Image

As the final process, a recognition process is applied to a generated super-resolution image. For road sign recognition process, it can employ an already proposed technique, such as template matching which is widely used in the image processing area, and so on.

3. Confirming Experiments

Some experiments are done for confirmation of the proposed technique.

3.1 Experimental Conditions and Preliminary Experiment

To confirm effectiveness of the proposed technique, some experiments were done. As input images, a camera SONY DFW-VL500 with 1/3 inch CCD sensor and vary-angle lens 5.5 to 64mm is used. The camera was placed on dashboard of a car. Images are taken at 15 fps in fine or cloudy daytime. The image sizes of all images are 640×480 pixels. Some of these images include occasionally lighting varying as affection by high reflection of daylight or occasionally clouding.
Tracking process was done from the frame that an extraction road sign's radius is more than 7 pixels. This extraction size is corresponding that the relative distance between the camera and the target sign is about 125 m. For tracking and extraction process from frames following the first extraction frame, CONDENSATION was employed which was described in 2.2. The parameters for fitness calculation were $W_{edge} = 1.0$, $W_{pos} = 1.25$, $W_{rad} = 1.0$. As the recognition technique, a template matching algorithm which is generally used was employed. In the extraction, tracking and recognition process, the technique proposed in (Yamauchi & Takahashi, 2004) is used for specific color detection.
Prior to verification experiments, a measurement experiment was done. In this experiment, relation of relative distance between the camera and the target sign and size of the extracted sign was obtained. The measurement was done by taking image at distance one by one meters from 100 m to 11 m, in which the target sign becomes frame-out. As a result of the measurement, a relation between relative distance and extraction radius was achieved. By

using regression analysis, the regression equation can be represented as eq. 9, where r is radius of extracted sign image in pixels, and d is relative distance in meters.

$$d = \frac{1348.8}{r^{1.2284}} \tag{9}$$

The R^2 value, which is used for an index of suitableness, of this equation is 0.9914, which represents that eq. 9 is a suitable equation. In the following, relative distance is used for evaluations, which is more important index in practice case, by using eq. 9.

3.2 Experimental Result

As an experimental results, Fig. 2 is expressing the result targeted a yellow precaution sign, left curbing caution (refer to Fig. 1 (b)). These experimental observed images were taken in a fine day of December at around 14:15. In this figure, results of 1) extracted images from each single frame, 2) superposed images of the extracted images as comparison targets, and 3) proposed super-resolution images, are shown. The superposing technique is proposed in (Yamauchi et al., 2007). These superposed images are generated by weighted averaging of extracted images according to sizes of extraction images.

In Fig. 2, extracted images from single frames that are picked up from a scene every 5 frames, and processing result images of these extracted images are shown. This picking up is from frames which is taken at relative distances range in 100 to 50 m. The columns are relative distances calculated by eq. 9 and extracted sizes, extracted images from single frames, superposed images, and super-resolution results, respectively. Moreover, correct recognition ratios on the images are shown under each image. These ratios are average value of 10 frames which consists of the target frame, formerly 5 frames and following 4 frames. In these relative distances, sizes of extracted images are lower than 30×30 pixels. Most of former road sign recognition techniques are hard to recognize correctly for such sizes. Then, it is relative difficult to recognizing by human observation. In this condition, not only extractions from a single frame but also superposed images are not clear because of blur affection. In spite of such condition, super-resolution results are clearly than these comparison images.

From the figure, the recognition results on extracted images from single frames are almost all failed in the relative distance range. The main reason of this may be low resolution resulted from long relative distance. On another comparison technique, superposition technique proposed in (Yamauchi et al., 2007), high recognition ratio is shown in range near than 60 m. However, results at more far distance are almost all failed, similar to results on extracted images from single frames. This means that blur affection resulted from low resolution may not be greatly improvable by using superposition.

As comparison with these results, results of the proposed technique using super-resolution are almost full correct recognitions. This due to resolution estimation from plural low resolution images which have less resolution information themselves. Thus, the estimated high resolution images may be properly estimation, and this leads correct recognitions.

Note that the correct recognition ratios on super-resolution images shown as (c8) and (c9) are low. At such relative distance, extracted images are in higher resolution themselves. These failures may be caused by position shifts because of low position detection accuracy in CONDENSATION of the experiments. Thus, these ranges of position shifts are exceeded search range of the registration process of super-resolution process. Then, super-resolution process was done using incomplete registration images.

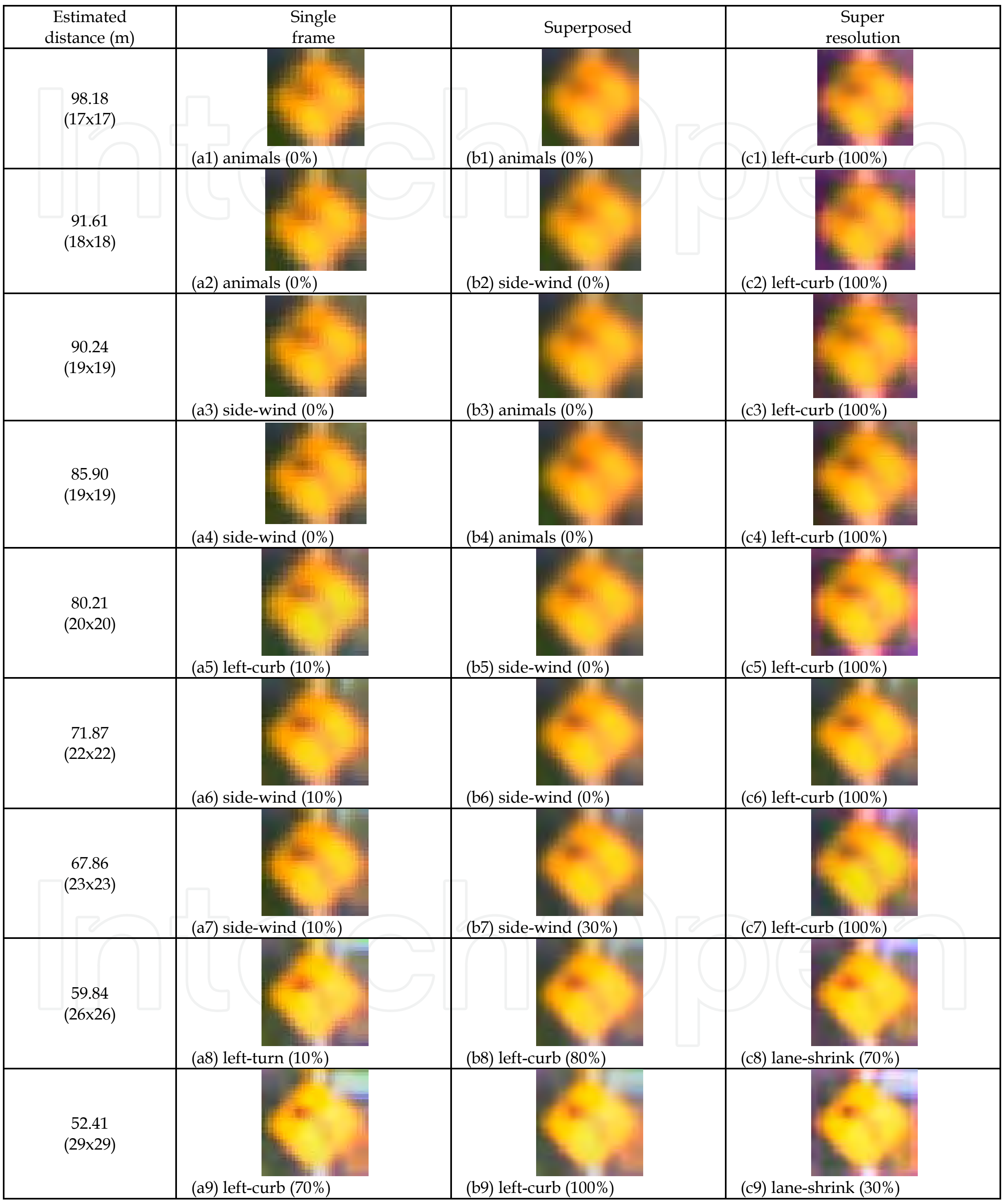

Fig. 2. An example result of extraction, superposition, super-resolution and recognition for a yellow rectangle sign.

4. Consideration

The characteristics of the proposed technique are described as follows.

1. Road sign images are extracted from successive frames by tracking. Using these extracted images, a relatively high resolution image is estimated by employing super-resolution technique. The experimental results show that this technique achieves improvement of recognition ratio.
2. In real driving scene, correct recognition of road sings at far relative distance is more useful and meaningful. It is shown that the proposed technique improves quite improvement at more far range 50 to 100 m than ordinary road sign recognition techniques' target ranges.
3. In the experiments, it was not in consideration that extraction images of a targeted road sign may be affected by angle changes. In such cases, estimating affine parameters is useful for more accurate recognition. But increasing processing time by adding this process is in trade-off relation against low recognition accuracy for real-time processing applications such as road sign recognitions. To overcome this increasing processing time, employing some object matching techniques, such as Scale Invariant Feature Transform (SIFT) (Lowe, 2004) based technique, may be usefulness.

5. Conclusion

In this paper, a road sign recognition technique applying super-resolution to extracted images from successive frames is proposed to achieve high recognition accuracy. The super-resolution process is applied on low resolution images extracted from successive frames by tracking using CONDENSATION.

As results of comparison to recognition ratio of extracted images from single frames and superposition images, high accurate recognition ratio is achieved for signs at far position by the proposed technique.

Verification of weights in super-resolution process, development of eliminating technique cased by occlusions or shadowing, are future works.

6. References

Gavrila, D.M (1998). Multi-feature Hierarchical Template Matching Using Distance Transforms, *Proceedings of IEEE International Conference on Pattern Recognition*, pp. 439-444, 0-8186-8512-3, Brisbane, Australia, September 1998.

Irani, M. & Peleg, S. (1999). Improving Resolution by Image Registration, CVGIP:Graphical Models and. Image Processing, Vol. 53, No. 3, May 1991, pp. 231-239, 1049-9652.

Isard, M. & Blake, A. (1998). CONDENSATION -- Conditional Density Propagation for Visual Tracking, *International Journal of Computer Vision*, Vol. 29, No. 1, 1998, pp. 5-28, 0920-5691.

Lowe, D.G. (2004) Distinctive image features from scale-invariant keypoints, *International Journal of Computer Vision*, Vol. 60, No. 2, 2004, pp. 91-110.

Matsuura, D.; Yamauchi, H. & Takahashi, H. (2002). Extracting Circular Road Signs Using Specific Color Distinction and Region Limitation, *IEICE Transactions on Information and Systems, Part 1 (Japanese Edition)*, Vol. J85-D-II, No. 6, June 2002, pp. 1075-1083.

Ohara, H.; Miki, S.; Yabuki, N.; Fukui, Y. & Sumi, Y. (2002). Detection and Recognition of Road Signs Using Neural Network, *Proceedings of the 15th Workshop on Circuits and Systems in Karuizawa*, pp. 505-510, Karuizawa, Japan, Apr. 2002 (in Japanese).
Park, S.C.; Park, M.K. & Kang, M.G. (2003). Super-Resolution Image Reconstruction: A Technical Overview, *IEEE Signal Processing Magazine*, Vol. 26, No. 3, May 2003, pp. 21-36, 1053-5888.
Uchimura, K.; Kimura, H. & Wakiyama, S. (1998). Extraction and Recognition of Circular Road Signs Using Road Scene Color Images, *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences (Japanese Edition),* Vol. J81-A, No. 4, April 1998, pp. 546-553.
Yabuki, N.; Matsuda, Y.; Kimura, H.; Fukui, Y. & Miki, S. (1998). Extraction of Special Color Region Using Feature Energy of Color Image, *Proceedings of the 11th Workshop on Circuits and Systems in Karuizawa*, pp. 101-106, Karuizawa, Japan, Apr. 1998 (in Japanese).
Yamauchi, H. & Takahashi, H. (2003). A Road Sign Recognition Technique by Tracing Outline Vectors, *Journal of The Institute of Image Information and Television Engineers*, Vol. 57, No. 7, July 2003, pp. 847-853 (in Japanese).
Yamauchi, H. & Takahashi, H. (2004). A Specific Color Detection Technique for Road Sign Recognition Using Outline Vectors, *Technical Report of IEICE*, No. IE2004-88, November 2004, pp. 11-16 (in Japanese).
Yamauchi, H.; Kojima, A.; Miyamoto, T.; Takahashi, H. & Fukunaga, K. (2007). A Robust Road Sign Recognition by Superposing Extracted Regions from Successive Frames, *IEICE Transactions on Information and Systems (Japanese Edition)*, Vol. J90-D, No. 2, February 2007, pp. 494-502 (in Japanese).
Zin, T.T. & Hama, H. (2005). A Robust Road Sign Recognition Using Segmentation with Morphology and Relative Color, *Journal of the Institute of Image Information and Television Engineers*, Vol. 59, No. 9, September 2005, pp. 1333-1342.

Mechatronic Systems Applications

Edited by Annalisa Milella Donato Di Paola and Grazia Cicirelli

ISBN 978-953-307-040-7

Hard cover, 352 pages

Publisher InTech

Published online 01, March, 2010

Published in print edition March, 2010

Mechatronics, the synergistic blend of mechanics, electronics, and computer science, has evolved over the past twenty five years, leading to a novel stage of engineering design. By integrating the best design practices with the most advanced technologies, mechatronics aims at realizing high-quality products, guaranteeing at the same time a substantial reduction of time and costs of manufacturing. Mechatronic systems are manifold and range from machine components, motion generators, and power producing machines to more complex devices, such as robotic systems and transportation vehicles. With its twenty chapters, which collect contributions from many researchers worldwide, this book provides an excellent survey of recent work in the field of mechatronics with applications in various fields, like robotics, medical and assistive technology, human-machine interaction, unmanned vehicles, manufacturing, and education. We would like to thank all the authors who have invested a great deal of time to write such interesting chapters, which we are sure will be valuable to the readers. Chapters 1 to 6 deal with applications of mechatronics for the development of robotic systems. Medical and assistive technologies and human-machine interaction systems are the topic of chapters 7 to 13.Chapters 14 and 15 concern mechatronic systems for autonomous vehicles. Chapters 16-19 deal with mechatronics in manufacturing contexts. Chapter 20 concludes the book, describing a method for the installation of mechatronics education in schools.

InTech Europe

University Campus STeP Ri

Slavka Krautzeka 83/A

51000 Rijeka, Croatia

Phone: +385 (51) 770 447

InTech China

Unit 405, Office Block, Hotel Equatorial Shanghai

No.65, Yan An Road (West), Shanghai, 200040, China

中国上海市延安西路65号上海国际贵都大饭店办公楼405单元

Phone: +86-21-62489820

www.intechopen.com

Fax: +385 (51) 686 166
www.intechopen.com
Fax: +86-21-62489821