

BICYCLE DETECTION USING PEDALING MOVEMENT BY SPATIOTEMPORAL GABOR FILTERING

KAZUYUKI TAKAHASHI, YASUTAKA KURIYA AND TAKASHI MORIE

Graduate School of Life Science and Systems Engineering
Kyushu Institute of Technology
2-4 Hibikino, Wakamatsu-ku, Kitakyushu 808-0196, Japan
morie@brain.kyutech.ac.jp

Received March 2011; revised October 2011

ABSTRACT. *For safety vehicle driving assistance, it is important to detect objects such as cars, pedestrians and bikes. Although high accuracy is achieved by HOG features for object detection, it is difficult to discriminate bicycles from similarly shaped objects such as motorbikes. In this paper, we propose a bicycle detection algorithm that utilizes relative motion detection of leg movements during bicycle pedaling using spatiotemporal 3D Gabor filtering as well as shape-based object detection using HOG and SVM. The pedaling movement enables us to discriminate between bicycles and motorbikes and to improve the bicycle detection accuracy. Using test video images, our proposed algorithm showed improved performance in discriminating between bicycles and motorbikes.*

Keywords: Bicycle detection, Spatiotemporal 3D Gabor filter, Leg movement, HOG, SVM

1. Introduction. To improve driving safety, object detection/recognition systems using a vehicle camera are being actively developed. The objects to be detected include cars, pedestrians, bicycles and motorbikes. Many systems have been proposed for detecting objects, and high accuracy in pedestrian detection has been achieved by foreground extraction or motion detection methods and HOG (Histogram of Oriented Gradients) features [1, 2, 3].

In order to represent object shapes, the HOG approach, which encodes an image by histogram-based visual features, is widely used [4]. This approach is considered suitable not only to human detection but also to bicycle detection. Previously, the HOG approach was compared with several features such as Harris operators, LSI (Latent Semantic Indexing) and SIFT (Scale Invariant Feature Transform) for detecting pedestrians and bicycles in traffic scenes and achieved the best performance among them [5]. An SVM (Support Vector Machine) was used as the classifier; it has a so-called maximum margin classification and regression approach. It is known that SVMs are suitable for high-dimensional data classification.

Although some bicycle detection algorithms have been proposed [5, 6, 7], there are very few algorithms that discriminate bicycles from similar objects such as motorbikes [8].

In the literature [6], bicycles and pedestrians on a sidewalk are detected using optical-flow based motion information and are tracked using the Harris corner edge detector. The utilization of motion information is effective for detecting objects on the road from stationary video cameras. After a moving object is detected, it is determined whether or not it is a bicycle. In previous studies [7, 8], the two circular wheels of a bicycle receive the focus; bicycles are detected using a circular Hough transform and the edge-based Hausdorff metric [7]. In [8], the object's classification as a bicycle or motorbike is performed based

on images of the wheel regions. Even though these methods use sophisticated approaches, the detection process fails in some situations where edges similar to circular patterns exist in an input image.

In order to recognize more detail about a traffic situation, a more accurate detection method is required, which can be achieved by combining shape features obtained by the above approaches and motion information included in video sequences. In a previous study [9], the body movement is used for person detection. We can detect bicycles from motorbikes and other objects not only using shape features but also by the characteristic leg movements that occur during bicycle pedaling [10].

In this paper, we describe a detailed bicycle detection algorithm using pedaling movement, and show its quantitative evaluation results. This paper is organized as follows. In Section 2, our proposed algorithm is described. Experimental results to verify our algorithm are shown in Section 3. Discussion is presented in Section 4. Finally, the conclusion is given in Section 5.

2. Proposed Approach. The process flows of our proposed approach are shown in Figures 1 and 2. As the first step, we need to detect two-wheel objects. Some methods for detecting two-wheel objects have been proposed. One method is shown in Figure 1(a): namely, moving object regions in video images captured by a camera are detected and extracted based on the motion information; then, the extracted regions are analyzed to determine whether the regions of motion represent any two-wheel objects based on the shape information. Another approach is shown in Figure 1(b): namely, two-wheel object detection is performed based on shape features such as HOG, and the detected two-wheel object is tracked.

In this paper, we evaluated the method shown in Figure 1(a). In order to calculate motion information, a spatiotemporal 3D Gabor filter was used. Moving object regions were detected and extracted from the image by nearest neighbor (NN) clustering. Details of the spatiotemporal 3D Gabor filtering are described in Section 2.1.

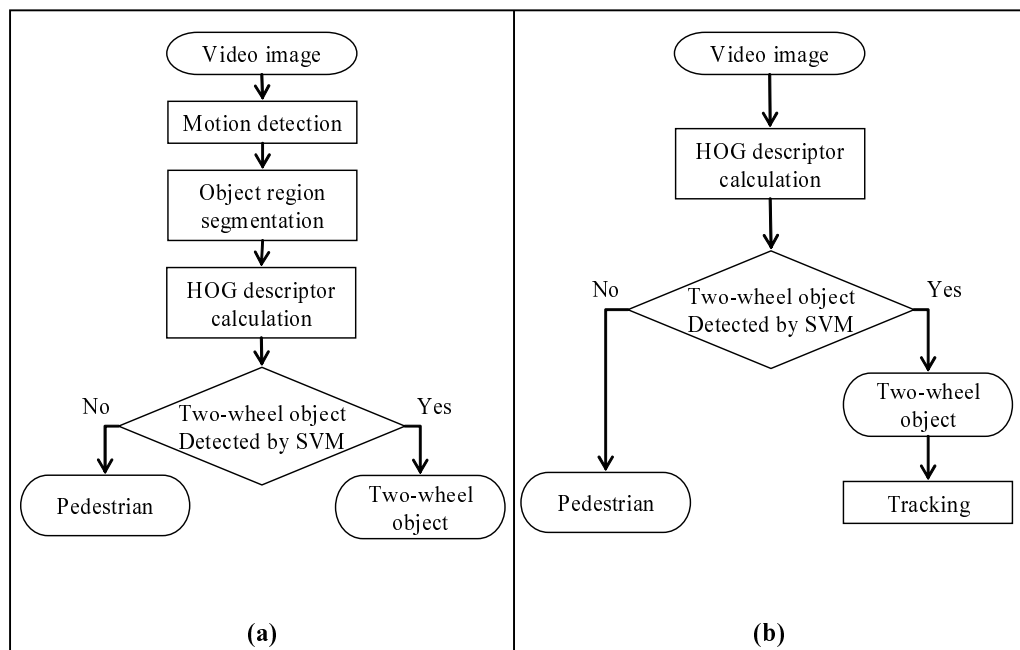


FIGURE 1. Two different flows of two-wheel object detection

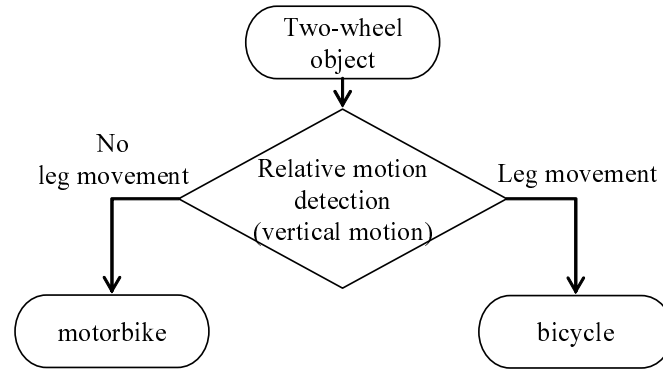


FIGURE 2. Flow of bicycle detection

The HOG feature is calculated for each extracted region, and it is determined whether the object in the region is a pedestrian or a two-wheel vehicle (bicycle/motorbike) by SVM using the HOG feature. When the object is judged to be a two-wheel vehicle, it is necessary to determine whether it is a bicycle or a motorbike. Leg movement by pedaling enables us to discriminate between a bicycle and a motorbike. When one is pedaling a bicycle, the trajectories of the legs are intrinsically circular. However, from a camera view independent of the motion of the whole bicycle, they do not look circular. In order to detect leg movement during pedaling more accurately, we transform the original image sequence to that observed from a viewpoint relative to the center of the bicycle. After this transformation, the relative leg motion becomes nearly circular.

For accurate motion detection, the Lucas-Kanade gradient-based optical flow method is widely used in computer vision [11]. It is known that such gradient-based methods can give accurate results despite their low computational costs, but that they are very sensitive to noise included in the input image [12]. On the other hand, spatiotemporal filters are robust to noise compared with other motion detection methods.

2.1. Spatiotemporal 3D Gabor filtering. The spatiotemporal 3D Gabor filter is a velocity-tuned spatiotemporal filter [13, 14, 15]. This filter is based on the spatial 2D Gabor filters, which have both frequency-selective and orientation-selective properties as well as optimal joint resolution in both spatial and frequency domains [12, 16]. Spatial 2D Gabor filters are described by the following equation in the x - y spatial plane:

$$g_{\theta,\phi,t_0}(x,y) = \exp\left(-\frac{x_0^2 + \gamma^2 y_0^2}{2\sigma_s^2}\right) \exp\left[i\left(\frac{2\pi}{\lambda}x_0 + \omega t_0\right)\right], \quad (1)$$

$$x_0 = x \cos \theta + y \sin \theta, \quad y_0 = -x \sin \theta + y \cos \theta, \quad (2)$$

where filter parameter γ determines the aspect ratio of the spatial filter size, σ_s is the spread of spatial Gabor filter, λ is the period of the sinusoidal plane wave, ω and t_0 are the angular velocity for shifting in the time domain and the shifting time, respectively, and i equals $\sqrt{-1}$. Angle parameter θ determines the preferred spatial orientation of the filter. The response of spatial 2D Gabor filters is computed by convolution with input image $I(x, y, t)$ at frame time t :

$$r_{\theta,\phi,t_0}(x,y,t) = I(x,y,t) * g_{\theta,\phi,t_0}(x,y). \quad (3)$$

The response of the spatiotemporal 3D Gabor filter is obtained by the combination of the convolution outputs from a set of spatial 2D Gabor filters mutually shifted with an appropriate time span.

In order to detect edges moving along a certain direction with high sensitivity, the orientation of the spatial 2D Gabor filters should be perpendicular to the direction, and

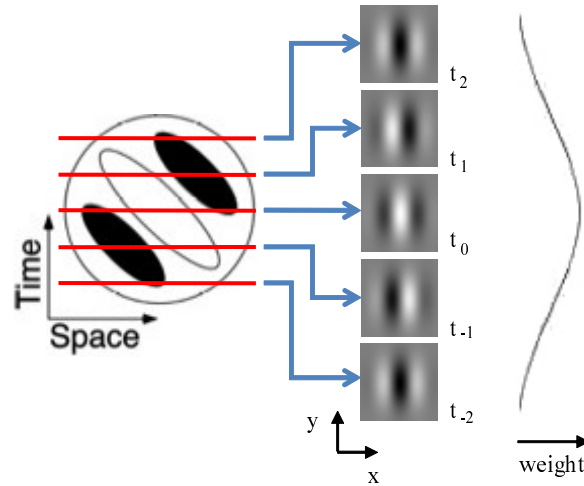


FIGURE 3. Spatiotemporal 3D Gabor filter (left) and x - y plane outputs of the filter at different timing (right)

the phase of the spatial 2D Gabor filters between two continuous frames should be varied linearly with an equal phase step in the range from -90 to 90 degrees by the term ωt_0 in Equation (1). Images processed with such phase-shifted filters are summed over several subsequent frames, as shown in Figure 3. A Gaussian temporal weight function is used across the subsequent frames:

$$s_{v,\theta,\phi}(x, y) = \sum_t r_{\theta,\phi,t}(x, y, t) \exp\left(-\frac{t^2}{2\sigma_t^2}\right), \quad (4)$$

where σ_t is the spread of the temporal Gaussian filter. Velocity v in Equation (4) is defined by the following equation:

$$v = \frac{\lambda\omega}{2\pi}. \quad (5)$$

Therefore, the detectable speed of the 3D Gabor filter is determined by λ and ω .

The spatiotemporal motion energy is introduced using a stable phase-invariant energy expression, which is defined as the square root of the sum of the squared outputs of spatiotemporal 3D Gabor filters with the real part or the imaginary part:

$$E_{v,\theta}(x, y) = |s_{v,\theta,\phi}(x, y)|. \quad (6)$$

The motion energy for all velocities in the same moving direction are compared, and the velocity of the filter $v^*(x, y)$ that outputs the maximal motion energy $l_\theta(x, y)$ is regarded as the detected velocity $V_\theta(x, y)$:

$$v_\theta^*(x, y) = \arg \max_v E_{v,\theta}(x, y). \quad (7)$$

$$l_\theta(x, y) = E_{v_\theta^*,\theta}(x, y). \quad (8)$$

$$V_\theta(x, y) = v_\theta^*(x, y). \quad (9)$$

The detectable range of motion velocity obtained using the spatiotemporal energy model is proportional to the spatiotemporal filter size. If the filter size is fixed, the detectable speed range is limited, but a multi-resolution approach overcomes this limitation as shown in Figures 4(a)-4(c).

Decreasing the image size is equivalent to increasing the filter size. High resolution in filtering leads to slow motion detection, and vice versa. The input image is scaled down with some lower resolutions $I_\tau(x, y)$, where τ represents the scaling factor, and motion detection is performed independently at each resolution. Among all motion energy values

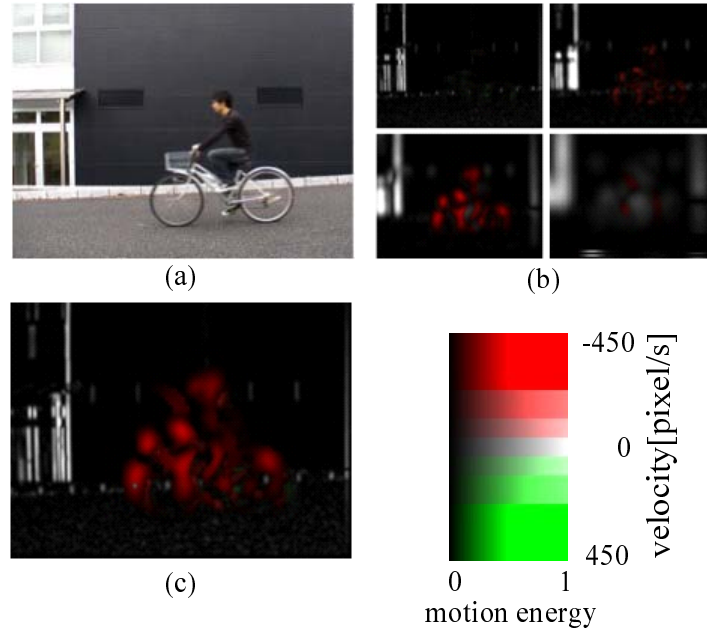


FIGURE 4. Spatiotemporal 3D Gabor filtering outputs: (a) input image, (b) detected motion at four different resolutions; image size is 640×480 ($\tau = 1$, upper left), 320×240 ($\tau = 1/2$, upper right), 160×120 ($\tau = 1/4$, lower left) and 80×60 pixels ($\tau = 1/8$, lower right), where $\theta = 90$ deg and $\sigma = 4$ pixels and (c) detected motion combining four different resolutions

with the different resolutions $l_{\theta,\tau}(x, y)$, the velocity that outputs the maximal motion energy $L_{\theta}(x, y)$, $V_{\theta,\tau^*}(x, y)$, is regarded as the detected velocity $V_{\theta}(x, y)$:

$$\tau^*(x, y) = \arg \max_{\tau} l_{\theta,\tau}(x, y). \quad (10)$$

$$L_{\theta}(x, y) = l_{\theta,\tau^*}(x, y). \quad (11)$$

$$V_{\theta}(x, y) = V_{\theta,\tau^*}(x, y)/\tau^*(x, y). \quad (12)$$

Although Gabor filtering exhibits good performance for detecting accurate velocities, the computational cost of this filter is extremely high [12]. By using parallel computation using GPUs (Graphic Processing Units), this computational cost issue may be overcome [17].

2.2. Tilt alignment. In order to calculate the relative motion of pedal movement accurately, it is best to obtain video images of a stationary two-wheel object from the viewpoint based on the center of the object. For this purpose, we first obtain the video images where the object moves on the actual horizontal plane. Because real object images usually move horizontally on a plane with a slight tilt angle, tilt adjustment is required. To obtain the adjusted images, we detect the tilt angle in the image movement. The tilt-alignment process is as follows:

1. Detection of horizontal edges in the image:

Using the border information between the bicycle wheels and the road surface, we could detect the tilting line. To detect the border, we first detected the horizontal edges in the images using a horizontally oriented spatial 2D Gabor filter. To reduce noise and background in the images, we subtracted each output image from the consecutive frame images, as shown in Figure 5. By accumulating sequential subtracted images, the pixel values on the border became higher, as shown in Figure 6.

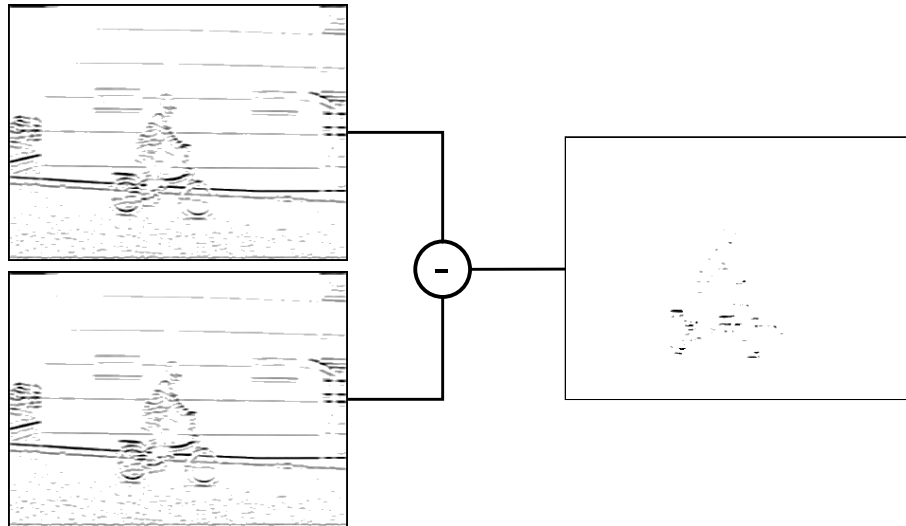


FIGURE 5. Temporal difference of spatial 2D Gabor filtered images

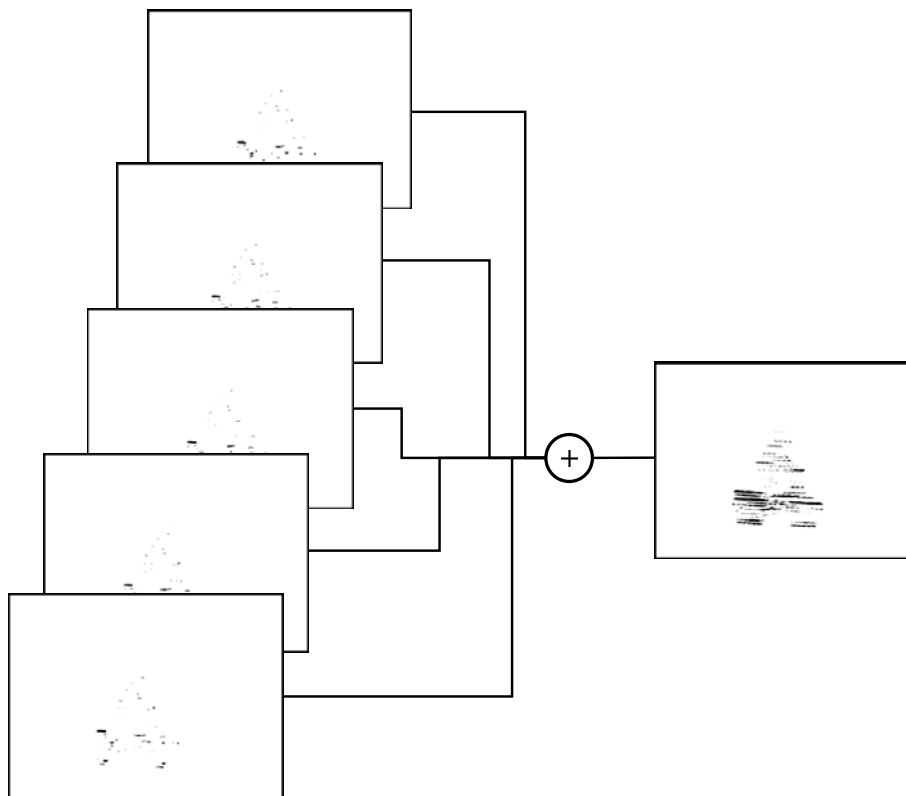


FIGURE 6. Accumulation image of temporal difference outputs

2. Reduction of boundary noise in the image:

To detect the border, boundary noise mainly due to the shadow must be reduced. We re-defined the object region by calculating projection histograms in the horizontal and vertical directions, as shown in Figure 7. First, we calculated the projection histogram in the vertical direction. If the projection values were above the predefined threshold, the region was determined as the object region with no shadow. After the object region along the horizontal axis was re-defined, the region along the vertical axis was also re-defined by the same process.

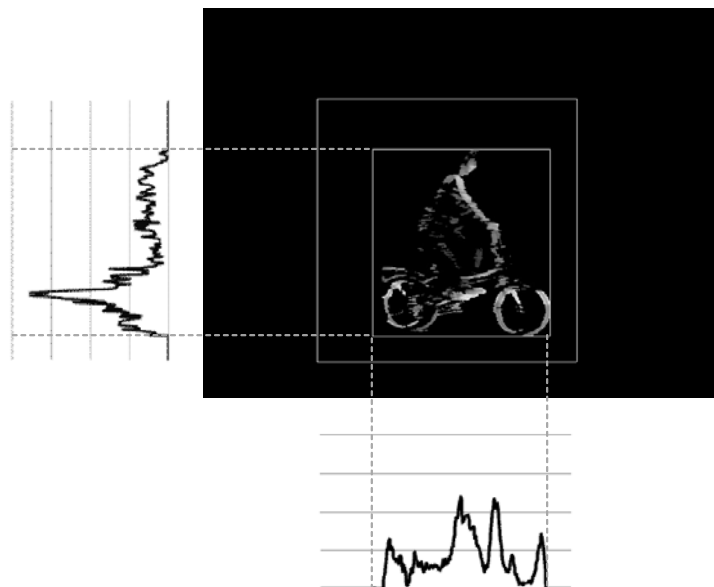


FIGURE 7. Accumulated moving-object region image

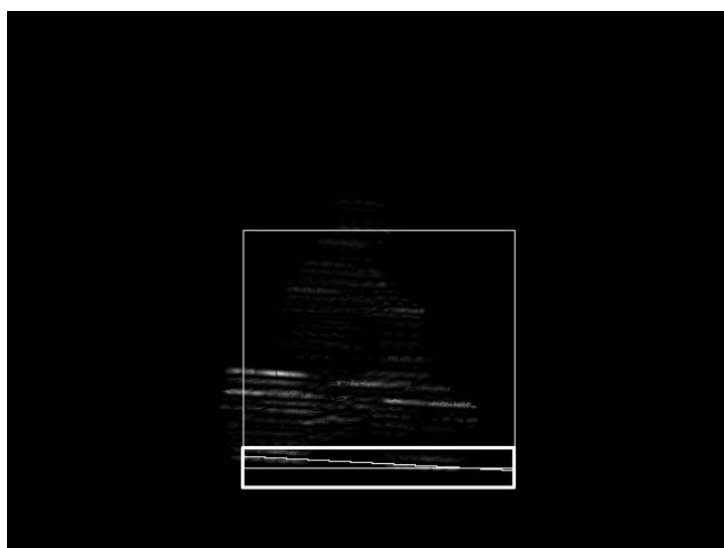


FIGURE 8. Wheels-road border detection result by Hough transform

3. Calculation of angle information based on the Hough transform:

To detect the border between bicycle wheels and the road surface, a Hough transform was applied to the predefined bottom region of the object image, as shown in Figure 8. We determined the border line as the point with the largest value in the Hough space. Then, we detected the object angle.

4. Tilt adjustment by rotating image:

The sequence of images processed with horizontally oriented spatial 2D Gabor filtering is rotated with the angle calculated by the Hough transform. Then, the object images are aligned along the actual horizontal axis. The object images in those adjusted images are segmented based on the object region width and height.

2.3. **HOG and SVM.** The extracted object is classified as a two-wheel object, pedestrian or another object using HOG features. These consist of histograms of gradient

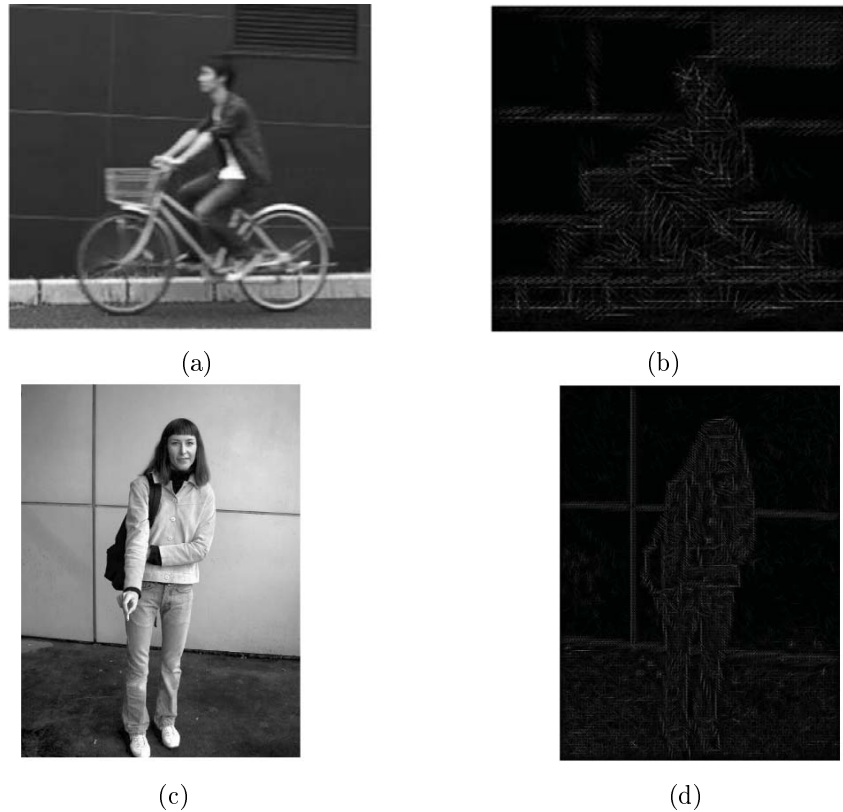


FIGURE 9. Typical outputs of HOG: (a), (c) input images, and (b), (d) HOG features

amplitude and orientation around each key point location. Orientations are quantized into a number of bins in the histogram. To provide reliable invariance against illumination, the histogram values are normalized by the total energy of all orientations. Typical outputs processed by the HOG approach in the segmented moving object is shown in Figure 9. Oriented short lines in Figures 9(b) and 9(d) are normalized local features. The luminance of these lines indicates the strength of the energy of the orientation.

The SVM, which is a two-class linear discriminator with soft margins, classifies the extracted images using HOG features. The SVM needs to be trained with a sufficient number of images of bicycles/motorbikes and pedestrians.

2.4. Relative motion detection. Images of bicycles can be detected not only by the shape but also by the characteristic leg movement during pedaling. To detect such movement, relative motion detection is required. In this work, it is assumed that the input image includes a side-view of a bicycle. We detect vertical motion during pedaling by spatiotemporal 3D Gabor filtering after removing the horizontal motion of the whole bicycle. To obtain the average velocity of leg movement, a window in which all velocities are averaged is defined. The ratios of the window width and height to the whole object-image width and height are defined as W_w and H_w , respectively. The relative coordinates of the center position of the window are defined as (w_x, w_y) .

3. Experimental Results. The proposed method was tested using numerical simulations with Intel Core i7 2.93 GHz CPU and nVIDIA GTX480 GPU. Our algorithm was implemented in C-language with an open-source computer vision library (OpenCV) for reading video streams. A convolution operation for filtering was performed using a GPU

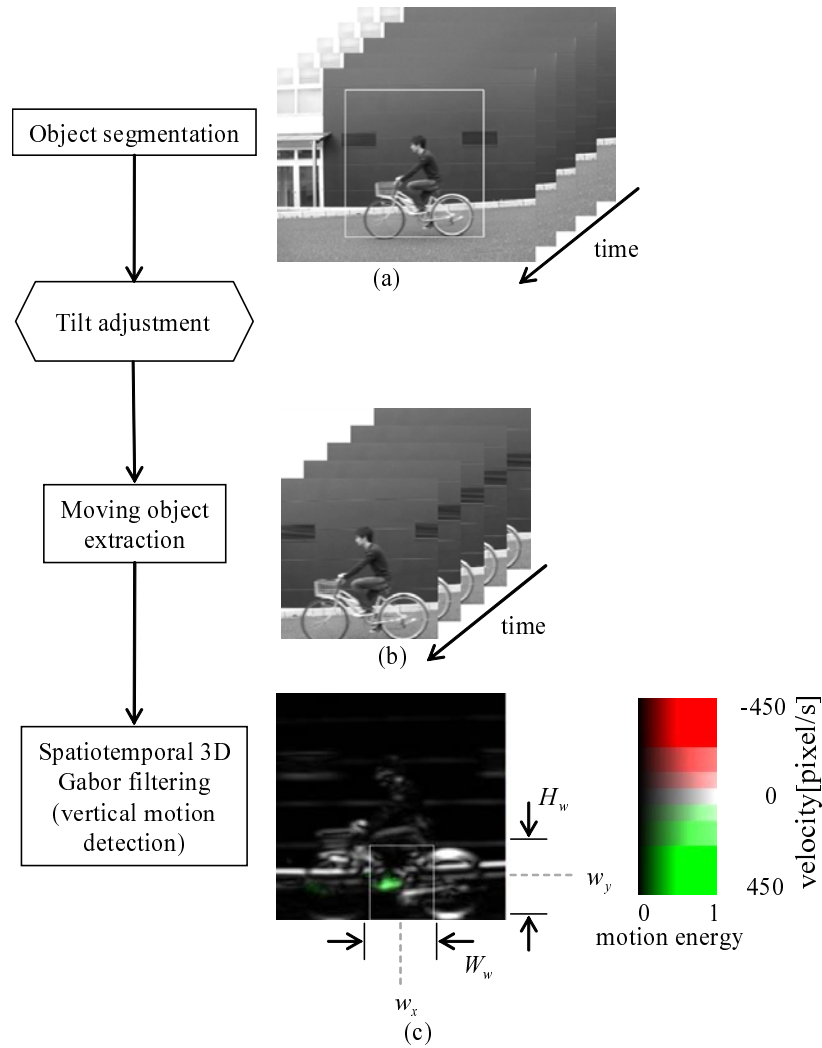
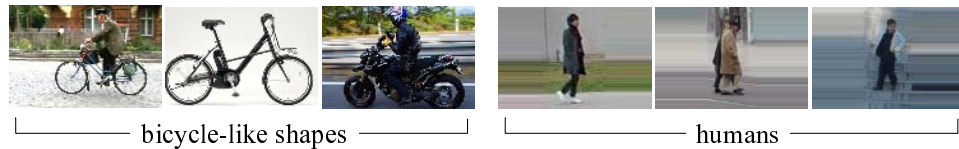


FIGURE 10. Flow of leg motion detection: (a) input image sequence with a rectangle that represents the region of a moving object, (b) segmented image sequence, and (c) spatiotemporal 3D Gabor filtering result that detects vertical movement, where the window detecting pedaling movement is indicated and the static edges are indicated by white color

processor to accelerate the processing speed by parallel processing. The GPU was controlled under a GPGPU environment (CUDA2.0) supplied by nVIDIA. The processing time per frame was as follows: 0.054 s for motion detection, 0.095 s for object region segmentation, 0.030 s for HOG feature calculation and two-wheel object detection by SVM, and 0.100 s for relative motion detection. Therefore, the total processing time per frame was 0.279 s. We used GPGPU only for convolution operation for filtering. To optimize some processes for GPGPU, we could accelerate the processing speed.

First, we evaluated the performance of the HOG/SVM approach that detects bicycle-like shapes. We used INRIA Person Database [4] and our original bicycle-like shape and pedestrian images. To train the SVM, we used 264 images for each category, and the total number of images used was 1056. A result of the discrimination task is shown in Figure 11. The image sets of bicycle-like shapes and pedestrians were classified with an average accuracy of 99.8 %.

Next, we evaluated the performance of the HOG/SVM approach that discriminates between bicycles and motorbikes. We used bicycle images included in INRIA Person



	No. for training	No. for test	Discrimination result		
			bicycle like shape	human	correct rate[%]
bicycle-like shape	264	264	263	1	99.6
human	264	264	0	264	100.0
total	528	528			99.8

FIGURE 11. Experimental result of discrimination between bicycle-like objects and pedestrians



	No. for training	No. for test	Discrimination result		
			bicycle	motorbike	correct rate[%]
bicycle	264	264	108	156	40.9
motorbike	264	264	101	163	61.7
total	528	528			51.3

FIGURE 12. Experimental result of discrimination between bicycles and motorbikes

Database and our original bicycle and motorbike images. For training the SVM, we used 264 images for each category, and the total number of images used was 1056. A result of the discrimination task is shown in Figure 12. The image sets of bicycles and motorbikes were classified with an average accuracy of 51.3 %. This result shows that it is difficult to discriminate between bicycles and motorbikes by HOG features.

Then, we evaluated the performance of leg motion detection during pedaling for bicycle and motorbike images. Image datasets of bicycles and motorbikes moving on the road were prepared. The total number of bicycle video was 30, and that of motorbike video was 50 files. Figure 13(a) shows examples of snapshots of images for moving bicycles and motorbikes used in the experiments. A spatiotemporal 3D Gabor filter where $\theta = 0$ deg and $\sigma = 4$ pixels was used for input images with four different resolutions, $\tau = \{1, 1/2, 1/4, 1/8\}$. The window parameters for detecting leg motion were set at $W_w = H_w = 1/3$ and $(w_x, w_y) = (1/2, 1/6)$, as shown in Figure 10(c). If the number of frames in which the average velocity of the leg motion exceeds the predefined threshold reaches a predefined value, we concluded that pedaling leg movement had been detected. When bicycle images were given, we defined $N(B, t)$ as the detection frequency of pedaling

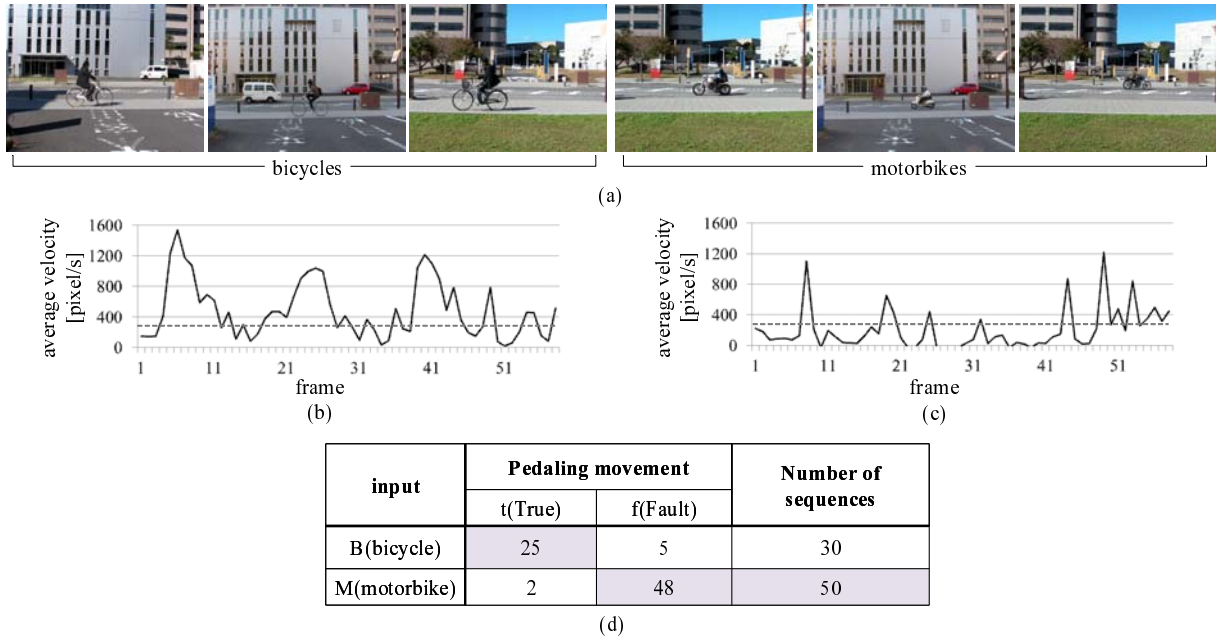


FIGURE 13. Leg motion detection results: (a) examples of snapshots of input image sequences, (b), (c) typical detected pedaling motion for a bicycle and a motorbike, and (d) bicycle detection ratio using pedaling motion

movement and $N(B, f)$ as the detection frequency of no pedaling movement. $N(M, t)$ and $N(M, f)$ are defined in the same manner for motorbike images. We calculated the bicycle detection ratio from the number of video images $N(B) + N(M)$ against $N(B, t) + N(M, f)$. Figures 13(b) and 13(c) show the average velocity change during pedaling movement on bicycle and motorbike images, respectively. Here, the threshold value for detecting pedaling movement was set at 300 pixel/s. Figure 13(d) shows the result of pedaling movement detection for each image sequence. The bicycle detection rate was 91.3%. From this result, we conclude that bicycles can be successfully discriminated from motorbikes using pedaling movement detection.

4. Discussion. We successfully detected relative vertical leg motion during pedaling in a nearly horizontally moving bicycle. This method can even detect bicycles that have motorbike-like shaped wheels such as a cruiser bicycle. It would be hard for the conventional methods focusing on the shape of wheels [8]. Therefore, the proposed method can improve bicycle detection accuracy.

As described in the previous section, for quantitative evaluation, we used 30 and 50 files of bicycle and motorbike videos, respectively. Usually, for object detection/discrimination tasks, some databases consisting of a huge number of images are commonly used. However, there are few bicycle/motorbike detection/discrimination algorithms and no standard video database for this task exists. Thus, we took movies of bicycles and motorbikes by ourselves.

As a condition for detecting pedaling movement, we need time for observing pedal movement which is about 10-15 frames (0.3-0.5 sec) according to Figure 13(b). For stable discrimination, it is preferable to use over 15 image frames. Because of applying tilt compensation, our approach can detect a bicycle with slightly tilting movement with less than about 10 degrees from the horizontal line.

Regarding computation time, the current processing time is much longer than the video rate (30 fps), although we partially used GPGPU, as described in the previous section.

However, improving the throughput will be possible using pipeline processing in dedicated parallel hardware, which may be implemented in an FPGA.

5. Conclusion. In this paper, we focused on detection of pedaling movement, which is a unique feature of a bicycle imaging. We proposed a detection algorithm for such a movement from an image sequence of a bicycle using spatiotemporal 3D Gabor filtering. The pedaling movement is a powerful key feature for discriminating a bicycle from similar objects such as a motorbike. The proposed approach will be useful for surveillance systems on the roads or in-vehicle camera systems for surrounding monitor with image recognition function.

In our future work, we will improve our algorithm for application to more general cases. For example, if bicycles move in the frontal direction, an additional process is needed. A more precise quantitative evaluation for leg movement detection performance is also required.

Acknowledgment. This work was partly supported by a grant of Knowledge Cluster Initiative from MEXT, Japan.

REFERENCES

- [1] N. Dalal, B. Triggs and C. Schmid, Human detection using oriented histograms of flow and appearance, *Proc. of European Conf. on Computer Vision*, pp.428-441, 2006.
- [2] J. Xiao, C. Yang, F. Han and H. Cheng, Vehicle and person tracking in aerial videos, *Proc. of Int. Workshop on Multimodal Technologies for Perception of Humans*, pp.203-214, 2009.
- [3] H. C. Zeng, S. H. Huang and S. H. Lai, Real-time video surveillance based on combining foreground extraction and human detection, *Proc. of Multimedia Modeling Conf.*, pp.70-79, 2008.
- [4] N. Dalal and B. Triggs, Histograms of oriented gradients for human detection, *Proc. of Conf. on Computer Vision and Pattern Recognition*, pp.886-893, 2005.
- [5] G. Somasundaram, V. Morellas and N. Papanikolopoulos, Counting pedestrians and bicycles in traffic scenes, *Proc. of IEEE Intelligent Transportation Systems*, pp.299-304, 2009.
- [6] Z. Qui, D. Yao, Y. Zhang, D. Ma and X. Liu, The study of the detection of pedestrian and bicycle using image processing, *Proc. of IEEE Intelligent Transportation Systems*, vol.1, pp.340-345, 2003.
- [7] S. Rogers and N. P. Papanikolopoulos, Counting bicycles using computer vision, *Proc. of IEEE Intelligent Transportation Systems*, pp.33-38, 2000.
- [8] S. Messelodi, C. M. Modena and G. Cattoni, Vision-based bicycle/motorcycle classification, *Pattern Recogn. Lett.*, vol.28, pp.1719-1726, 2007.
- [9] T. T. Zin, H. Takahashi and H. Hama, Robust person detection in far infrared images – Methods based on multi-slits and GC movement patterns –, *International Journal Innovative Computing, Information and Control*, vol.5, no.3, pp.751-761, 2009.
- [10] K. Takahashi, Y. Kuriya and T. Morie, Bicycle detection using pedaling movement by spatiotemporal Gabor filtering, *Int. Tech. Conf. of IEEE Region*, vol.10, pp.918-922, 2010.
- [11] B. D. Lucas and T. Kanade, An iterative image registration technique with an application to stereo vision, *Proc. of Int. Joint Conf. Artificial Intelligence*, vol.3, pp.674-679, 1981.
- [12] T. Camus and H. B. Bühlhoff, Real-time optical flow extended in time, *Max-Planck-Institute Für Biologische Kybernetik, Tech. Rep.*, no.13, pp.1-17, 1995.
- [13] E. H. Adelson and J. R. Bergen, Spatiotemporal energy models for the perception of motion, *J. Opt. Soc. Am.*, vol.A2, no.2, pp.284-299, 1985.
- [14] D. Heeger, Optical flow using spatiotemporal filters, *Int. J. Computer Vision*, vol.1, no.4, pp.279-302, 1988.
- [15] T. Gautama and M. Van Hulle, A phase-based approach to the estimation of the optical flow field using spatial filtering, *IEEE Trans. Neural Networks*, vol.13, no.5, pp.1127-1136, 2002.
- [16] J. Daugman, Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters, *J. Opt. Soc. Am.*, vol.A2, pp.1160-1169, 1985.
- [17] A. Hanazawa, Real-time multi-resolution visual motion detection by spatiotemporal energy model implemented on GPU, *Proc. of Int. Workshop on Computational Intelligence and Applications*, pp.213-218, 2009.