PhD Thesis

Vehicle Detection Based on Spatial Saliency and Local Image Features in H.265 (HEVC) 4K Video and Evaluation Model for Quality of Detection Accuracy

By

MOST. SHELINA AKTAR

Doctoral thesis paper for the degree of
Doctor of Philosophy

Advisor: ***Professor Dr. Yuukou Horita***

Graduate School of Science and Engineering for Education
Faculty of Engineering
University of Toyama
Toyama, Japan

ABSTRACT

In order to realize a safe and secure road transportation system, research on intelligent transportation systems (ITS) is widely conducted. In the optimization and management of traffic, technology for detecting vehicles is important, and research on detecting objects using information obtained from images, still images, and sensors has been widely conducted. In this study, one of the main challenges is to develop vehicle detection. Most of the existing visual saliency models use the input images, in which salient objects are to be detected, are free from complex background and overlapping areas. Moreover, they are very sensitive to the complex scene and different illuminations. They cannot detect their interest objects from the input video. This study develops a vehicle detection method by using spatial saliency and local image features. The Scale Invariant Feature Transform (SIFT) and Harris features in combination with spatial saliency model play an important role to detect vehicle from the scene. One-to-one symmetric search is performed on the descriptors to select a set of matched interest point pairs for vehicle detection. The one-to-one symmetric search on the descriptors is useful for detection of the interest object in the context of saliency detection. We use 4K video of a road scene with different types of vehicles. The propose method is able to detect desired overlapping objects from the road scene without heavy computation like other training based methods. In the second, the detection performance is analyzed with another saliency based methods. Our methods have better performance as compared to the other conventional methods.

In the images/videos based applications over internet are typically stored in the compressed domain such as MPEG2, H.264, MPEG4, since they can reduce the storage space and greatly increase the delivering speed for Internet users. Most of the systems require transmission of data to some central server and have to deal with some issues such as limited bandwidth and quality. Consequently, they require to transmit videos with a reasonable high-quality in compressed domain for further processing by vision-based systems, such as person identification, fraud detection, and vehicle detection for road monitoring. Furthermore, existing saliency detection models are implemented in uncompressed domain and lack of analysis their performance. Therefore, there still have challenging research issues to detect interest objects with the conventional saliency based methods, and determined the reasonable high-quality video in compressed domain. From these contexts, we analyze the proposed detection method in compressed domain, and it shows better result in compared with conventional methods and single feature based detection.

During vehicle detection, it is necessary to know the correct vehicle position considered as “ground truth” in order to evaluate the vehicle detection method. For this reason, many detection models define areas of the targeted object, where people considered areas of the objects. In many studies, the ground truth is represented by a rectangle. We consider the relationship between Intersection over Union (IoU) and subjective vehicle detection by considering shifted from the ground truth position. In this study, subjective evaluation experiments have been carried out with respect to misalignment from ground truth in vehicle detection. We also investigate subjective evaluation model with respect far and near view in vehicle detection. Based on the experimental results, we see that there is a significant difference in left and right misalignment even if the Intersection over Union (IoU) value was the same. Finally, we propose indices considering subjective evaluation model in vehicle detection utilizing IoU.

ACKNOWLEDGEMENTS

First of all, I would like to express my sincere gratitude to my advisor ***Professor Yuukou Horita*** for his patience, motivation, immense knowledge, and continued support of my Ph.D. study and related research. His guidance helped me in all the time of research and writing of this thesis. I appreciate all his contributions of time, and ideas to make my Ph.D. experience productive and stimulating. I could not have imagined having a better advisor and mentor for my Ph.D. study.

I would like to express my gratitude to other members of our Laboratory for their continued support and all office staffs of University of Toyama for their assistance and cooperation. They have contributed immensely to my personal and professional time at University of Toyama. The group has been a source of friendships as well as good advice and collaboration. I would like to offer my special thanks to ***Koutaro Katayama, Naho Ito*** and ***Takano Shina*** for their support throughout my research work.

A very special gratitude goes out to the **Nikki-SaneYoshi** and **Nakamura Sekizenkai Scholarship** for providing me the financial support without which this research would have been impossible.

I am deeply grateful to my husband ***K.M Ibrahim Khalilullah*** *and my son* ***K.M Sadik Wadud***, for their caring, patience and continuous support throughout the years. I also thanks to my parents ***Most. Feroza Begum*** and ***Md. Shamsul Haque*** and my two little brothers and one elder sister for their strong encouragement in pursuing my research and for giving me the strength to succeed.

Dedicated to my parents, my husband
K. M. Ibrahim Khalilullah, and my son
K. M. Sadik Wadud

Table of Contents

Chapter 1

INTRODUCTION

Chapter 2

SPATIAL SALIENCY AND LOCAL IMAGE FEATURES BASED VEHICLE DETECTION FROM 4K IMAGE SEQUENCES

Chapter 3

PERFORMANCE ANALYSIS OF VEHICLE DETECTION IN COMPRESSED AND UNCOMPRESSED DOMAIN

Chapter 4

CONSTRUCTION OF SUBJECTIVE VEHICLE DETECTION EVALUATION MODEL CONSIDERING SHIFT FROM GROUND TRUTH POSITION

Chapter 5

HUMAN PERCEPTION AND IOU BASED ON VARIATION OF ANGLES AND DISTANCE

Chapter 6

List of Tables

List of Figures

Chapter 1
INTRODUCTION

1.1 Background of this research

The number of vehicles is increasing day by day, which makes a lot of pressure on roads capacity and infrastructure. As a result, he traffic hazards increases at a high rate and the traffic management is too difficult, moreover it causes a huge loss of life and property due to the road accidents. Automated vehicle detection systems can contribute significantly to reduce the harmful side effects of traffic, as a part of many traffic applications such as road monitoring, road traffic control, traffic response system, traffic signal controller, lane-departure warning system, automatic vehicle accident detection and automatic traffic density estimation [1].

Compared to still images, video processing techniques have attracted researchers for vehicle detection [2-7] in the recent years. Video-based vehicle detection has played a major role in real-time traffic management systems over the past years. Traffic surveillance is one of the significant applications of video-based supervision systems. Many researchers have investigated in the Vision-Based Intelligent Transportation System (ITS), transportation planning and traffic engineering applications to extract useful and precise traffic information such as vehicle count, vehicle trajectory, vehicle tracking, vehicle flow, vehicle classification, traffic density, vehicle velocity, traffic lane changes, license plate recognition, etc. [7-11].

The long-term scope of our research is to develop a vehicle detection method using image processing and construct a performance evaluation model for improving detection quality accuracy. In this research work, saliency model combined with local features is utilized for vehicle detection. Various methods have already been proposed for saliency detection over the past few years. A hierarchical approach for saliency-based visual attention is proposed by Itti et al. [12]. Their algorithm obtains the saliency map based on local and global information by global non-linear normalization and iterative filtering with the difference of Gaussian. However, these normalization and iterative filtering procedures are effective good even though their normalization schemes provide some global reasoning, since each map is normalized regardless of the statistical information of other maps. E. Rahtu et al. [13] introduced a salient object segmentation method based on combining a saliency measure with a conditional random field. The proposed saliency measure is developed using a statistical framework and local feature. N. Vikram et al. [14] proposed a simple and efficient method to compute an image saliency map based on computing local saliencies over random rectangular regions of interest. However, they calculate saliency map of each pixel independently, so the correlation between pixels is not considered properly. In addition, how to properly use the sampling strategy to help the detection task remains a problem. N. Imamoglu et al. [15] proposed a saliency detection model by introducing local and global saliency map in frequency domain. W. Wang et al. [16] developed a visual saliency detection method based on region descriptors and prior knowledge. Region descriptors and prior knowledge were introduced in their proposed method to provide more accurate visual cues. Spectral analysis and multiresolution image processing based approaches were proposed in [17,18]. N.D.B. Bruce and J. K. Tsotsos [19] presented entropy-based approach based on the principle of maximizing information sampled from a scene by employing independent component analysis. A discriminant center-surround entropy contrast based saliency map was proposed in [20]. In the center-surround entropy based method, selecting an appropriate window size is a challenging issue for obtaining a higher quality saliency map. To overcome this issue, many machine learning based approaches were proposed. W.Kienzle et al. [21] proposed a nonparametric approach to visual saliency using Support Vector Machine (SVM). A visual saliency detection method is proposed by H. J. Seo and P.Milanfar [22] for automatic target detection and image quality assessment. However, their method performs on simple image database without overlapping. J. Harel et al. [23] proposed a graph-based visual saliency approach based on a Markovian representation of the feature maps. It is observed from the literature review that most of the existing approaches process pixels of the input images sequentially with fixed sliding windows. However, target objects or salient regions can come with different shapes, scales, and arbitrary positions. The outdoor scene is also combined with complex background and different types of objects. As a result, some regions of the outdoor scene might be detected by the visual saliency models, which are not the salient or target regions according to the user needs. In addition, most of the existing saliency detection models are performed their tasks on the simple image databases, which are simple in the sense of contents, background, and low-resolution. The background is complex when compared to other visual saliency based detection. In this study, we propose a spatial saliency model combined with SIFT and Harris features for vehicle detection from outdoor 4K video, and analysis detection performance in compressed domain. The main advantage of this study involves combining spatial saliency model with the SIFT and Harris features to detect the desired objects and separating the overlapping objects from complex background using one-to-one symmetric search.

Another contribution is to analyses the detection performance of the proposed model in H.265 4k video for developing a relationship between IoU and subjective evaluation value. The object detection method proposed by Cheng et al. have succeeded in separating the target object region and the non-target object region at the pixel level [24]. Borji et al. published an object region detection method and dataset using a saliency map [25]. In object detection research, the region considered a vehicle is labelled using a rectangle, or the target region and non-target region are labelled at the pixel level. According to the reference [24], if the area considered as an object is a rectangle, it is too rough for detailed evaluation in knowledge, and they states that segmentation at the pixel level is important. In vehicle detection research, the region considered a vehicle is often defined by a rectangle, and data sets and detection methods that divide the target region at the pixel level have been released. Geiger et al. have released a publicly available data set for vehicle detection, and defined the area considered as a vehicle in three dimensions [26]. Geiger et al. and Cordts et al. have published data sets that define the area considered vehicles at the pixel level [27, 28].

Figure 1.1 shows an example of a region that is considered a vehicle. Figure 1.1 shows the area considered as a vehicle in various shapes and sizes. It takes from a still image. As shown in Figure 1.1, there are various ways of thinking about the area considered as a vehicle.

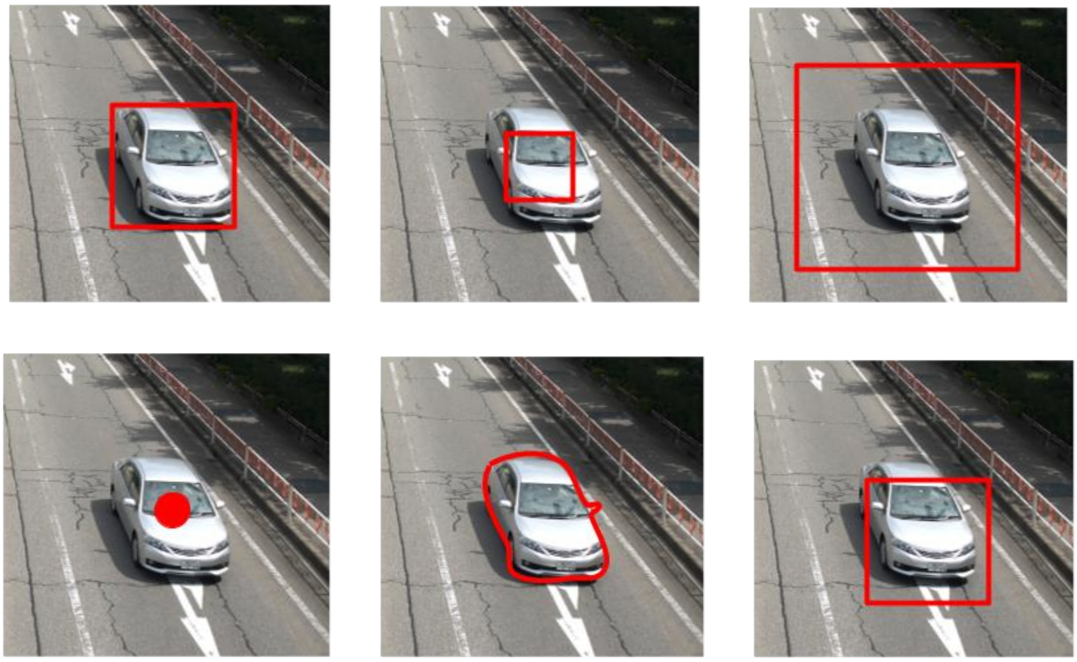
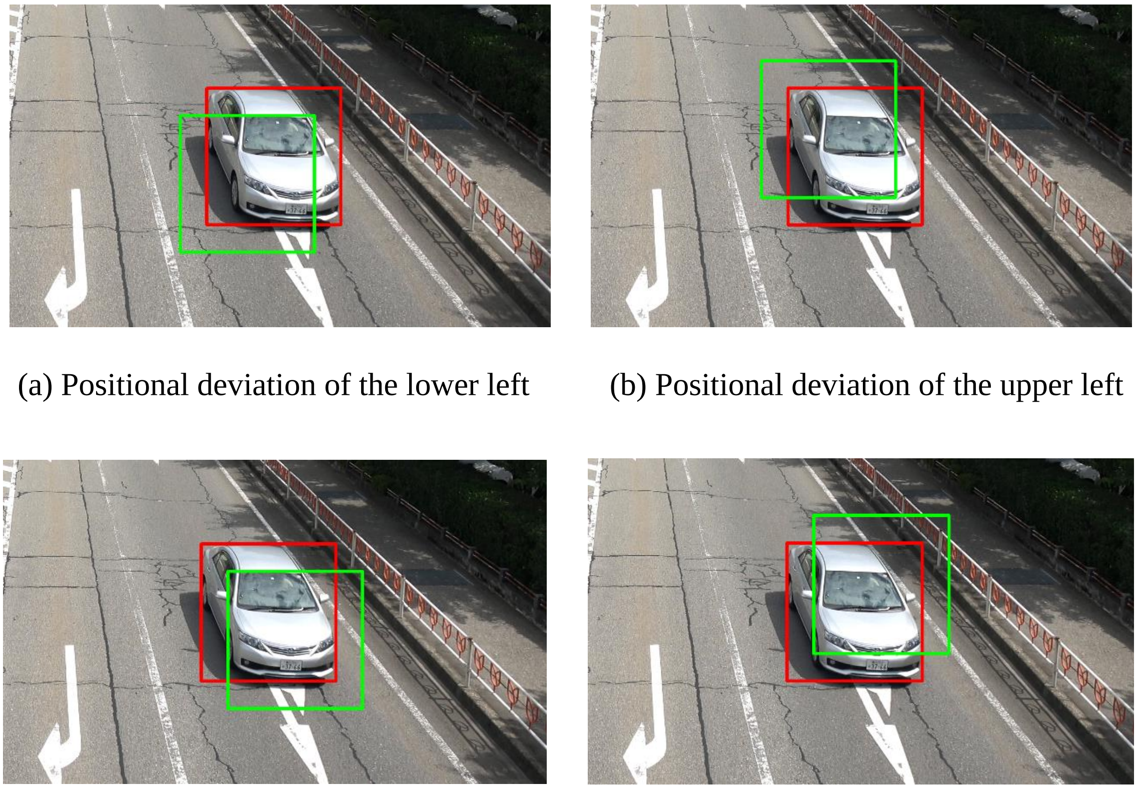


Figure 1.1: Example of various vehicle detection.

When a computer detects a vehicle region, it first outputs all regions considered as objects in the still image. Then, the area of the vehicle is narrowed down from the area considered as an object. In order to narrow down the vehicle area, it is necessary to define the vehicle area. However, the output of the area considered as a vehicle in image processing is not clear. For this reason, many vehicle detection models define areas, where people considered vehicles using rectangle. From the above, it can be considered that the design of the vehicle detection model depends on people.

Intersection over Union (IoU) [29] is a widely used evaluation method for vehicle detection. IoU is a method that evaluates the area ratio of overlapping regions, and is widely used in the threshold of the region considered as a vehicle in the vehicle detection model and in the performance evaluation of the vehicle detection model. An example of vehicle detection is shown in Figure 1.2. The red rectangle in Figure 1.2 is the vehicle region defined by the person, and the green rectangle is the vehicle region predicted by the vehicle detection model. From Figure 1.2 (a) to Figure 1.2 (d), the ratio of the area where the region defined by the person and the region predicted by the vehicle detection model overlap is the same. Therefore, the evaluation value in IoU is the same. However, the area



(a) Positional deviation of the lower left

(b) Positional deviation of the upper left

(c) Positional deviation of the lower right

(d) Positional deviation of the upper right

Figure 1.2: Examples of the detection results when posiitonal displacement of the vehicle detection occurs.

predicted by the vehicle detection model is not always the same as the subjective evaluation because the positions are shifted in various directions with respect to the area defined by the person. Therefore, it is necessary to analysis human perception based evaluation and IoU.

1.2 Objectives

The main objective of this study is to develop a vehicle detection method using image processing without utilizing machine learning for relaxing heavy computation, and to construct a subjective vehicle detection model by developing a relationship between human perceptions based evaluation and IoU. In this study, we propose a vehicle detection method based on visual saliency and combination of local image features from 4k video.

Another objective of this study is to propose a vehicle detection evaluation index that takes into account subjective evaluation by investigating subjective evaluation in vehicle detection and investigating the relationship with IoU. We also investigate the relationship between subjective evaluation based on the direction of deviation in vehicle detection, the distance of the target vehicle, and the color of the vehicle, and IoU, which lead us to develop human perception based evaluation model for overcoming limitations of IoU.

1.3 Scope of the research

It is clear from the discussion above that without a vehicle detection the current study and its findings would have not been possible. Since vehicle detection is important, one part of this work deals with the detection of vehicle. Another part of this work is to develop evaluation method for detection performance analyses and comparison with another method. In order to evaluate the vehicle detection method, it is necessary to know the correct vehicle position considered as "ground truth". Finally, a subjective detection evaluation model is developed considering shift from ground truth position. It is easy to be applied for any objects detection evaluation methods. The analysis and developed evaluation model maybe utilizes in industrial application instead of IoU. The main contributions of this research are summarized below:

- The developed method detects vehicle on a road from the image sequences of a video.
- Combination of saliency model with SIFT and Harris features for vehicle detection.
- It can detect the desired objects and separating the overlapping objects from complex background using image processing by combining spatial saliency model and local features.
- Analysis of subjective evaluation model and IoU for detection
- Developed human perception based evaluation method for overcoming the limitations of IoU.

1.4 Structure of the thesis paper

The structure of the thesis is organized with literature review, methodological steps of the proposed methods, and their performance analysis. Each of the proposed methods are described in a separate chapter. A short description of each chapter is provided here to ease the understanding and use of the thesis.

Chapter 1 discusses the background, motivation, objectives, and contribution of this research work.

Chapter 2 presents the proposed vehicle detection method by combining spatial saliency and local features.

Chapter 3 describes the performance analysis in compressed and uncompressed domain.

Chapter 4 discusses about the vehicle detection subjective evaluation model with various misalignment with ground truth rectangles.

Chapter 5 analyzes the human perception based evaluation and IoU on various angle and distance.

Chapter 6 presents conclusions of the research works. The further direction for the development of this research is also proposed.

Chapter 2

SPATIAL SALIENCY AND LOCAL IMAGE FEATURES BASED VEHICLE DETECTION FROM 4K IMAGE SEQUENCES

2.1 Introduction

A vehicle detection method fusing spatial saliency and local image features is presented in this chapter. Most of the existing visual saliency models use the input images, in which salient objects are to be detected, are free from complex background and overlapping areas. Moreover, they are very sensitive to the complex scene and different illuminations. In this method, the Scale Invariant Feature Transform (SIFT) and Harris features in combination with spatial saliency model play an important role to detect vehicle from the scene. One-to-one symmetric search is performed on the descriptors to select a set of matched interest point pairs for vehicle detection. We use 4K video of a road scene with different types of vehicles. The propose method is able to detect desired overlapping objects from the road scene without heavy computation like other training based methods.

2.2 What is visual saliency

It is widely known that humans cannot perceive every detail on an entire scene at first sight. Human visual system works as a filter to allocate more attention to the attractive and interesting regions or objects according to their saliency. Humans can exhibit visual fixation, which is maintaining of the visual attention on a single location. Koch et al. [30] defined saliency as the distinctive perceptual quality, which makes some objects in the world stand out from their surroundings and immediately take our attention. It is one of the classical ways to find the regions of interest in the image.

A salience computational model refers to how visual features such as color, orientation, luminance and motion are combined into a single global map representing the relative ‘salience’ of each point on the map. The concept of the saliency map was originally proposed by Koch & Ullman [1] and was later implemented by Itti *et al.* [12, 31].

2.3 Image features

Typically, the process of the visual saliency model may be organized into three steps: feature extraction, activation map, and normalization/combination. In feature extraction stage, we want to ultimately highlight vehicle’s location in the image, where features are more stable even under changes in image scale, rotation, noise and illumination. In this study, we utilize motion features, SIFT and Harris descriptors, which are discussed in the following section:

2.3.1 Motion feature

Given an input frame I, an intensity image is computed as,

$$I = (r + g + b)/3, \tag{2.1}$$

where r, g, and b denote red, green, and blue channels of the input frame. The first processing step consists of decomposing it using dyadic Gaussian pyramids as in [12], which progressively low-pass filter and subsample the input frame. After that, the motion feature is computed from spatially-shifted differences between Gabor pyramids from the current and previous frames. Local orientation information is obtained from I using oriented Gabor pyramids $O_n(\delta, \theta)$, where δ represents scale and θ is the preferred orientation. For motion feature computation, the local orientation information is used, and only shifts of one pixel orthogonal to the Gabor orientation are considered, yielding one shifted pyramid $S_n(\delta, \theta)$ for each Gabor pyramid $O_n(\delta, \theta)$. The Reichardt [32] model is then used to compute the motion feature:

$$R_n = |O_n(\delta, \theta) * S_{n-1}(\delta, \theta) - O_{n-1}(\delta, \theta) * S_n(\delta, \theta)| \tag{2.2}$$

where * denotes a point-wise product and R_n is motion feature image of the n^{th} Gabor pyramid.

2.3.2 SIFT and Harris features

The SIFT is an algorithm in computer vision to extract and describe local features in images. This algorithm is presented in [33]. The important characteristic of the SIFT features is that the relative positions of objects in the original scene should not change from one image to another. As the SIFT feature descriptor is invariant to scaling, orientation, illumination changes, and partially invariant to affine distortion, it can robustly identify objects from the outdoor scene under partial occlusion and clutter environments. To make a database, SIFT keypoints and their descriptors of the vehicles are extracted from the Gaussian pyramids of some reference 4K image sequences.

From our empirical evidence, we observe that SIFT algorithm cannot detect corner points in some image sequences. However, the corner points are more discriminative and stable feature that can be matched well in spite of changes in viewing conditions [21, 34]. Besides, the Harris corner detector is an acceptable starting point for the computation of scale-positions and affine invariant features. To take large number of features, some other keypoints of the vehicles and their descriptors are extracted using Harris corner detector and SIFT algorithm, respectively. The SIFT and Harris features are used to highlight significant locations of the objects in outdoor scene.

2.4 Activation map and normalization

After feature extraction process, our goal is to compute an activation map, such that locations on the image *I* is somehow unusual in its neighborhood will correspond to high values of the activation map [23]. Then, normalization and combination operations are performed on the activation maps to make a single saliency map. Both activation and normalization operations use Markov chain interpretation of the image [14].

2.5 System flow diagram

In this work, the SIFT and Harris features are stored in two separate databases called SIFT feature database and Harris feature database, respectively. During detection, the motion, SIFT, and Harris features are extracted from an input image. After that, the SIFT and Harris features on the vehicle regions are detected using feature matching technique. In feature matching technique, nearest neighbor search algorithm is utilized for matching the feature database. Morphological operations are applied to enhance the vehicle regions and compute bounding boxes of that regions from the SIFT and Harris feature images. Saliency map is obtained by applying Markovian approach on all of the feature images. Bounding boxes of the saliency map is computed again using thresholding and morphological operations. Combining all of the bounding boxes from SIFT, Harris, and saliency map, we usually get multiple overlapping bounding boxes for each vehicle. We use a greedy procedure for eliminating repeated bounding boxes via non-maximum suppression filtering technique. We sort the bounding boxes by the bottom-right y-coordinate of the bounding box, and greedily select the largest coordinates for the start of the bounding box and the smallest coordinates for the end of the bounding box while skipping bounding boxes that are at least 50\% covered by a bounding box of a previously selected box. The overall structure of the propose approach is shown in Figure 2.1.

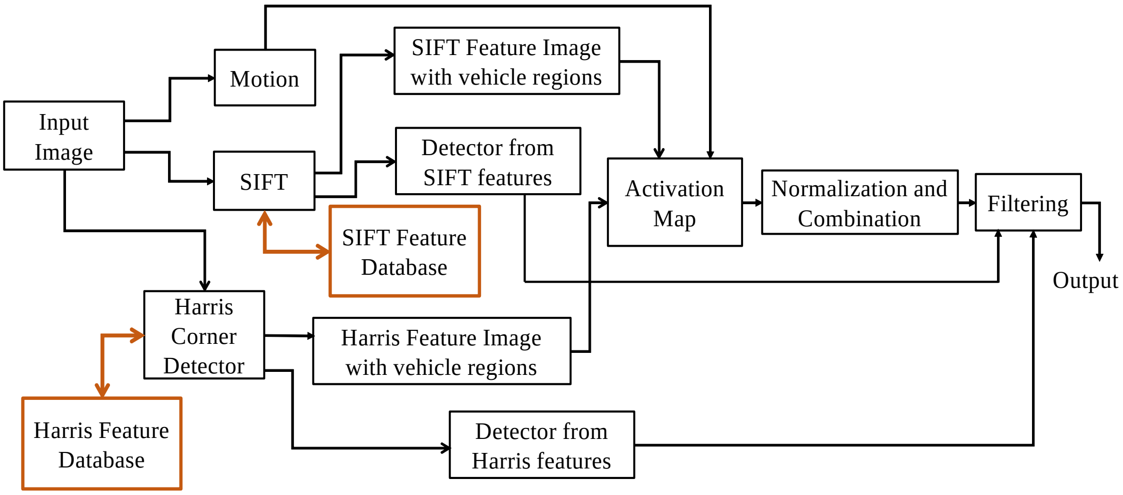
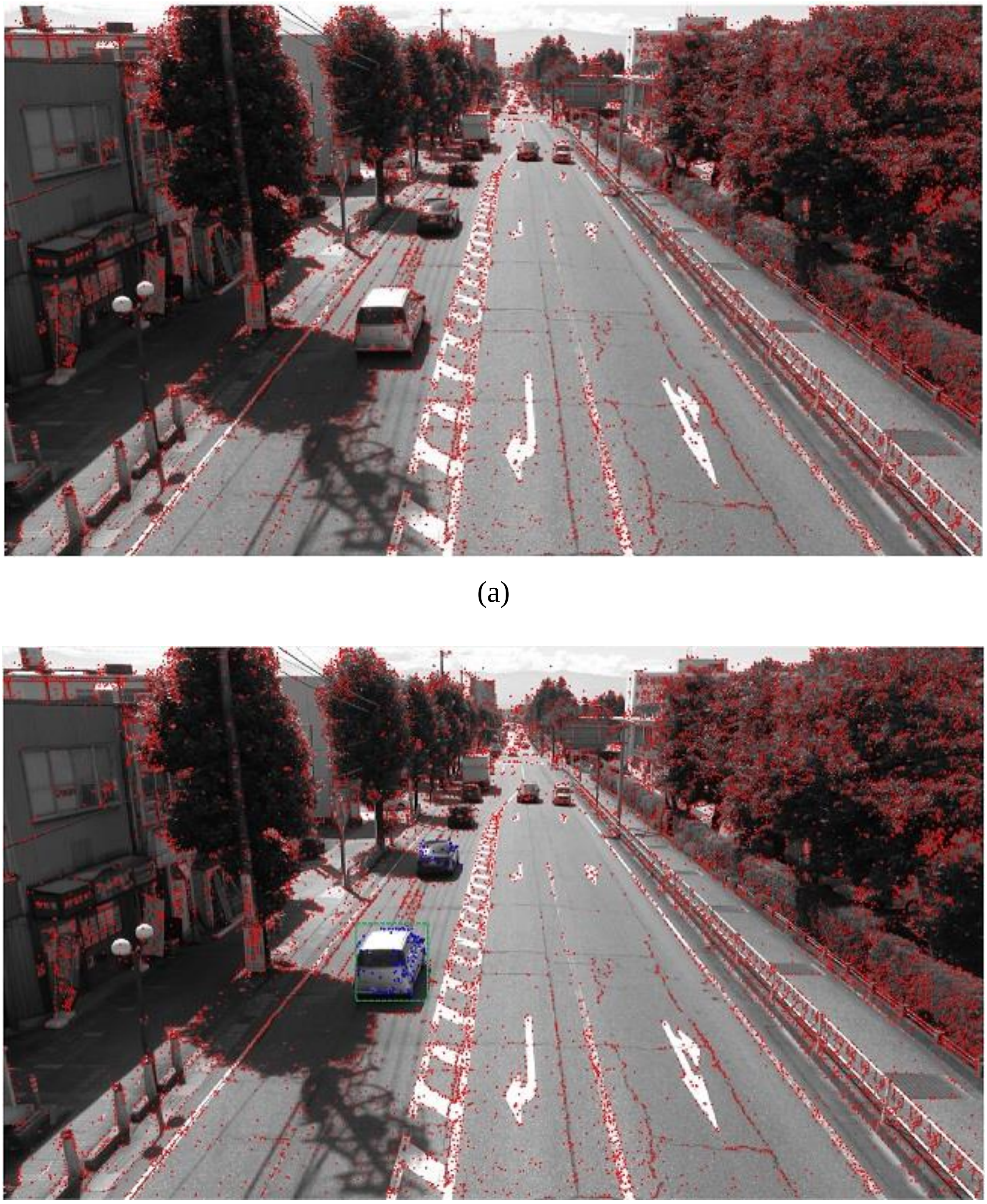

Input Image
Motion
SIFT
SIFT Feature Image with vehicle regions
Detector from SIFT features
SIFT Feature Database
Activation Map
Normalization and Combination
Filtering
Output
Harris Corner Detector
Harris Feature Image with vehicle regions
Harris Feature Database
Detector from Harris features

Figure 2.1: Overview of the proposed model

2.6 Data collection and pre-processing

SIFT and Harris feature detectors are used to detect key points of the arbitrarily selected images from a video. From the detected key points, we select feature points of the vehicle region using mouse event to create feature database. The following figure, Figure 2.2, represent the procedure of the database creation from SIFT feature detector. The left image represents detected key points and the right image represents selection procedure. The selected feature points of the vehicle region are marked using blue color. The Harris feature database is created using the same procedure.



(a)

(b)

Figure 2.2: Procedure of the SIFT feature database; (a) SIFT key points (b) Selected feature points using mouse event.

2.7 Experimental results

The proposed method is implemented using Matlab on Windows platform in CPU. In order to test the performance of the proposed model, we used sample outdoor video clip of traffic scene taken by a SONY FDR-AX55, 4K-UHD video camera. The frame rate of the captured video is 30 fps. All video frames are extracted from the captured video. The total number of video frames is 1500. Among them, 100 frames are randomly chosen to extract SIFT and Harris features, which are used for feature database. A sequence of frames from 1000 to 1100 is used to show the experimental results. The size of each frame is 3840×2160 pixels.

2.7.1 Performance evaluation

To test the detection performance, we manually created ground truth images. Some ground truth images and their corresponding binarized visual saliency maps are shown in Figure 2.3. A ground truth image from the sequences, corresponding binarized saliency map, and detection result are shown in Figure 2.4.

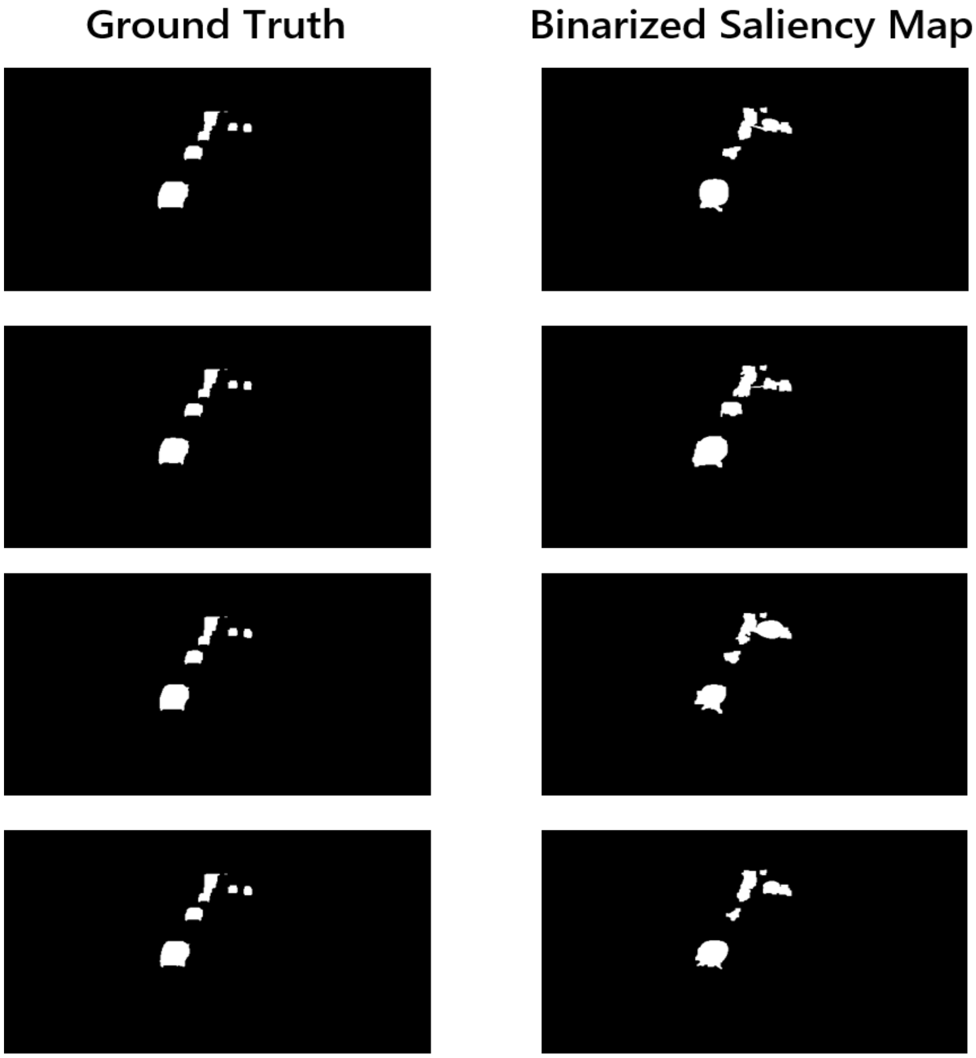
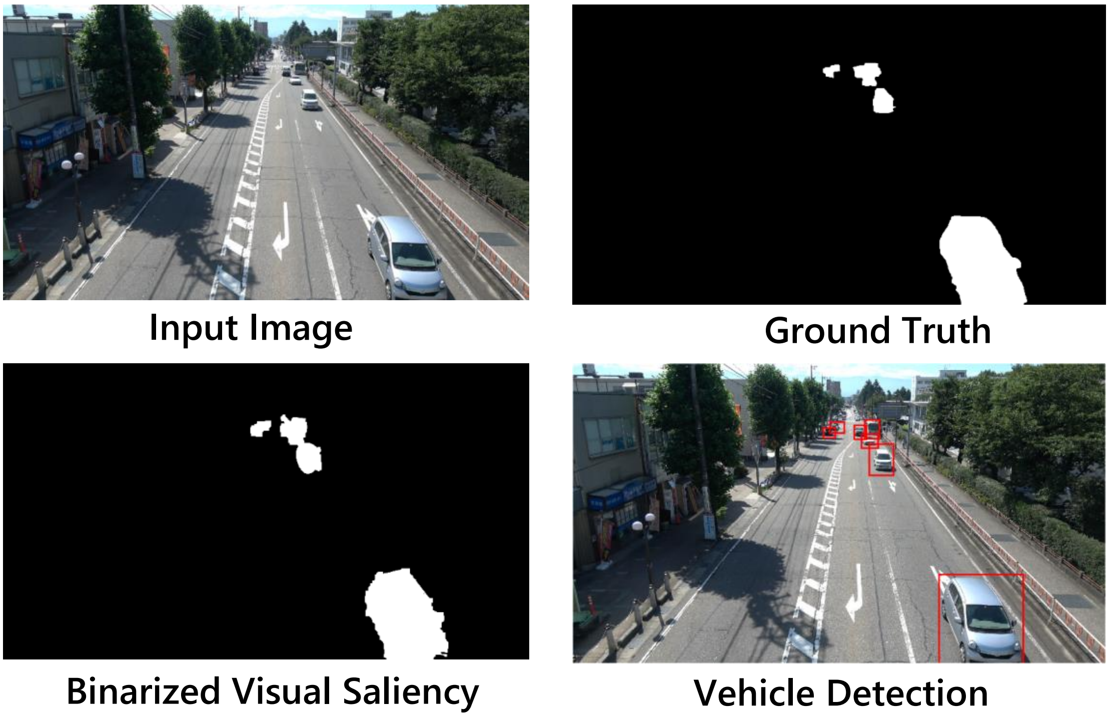


Figure 2.3: Ground truth and their binary saliency maps



Input Image

Ground Truth

Binarized Visual Saliency

Vehicle Detection

Figure 2.4: Vehicle detection

2.7.2 Evaluation method

In order to quantitatively evaluate the performance for the task of vehicle detection, we calculate pixel-based ROC-AUC and *F*-measure (F_1 and F_β) metric as recommended in [35], where the saliency map is binarized and compared with the ground truth mask. The ROC curve is created by plotting False Positive Rates(FPRs) and True Positive Rates (TPRs), and the AUC is calculated. The AUC will always be between 0 and 1.0. All of the evaluation metrics are computed using the following equations:

$$PRECISION = \frac{TP}{TP+FP}, \tag{2.3}$$

$$RECALL = \frac{TP}{TP+FN}, \tag{2.4}$$

$$TPR = \frac{TP}{TP+FP}, \tag{2.5}$$

$$FPR = \frac{FP}{FP+TN}, \tag{2.6}$$

$$F_1 = \frac{2 \times PRECISION \times RECALL}{PRECISON + RECALL}, \tag{2.7}$$

$$F_\beta = \frac{(1+\beta)^2 \times PRECISION \times RECALL}{\beta^2 \times PRECISON + RECALL}, \tag{2.8}$$

where *TP* is the number of true positives, *TN* is the number of true negatives, *FP* is the number of false positives and *FN* is the number of false negatives. During performance evaluation, we consider β^{2}=0.3. The ROC-AUC and *F*-measure (F_1 and F_β) performance of the saliency map on averages over all images (mean statistics) are given in Table 2.1.

Table 2.1: Mean Statistics

ROC-AUC	$\boldsymbol{F_1}$	$\boldsymbol{F_\beta}$
0.91	0.77	0.75

Figure 2.5 and Figure 2.6 represent the performance of the proposed approach in terms of *F*-measure (F_1 and F_β) and ROC-AUC, respectively.

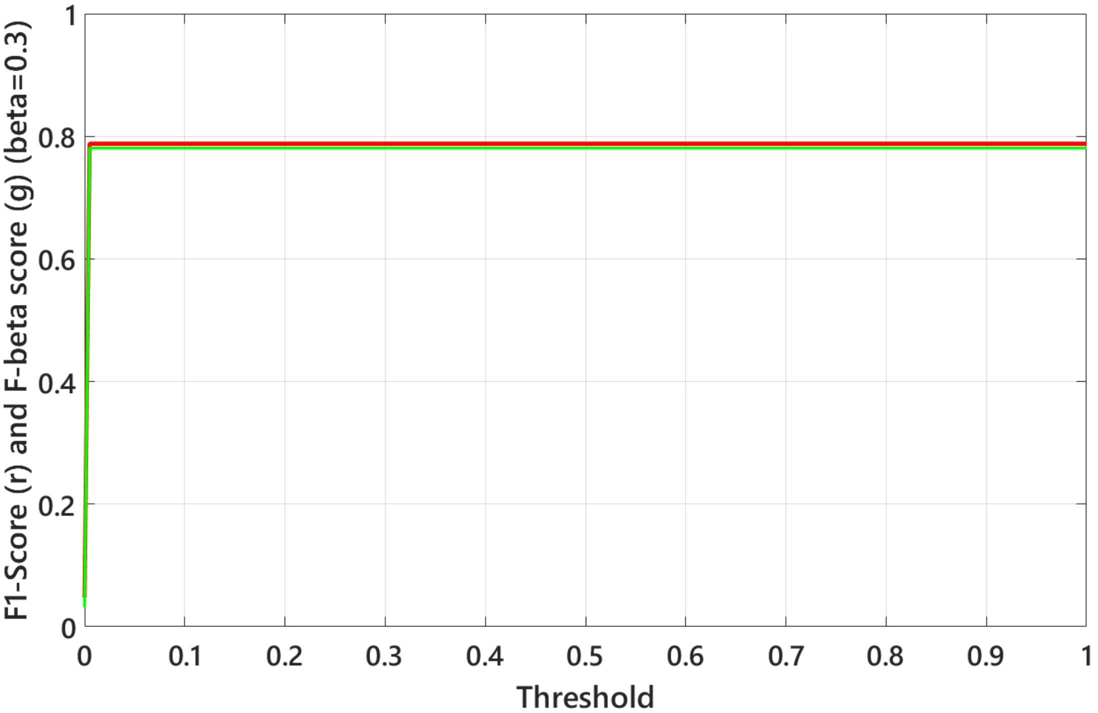


Figure 2.5: Performance evaluation using *F*-measure (F_1 and F_β).

J. Harel et. al. [23] and T. N. Vikram et. al. [14] did not explain about overlapping salient object detection. As compared to them, the propose approach detects overlapping salient object from 4K image sequences. Figure 2.7 represents some detection results of the proposed model. The results show that the propose training free method can detect overlapping vehicles from 4K image sequences.

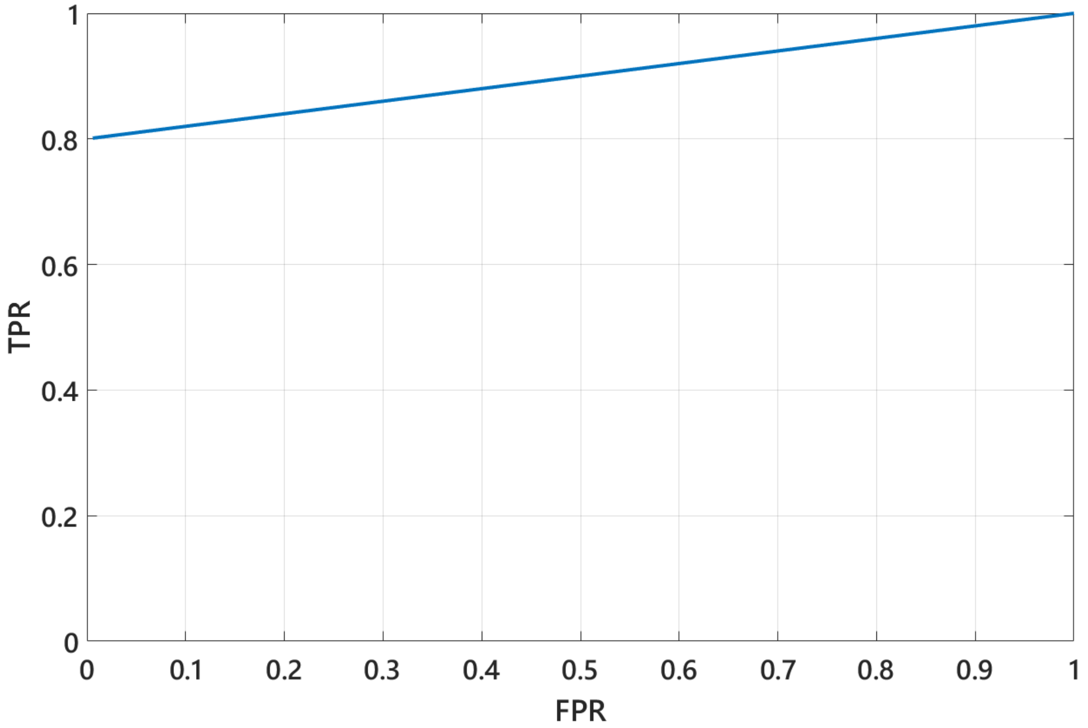


Figure 2.6: Performance evaluation using ROC-AUC.

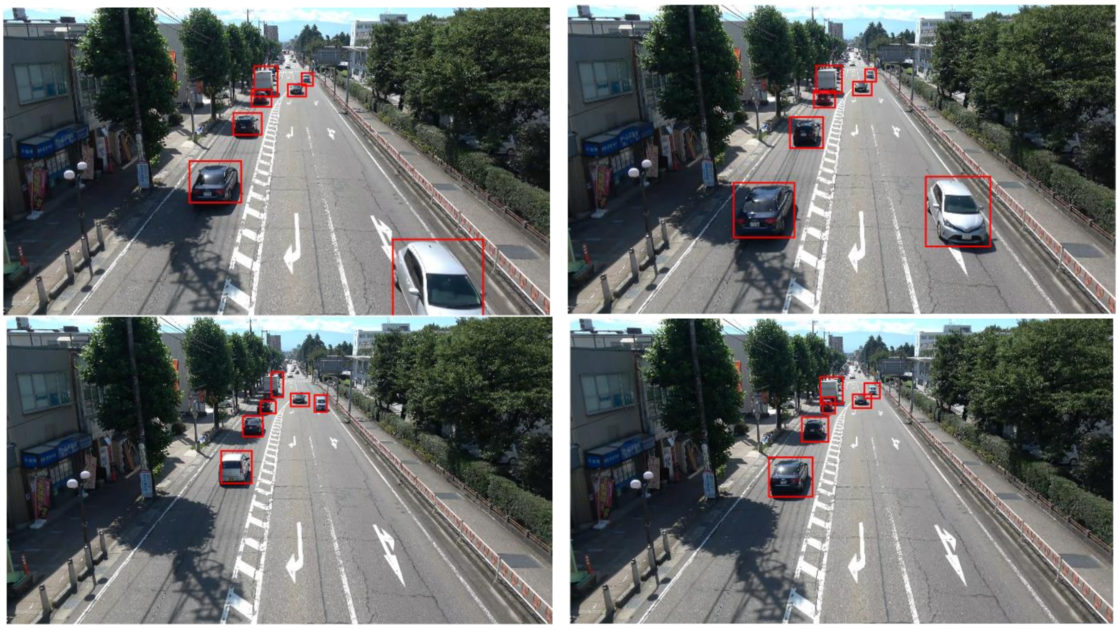


Figure 2.7: Some detection results

2.8 Conclusion

In this study, vehicle detection is carried out using a novel biologically inspired approach in combination with local image features. We use 4K image sequences due to their high resolution and good image quality. Our main goal is to develop vehicle detection method using image processing by taking advantages of the visual attention system.

Chapter 3
PERFORMANCE ANALYSIS OF VEHICLE DETECTION IN COMPRESSED AND UNCOMPRESSED DOMAIN

3.1 Introduction

In this chapter, we presented performance analysis of the vehicle detection method in uncompressed and compressed domain. We use 4K video of a road scene with different types of vehicles as described in chapter 2. The detection performance is analyzed in H.265 compressed domain for assisting to construct a relationship between objective and subjective evaluation value. The performance of this method is demonstrated by the experimental results.

3.2 Database preparation

In order to prepare database, we collected some video at Gofuku near the university from a pedestrian bridge on a road, and the camera was set up with a tripod on the pedestrian bridge on the central line. We used sample outdoor video clip of traffic scene taken by a SONY FDR-AX55, 4K-UHD video camera, which is shown in Figure 3.1.

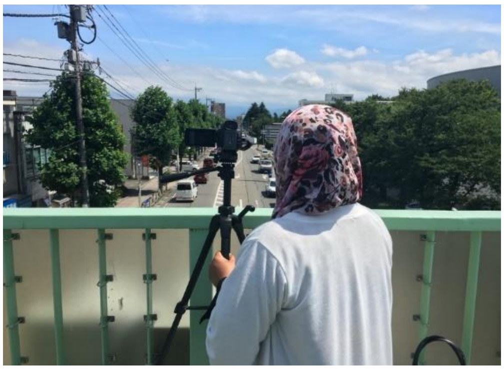


Figure 3.1: Data collection from a pedestrian bridge on a road

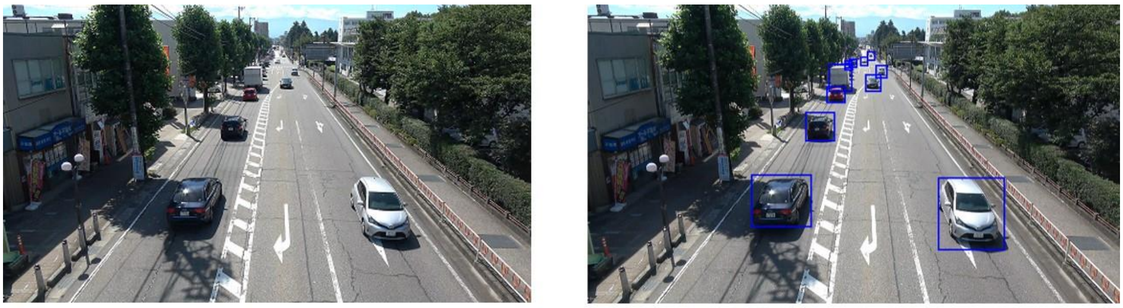
To analyze performance in compressed domain, we created different set of data for different bit rate of the original video. The compression method with different bit rates are shown in Table 3.1.

Table 3.1: Compression method with different bit rate.

Compression method	Bit rate
H.265 (HEVC)	9236 kbs
	739 kbs
	554 kbs
	462 kbs

3.3 Performance analysis

The proposed method is implemented using Matlab on Windows platform in CPU. In order to test the performance of the proposed method, a sequence of frames from 1000 to 1100 is used to show the experimental results. The ground truth bounding boxes are created manually from the uncompressed frames for measuring detection performance. Some ground truth bounding boxes of an input image frame are shown in Figure 3.2.



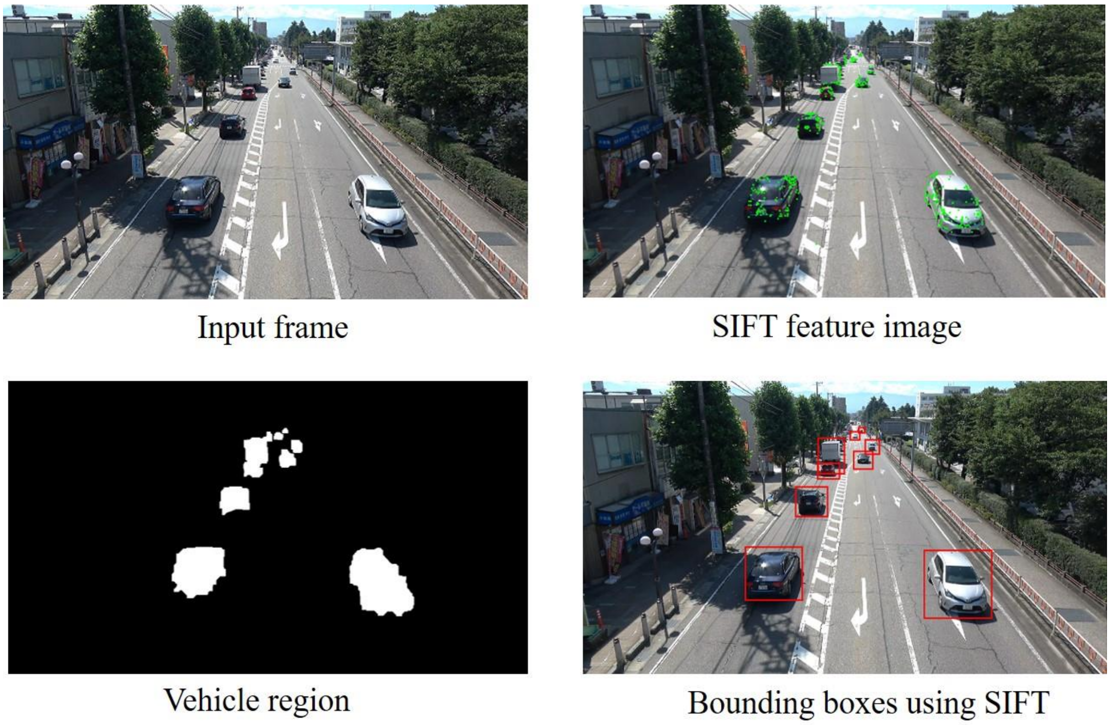
Input frame

Ground truth bounding boxes

Figure 3.2: Ground truth bounding boxes of an image frame.

3.3.1 Comparison of detection results

The Harris feature image, SIFT feature image with vehicle regions of an imput image frame, and their corresponding bounding boxes are shown in Figure 3.3 and Figure 3.4, respectively. The combination of all detected bounding boxes and the filtered bounding boxes are presented in Figure 3.5. The filtered bounding boxes are obtained using non-maximum suppression filtering technique. To test detection performance on each feature, the quantitative evaluation with each feature is shown in Figure 3.6. The y-axis of the figure represents sorted values of the average IoU per frame. The average IoU is calculated for every frame separately by dividing IoU of the detected vehicles and number of vehicle of the frame in the ground truth. Due to the sorted values, the indices of the xaxis does not represent to their corresponding values of the average IoU. It mentions only frame number. However, from this figure, we noticed that SIFT features perform better than other types of features. Our target is to use more features for covering as good shape of the vehicle as possible. The mean of the average IoU (MIoU) and detection percentage combined and single type feature are shown in Table 2. Figure 3.6 and Table 2 indicate that the combined features give best results among the single type feature.



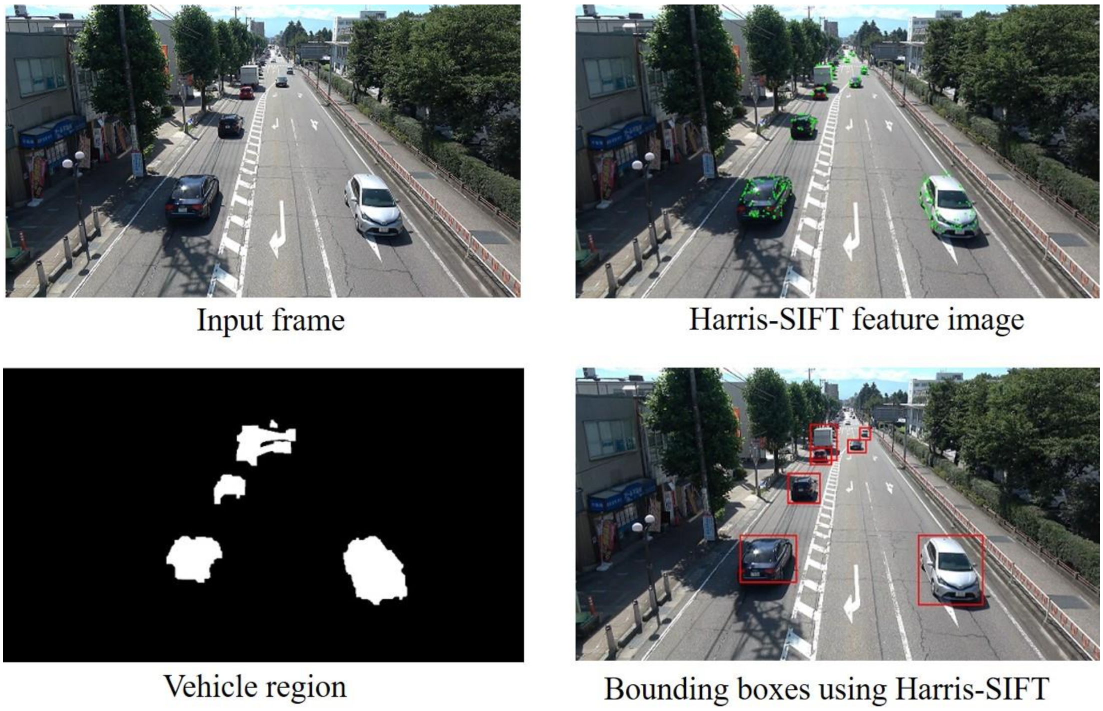
Input frame

SIFT feature image

Vehicle region

Bounding boxes using SIFT

Figure 3.3: SIFT features and corresponding bounding boxes.



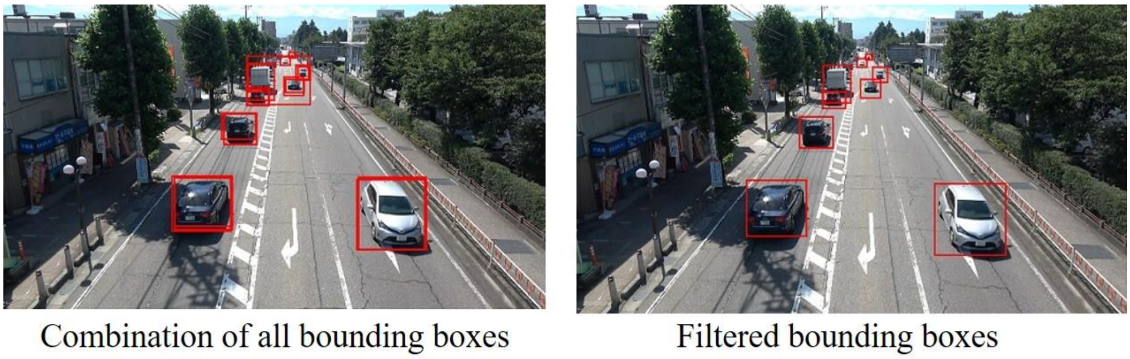
Input frame

Harris-SIFT feature image

Vehicle region

Bounding boxes using Harris-SIFT

Figure 3.4: Harris features and corresponding bounding boxes.



Combination of all bounding boxes

Filtered bounding boxes

Figure 3.5: Combination of all bounding boxes and filtered bounding boxes.

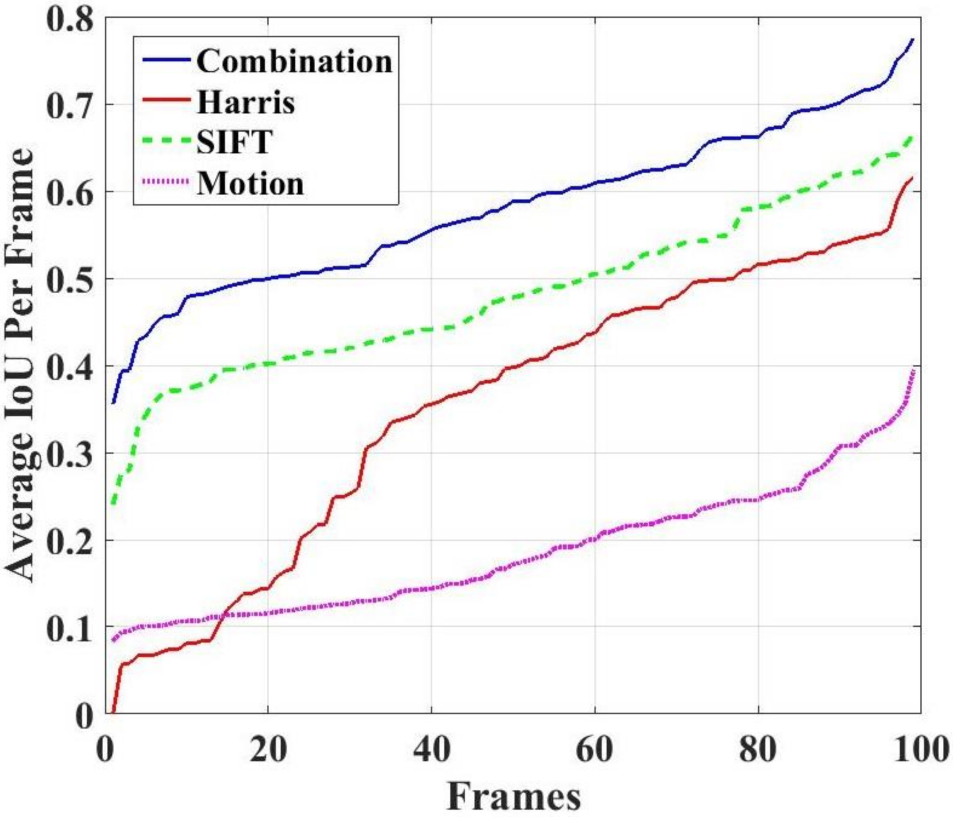
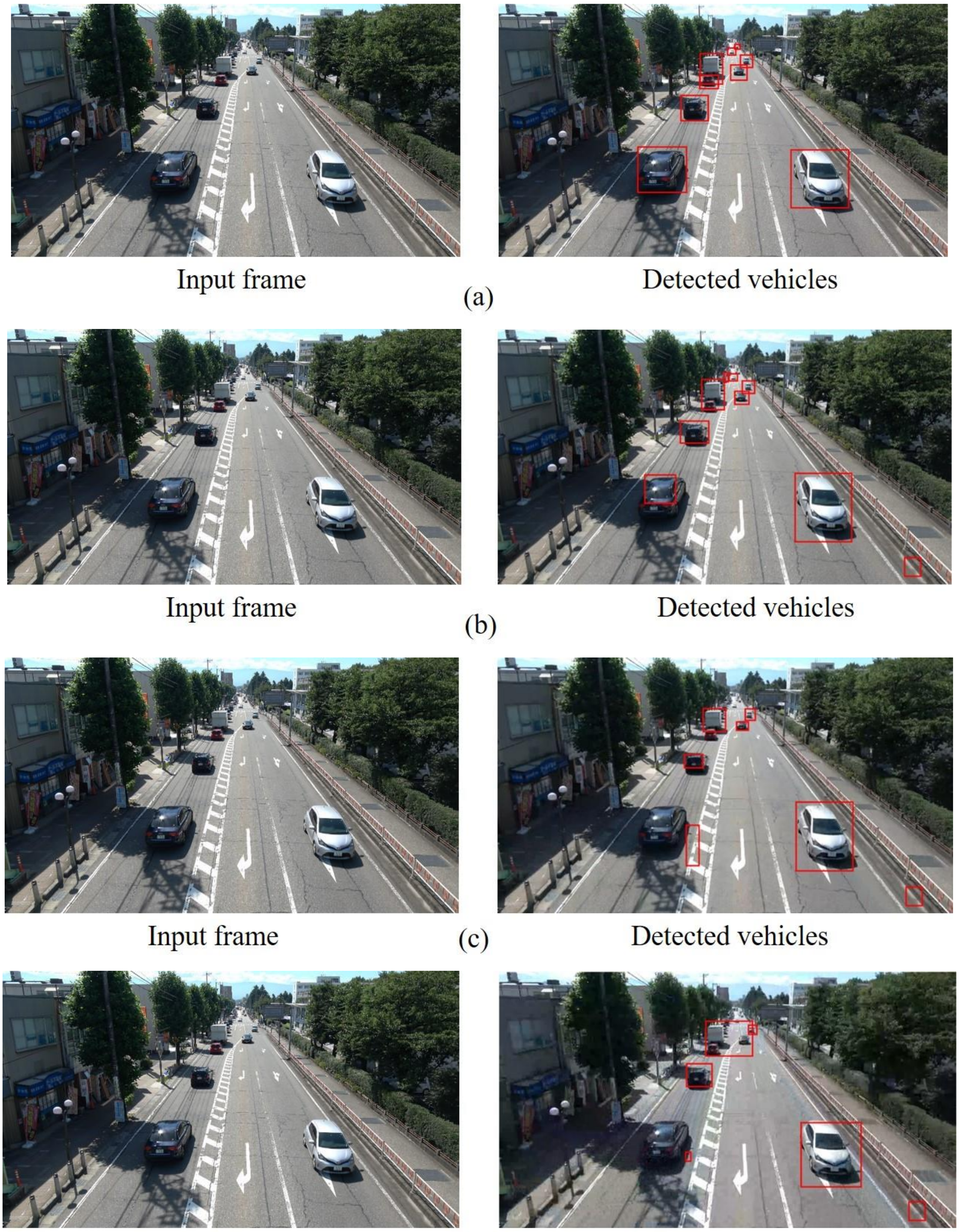


Figure 3.6: Quantitative evaluation with each feature.

Table 3.2: Mean of the average IoU (MIoU) and detection percentage among them.

	Combination	**Harris**	**SIFT**	**Motion**
MIoU	0.58	0.36	0.48	0.19
Percentage [%]	36.2	22.15	30.02	11.67

Motion features are more sensitive on shadow and changing intensity values of the images. Therefore, SIFT and Harris features perform better than motion feature. Some detected results of an input image frame from the original 4K video and compressed video with different bitrates are presented in Figure 3.7. From these results, it is noticeable that the detection performance is decreased due to compression. Figure 3.8 represents some detection results of the proposed model. The results show that the propose method can detect overlapping vehicles from 4K image frames.



Input frame (a) Detected vehicles

Input frame (b) Detected vehicles

Input frame (c) Detected vehicles

Input frame (d) Detected vehicles

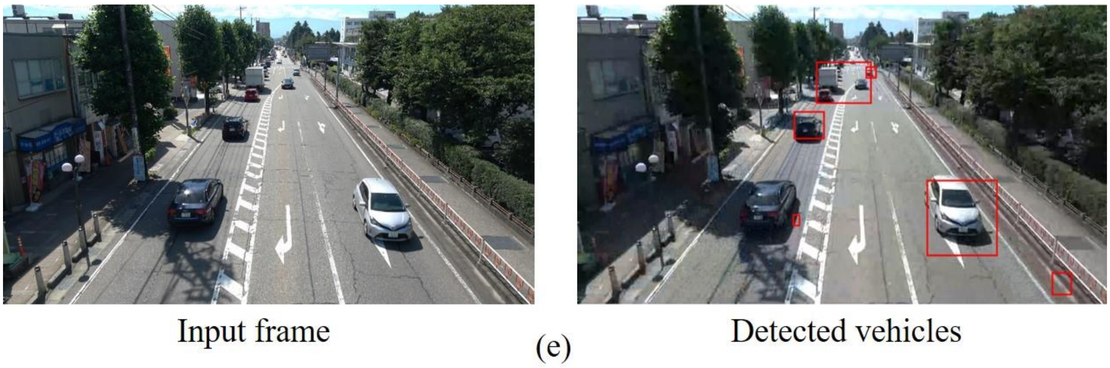


Figure 3.7: Detection results in uncompressed and compressed domain; (a) original video (92362 kbs), (b) compressed (9236 kbs), (c) compressed (739 kbs), (d) compressed (554 kbs), (e) compressed (462 kbs).

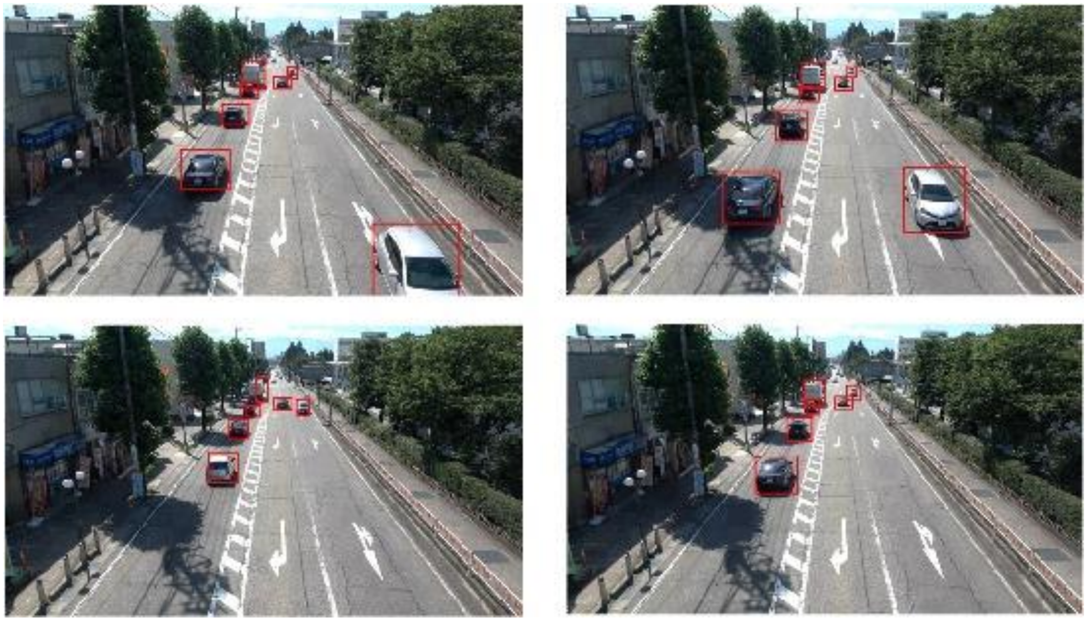
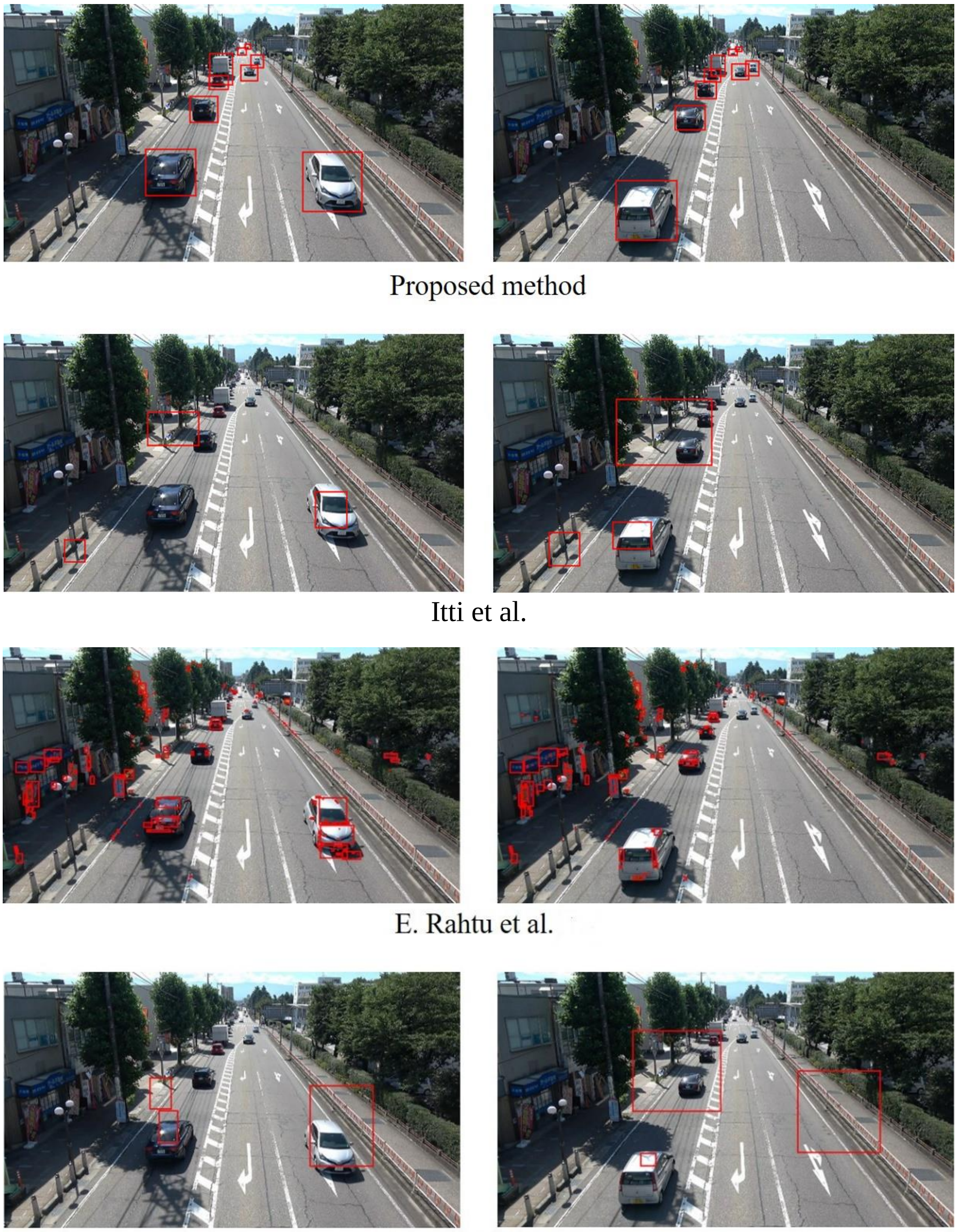


Figure 3.8: Some detection results in uncompressed domain.

The proposed method is compared with Itti et al. [12], E. Rahtu et al. [13], and J. Harel et al. [23]. The proposed detect interest objects like vehicles. However, Itti et al. [12], E. Rahtu et al. [13], and J. Harel et al. [23] methods cannot detect vehicle properly. These conventional methods used intensity, color, orientation, and motion features for extracting saliency information.



Proposed method

Itti et al.

E. Rahtu et al.

J. Harel et al.

Figure 3.9: Detection results of the proposed and existing method

For our experiment, we used different combination of all of the features to test detection performance recommended by the references [12, 13, and 23]. Based on the experiments, we chose all of the four features for evaluation. Some detection results of the conventional and proposed methods are shown in Figure 3.9. The quantitative results of the conventional methods are represented in Figure 3.10. The results of this figure indicate that the proposed method performs better than the conventional methods combining local features and saliency information.

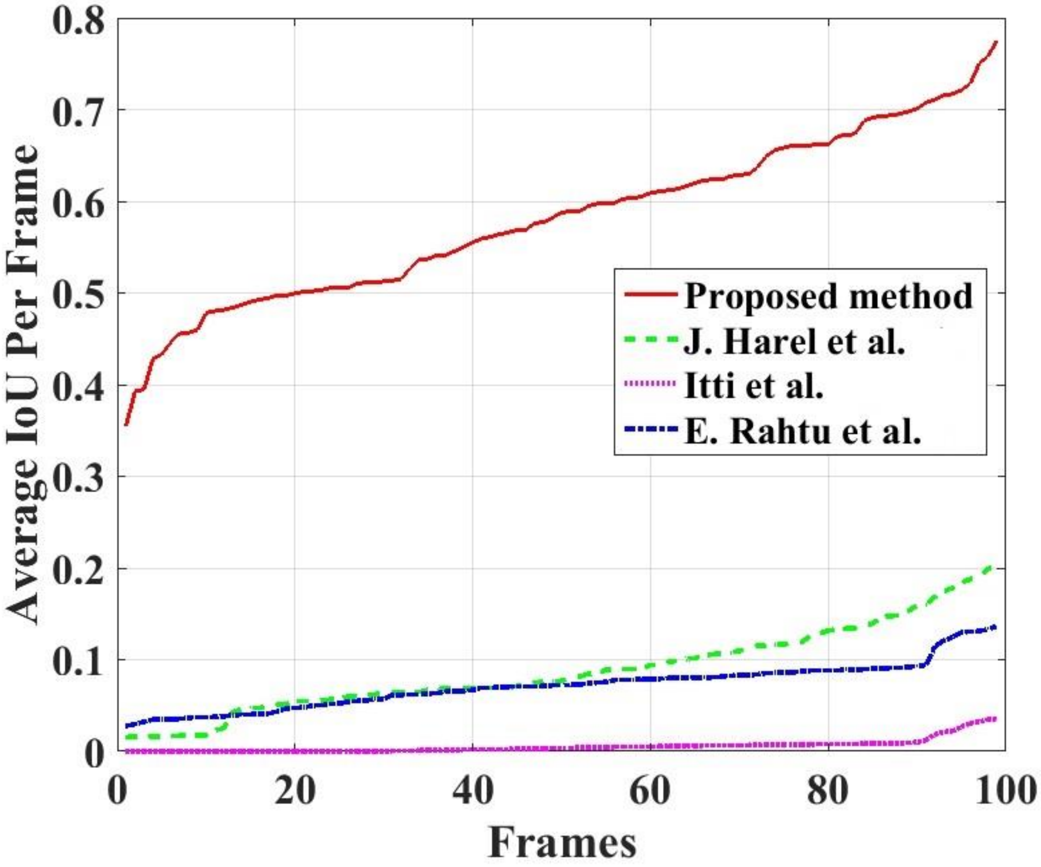


Figure 3.10: Quantitative of the proposed and conventional methods.

3.3.2 Quantitative performance analysis in compressed domain

In order to quantitatively compare the performance of the method in compressed domain, Precision, Sensitivity, and IoU (Intersection over Union) metrics are used to measure the accuracy of bounding box overlap between detected bounding boxes and ground truth bounding boxes. The Precision is a ratio of true positive instances to all positive instances of objects in the detector, based on the ground truth. The Sensitivity is a ratio of true positive instances to the sum of true positives and false negatives in the detector, based on the ground truth. The IoU is a ratio of the area of overlap to the area of union based on the ground truth box. The detection accuracy is also calculated using F Score metric. They are calculated by the following equations:

$$Precision = \frac{TP}{TP+FP} \tag{3.1}$$

$$Sensitivity = \frac{TP}{TP+FN} \tag{3.2}$$

$$F = \frac{2\times Precision\times Sensitivity}{Precision+Sensitivity} \tag{3.3}$$

$$IoU = \frac{D \cap G}{D \cup G} \tag{3.4}$$

where *TP* is the number of true positives, *FP* is the number of false positives, *FN* is the number of false negatives, *D* is the detected bounding box, and *G* is the ground truth bounding box.

The *F* Score value of the proposed method is 0.75. Figure 3.11 represents the values of the *IoU* (Intersection over Union) for each detected bounding boxes.

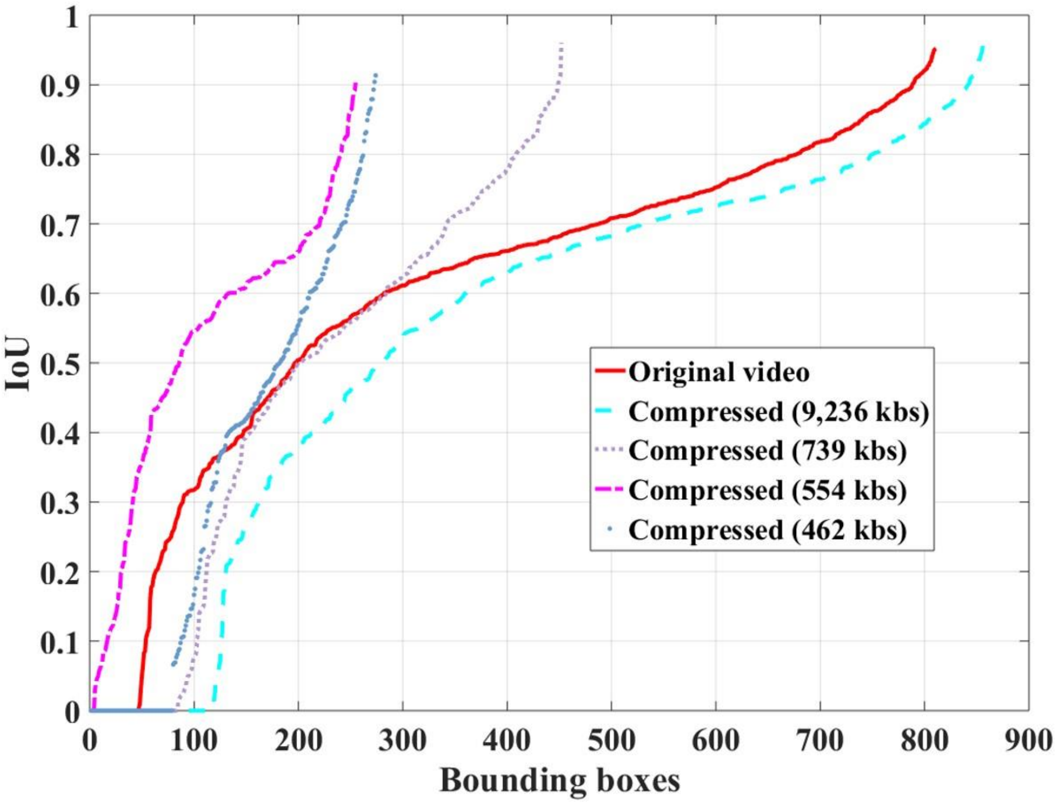


Figure 3.11: *IoU* for each detected bounding boxes.

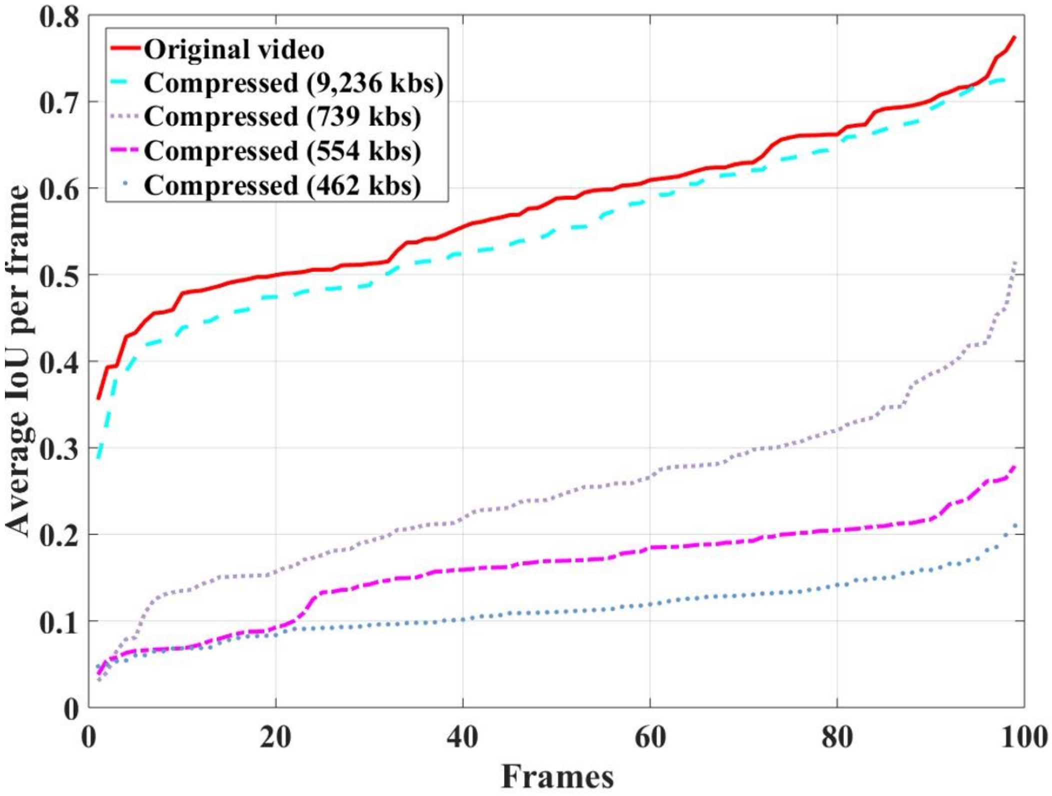


Figure 3.12: Average values of the *IoU* per frame.

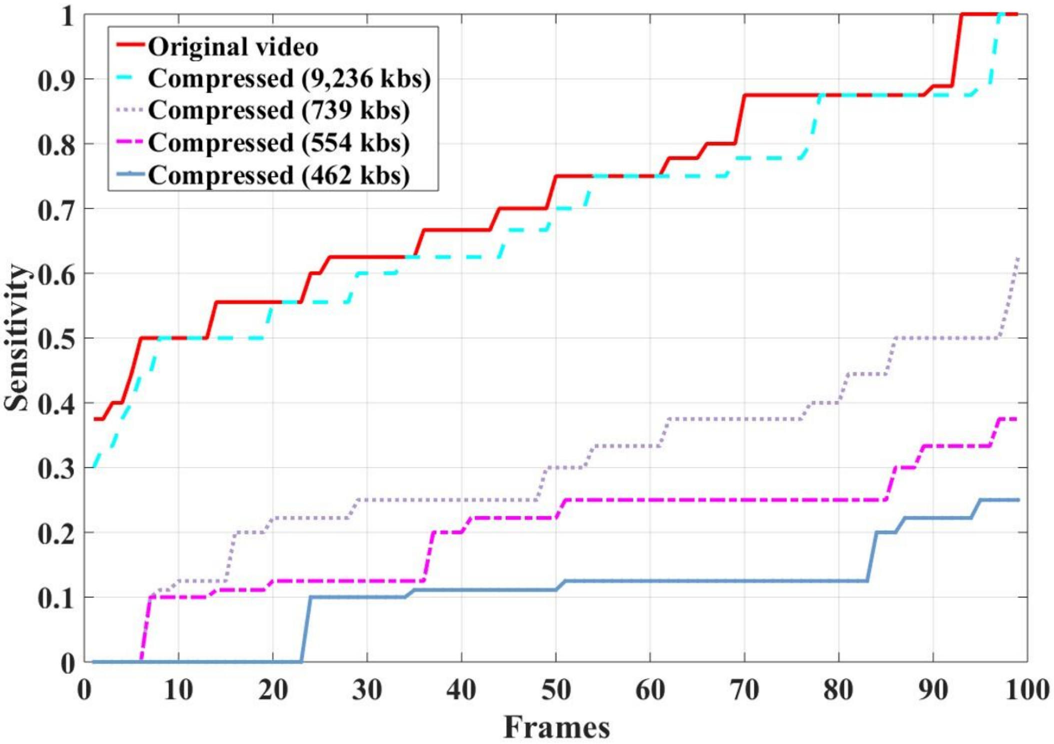


Figure 3.13: Sensitivity per image frame.

The zero values represent false positive, which wrongly indicates that a vehicle is present. The average values of the IoU per image frame are shown in Figure 3.12. The sensitivity per image frame is presented in Figure 3.13. From Figure 3.11, it is noticeable that the number of detected bounding box is decreasing according to size of the bitrate. The results of the Figure 3.12 and Figure 3.13 indicate that the detection accuracy is decreased in compressed domain.

To see robustness of the video compression with the conventional methods, the detection performance using conventional method and the proposed method in compressed domain is represented in Figure 3.14.

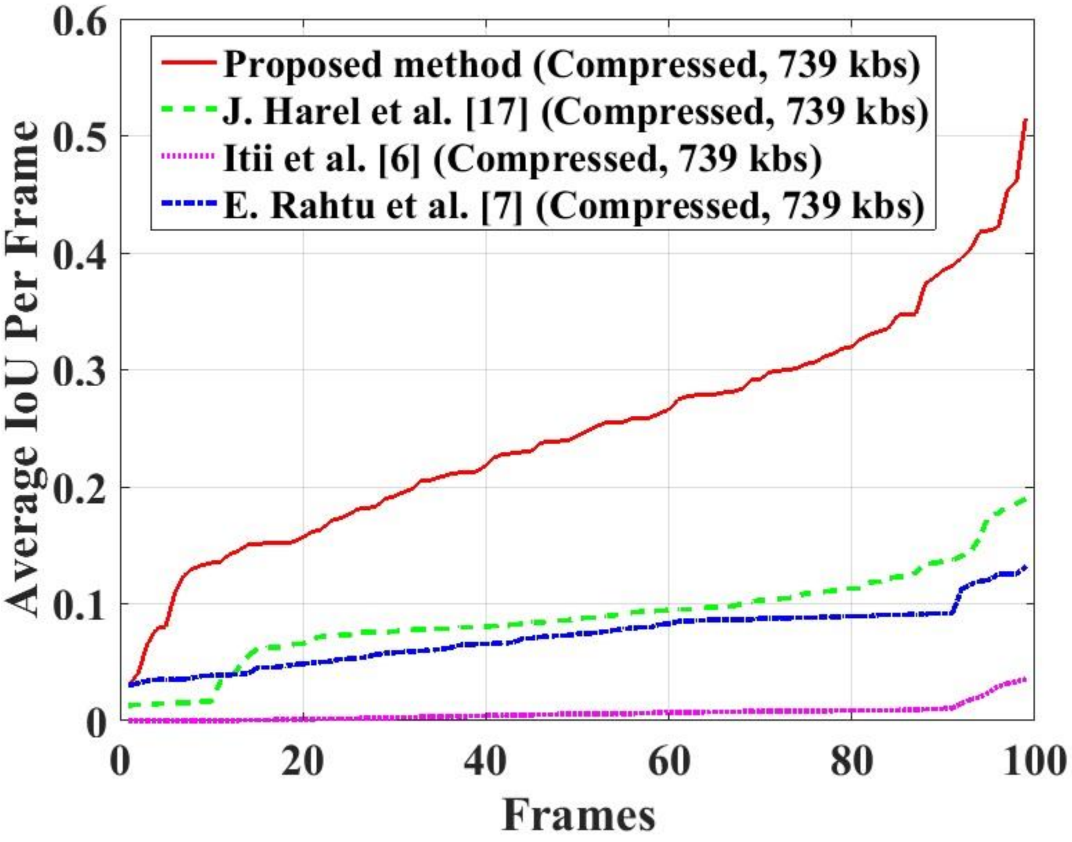


Figure 3.14: Detection performance using conventional and proposed method in compressed domain.

The Figure 3.14 indicates that the proposed method achieves higher robustness of the video compression compared with that of conventional methods. We also calculated processing time of each frame for combined feature and individual feature. The average processing time per frame of the proposed method with combined features is 15.0 minutes.

On the other hand, the average processing times for SIFT, Harris, and motion are 7.5, 3.9, and 0.04 minutes, respectively.

3.4 Conclusion

In this study, vehicle detection and performance analysis in compressed domain are carried out using a novel biologically inspired approach in combination with local image features. Although the processing time increased due to the feature combination, this analysis result of the detection performance in compressed domain will lead us to develop video quality model for detection by transmitting reasonable high-quality video. It will also facilitate to develop subjective evaluation model for detection.

Chapter 4

CONSTRUCTION OF SUBJECTIVE VEHICLE DETECTION EVALUATION MODEL CONSIDERING SHIFT FROM GROUND TRUTH POSITION

4.1 Introduction

In general, object recognition research, the region considered a vehicle may be labeled using a rectangle, or the target region and non-target region may be labeled at the pixel level. The detected region is evaluated using Intersection over Union (IoU), which is widely used method in detection. However, the area predicted by the vehicle detection model is not always the same as the subjective evaluation because the position is shifted in various directions with respect to the ground truth, which is defined by human. In this chapter, we focused on the subjective evaluation method by investigating the effect of misalignment or positional deviation from human generated ground truth.

4.2 Subjective evaluation method in vehicle detection

In this section, we discussed subjective evaluation methods, called Mean Opinion Score (MOS) and Differential Mean Opinion Score (DMOS), for vehicle detection. These two methods have been described shortly in the following sub-sections.

4.2.1 Mean opinion score (MOS)

Mean Opinion Score (MOS) is a subjective measurement used in the domain of Quality of Experience and telecommunications engineering, representing overall quality of a stimulus or system. In our research work, we used MOS to measure quality of vehicle detection of the proposed method. It is the arithmetic mean over all individual values on a predefined scale that a subject assign to his opinion on the detected vehicle.

4.2.2 Differential mean opinion score (DMOS)

Differential Mean Opinion Score (DMOS) is also a subjective measurement, which is calculated by the score of ground truth and the score of evaluation image. It is defined by the following expression:

$$DMOS = Ground\ truth\ image\ rating - evaluation\ image\ rating \qquad (4.1)$$

4.3 Subjective evaluation experiments

In a subjective evaluation experiment to evaluate the detection accuracy of a vehicle, some arbitrary still images are extracted from a 4K (3840 x 2160 pixels) video. The video was taken from a pedestrian bridge on a road, and the camera was set up with a tripod on the pedestrian bridge on the central line. In this study, 20 arbitrary still images are selected from the captured video. The conditions for the selected still image are as follows:

- Vehicles are clearly recognized
- There are no other vehicles near the vehicle, whose detection accuracy is to be evaluated.
- Near and far distant scenes are extracted for the same vehicle under the above conditions.
- Close view: About 13m from the overpass.
- Far view: About 45m from the overpass.
- Vehicle type: Passenger car
- Vehicle color: 5 types (black, white, gray, red, blue)
- Only one vehicle per image

Figure 4.1 shows the evaluation image for the left vehicle, and Figure 4.2 shows the evaluation image for the right vehicle. In the target vehicle of each evaluation image, the region indicated by the red rectangle is the vehicle region defined in this experiment. The target vehicle colors for the left and right vehicles are 5 types (black, white, grey, red, and blue). Near and far scenes were selected for each target vehicle in the left and right lanes. There are 10 scenes selected for the vehicle on the left and 10 scenes selected for the vehicle on the right.

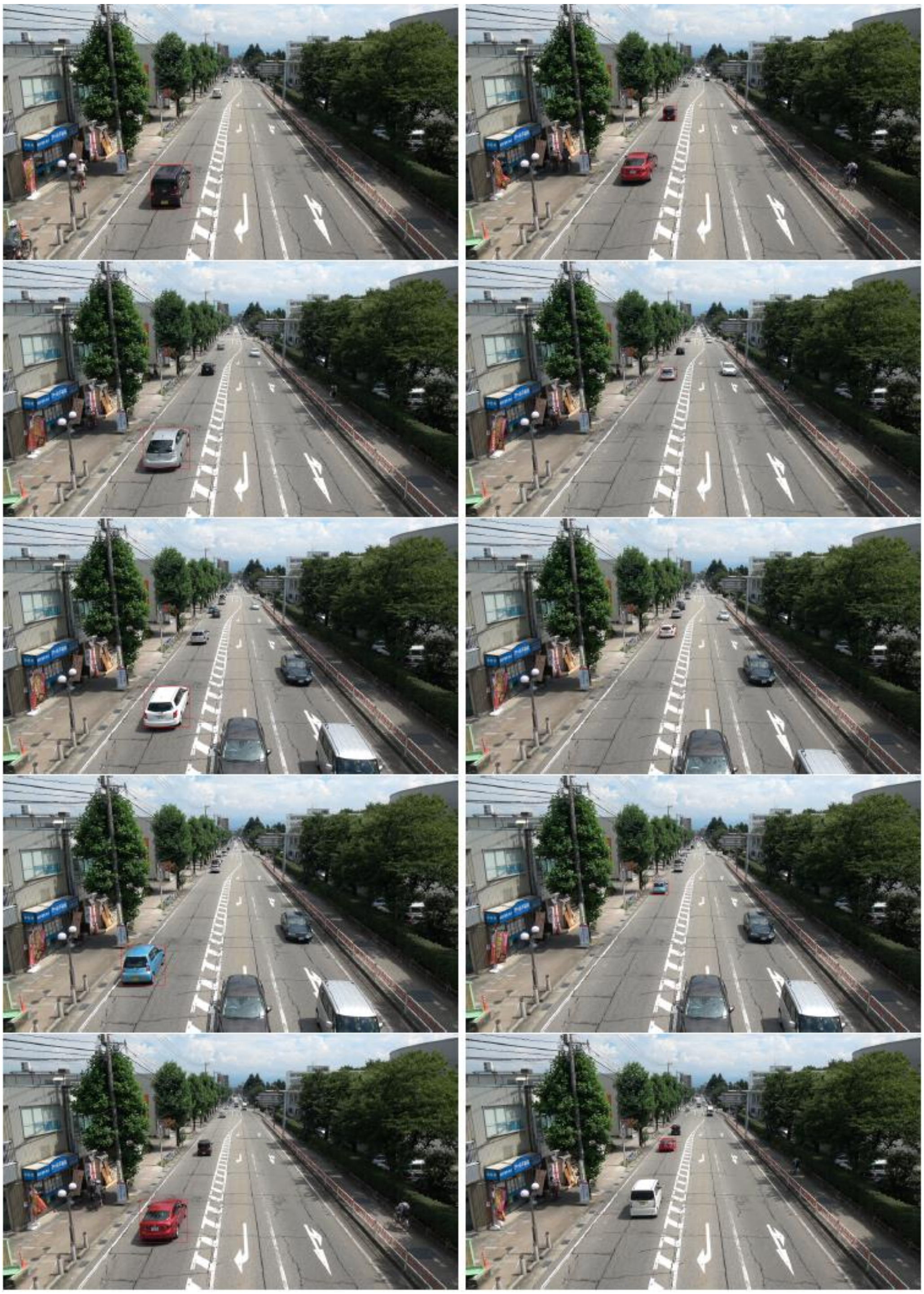


Figure4.1: Evaluation images of the vehicle on the left

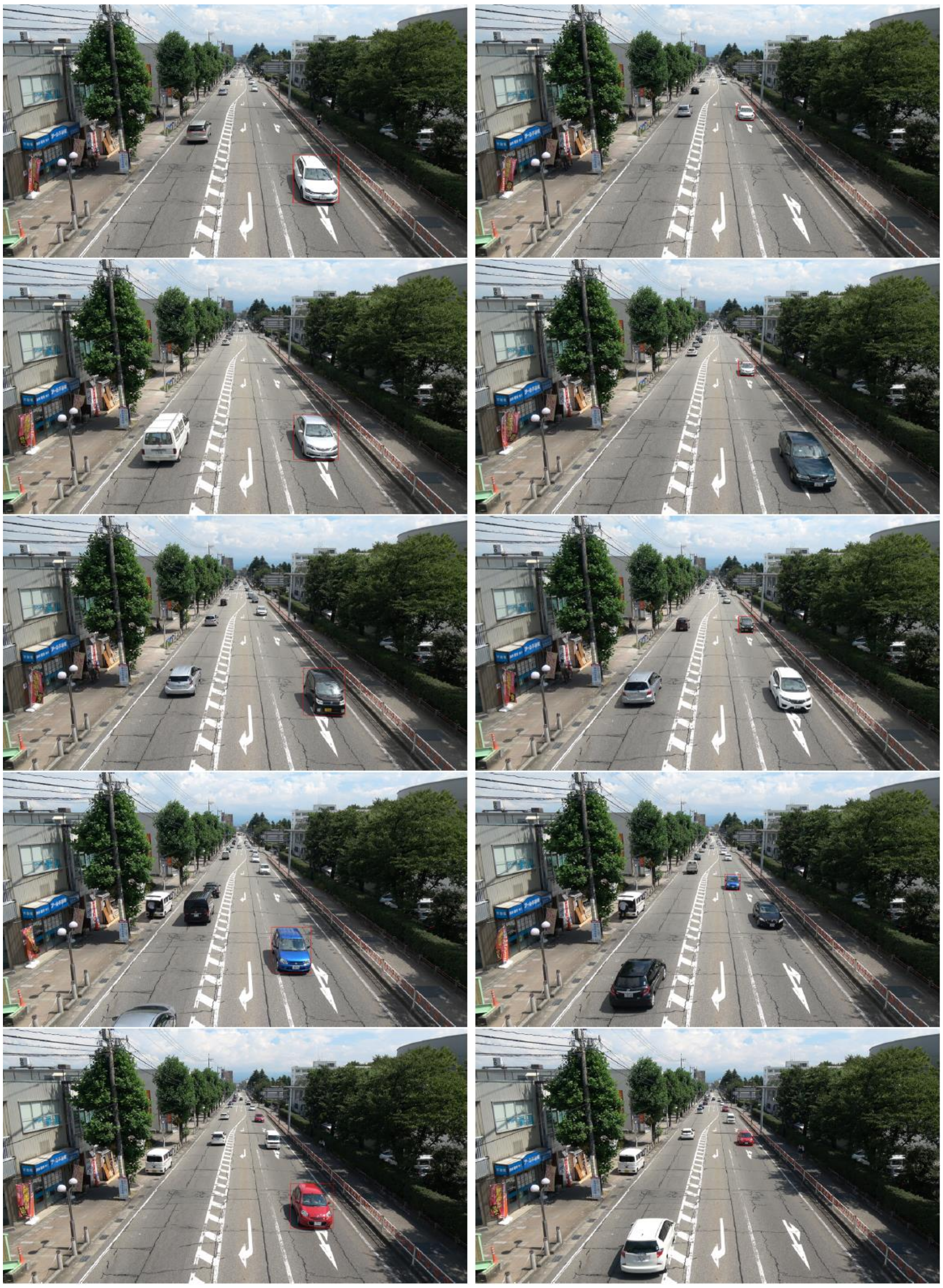


Figure 4.2: Evaluation images of the vehicle on the right

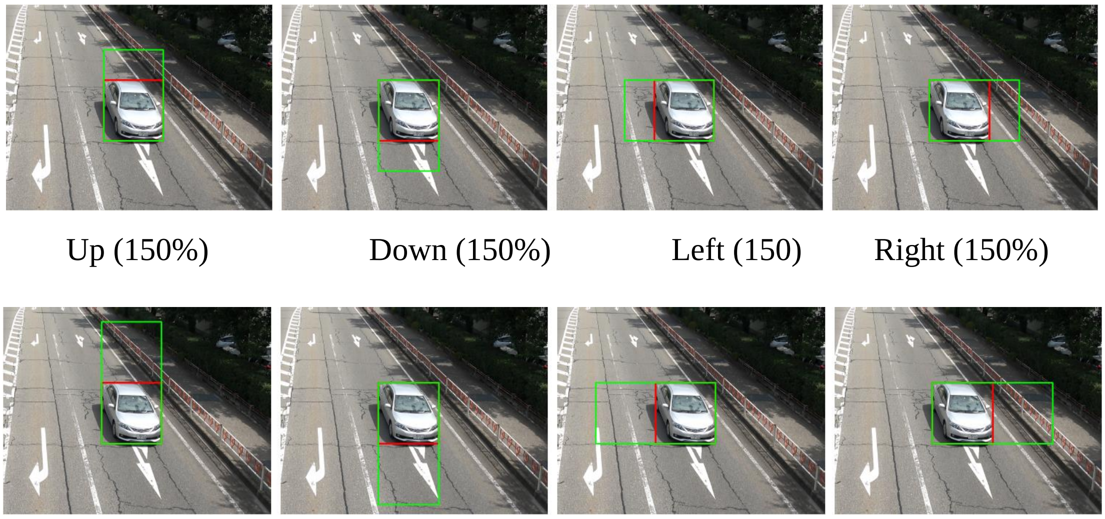
4.3.1 Types of evaluation images

Focusing on the positional deviation from the defined vehicle area, some evaluation images were created for the displacement in order to investigate how the positional deviation in vehicle detection affects the subjective evaluation. The following three types of evaluation images were created with reference to the positional deviations:

- Misalignment in only one direction with respect to the vertical and horizontal directions
- Misalignment in two directions with respect to the vertical and horizontal directions
- Misalignment by enlarging or reducing with respect to the original area of the vehicle

4.3.1.1 Misalignment in only one direction with respect to the vertical and horizontal directions

A unidirectional misalignment is a misalignment in which the position is deviated only in one direction with respect to the up / down / left / right direction from the area of the vehicle. The displacement rate was 150% and 200%. A 150% misalignment in a unidirectional misalignment is a unidirectional misalignment with an area of 150% when the area of the vehicle area without misalignment is 100%. Similarly, a displacement of 200% is a displacement of 200% when the area without displacement is 100%. For the unidirectional misalignment, the misalignment in the unidirectional region including the vehicle area without misalignment was referred to the misalignment of the object detection method in the previous research on object recognition. The following figure shows an example of misalignment in only one direction with respect to the top, bottom, left, and right directions (Figure 4.3).

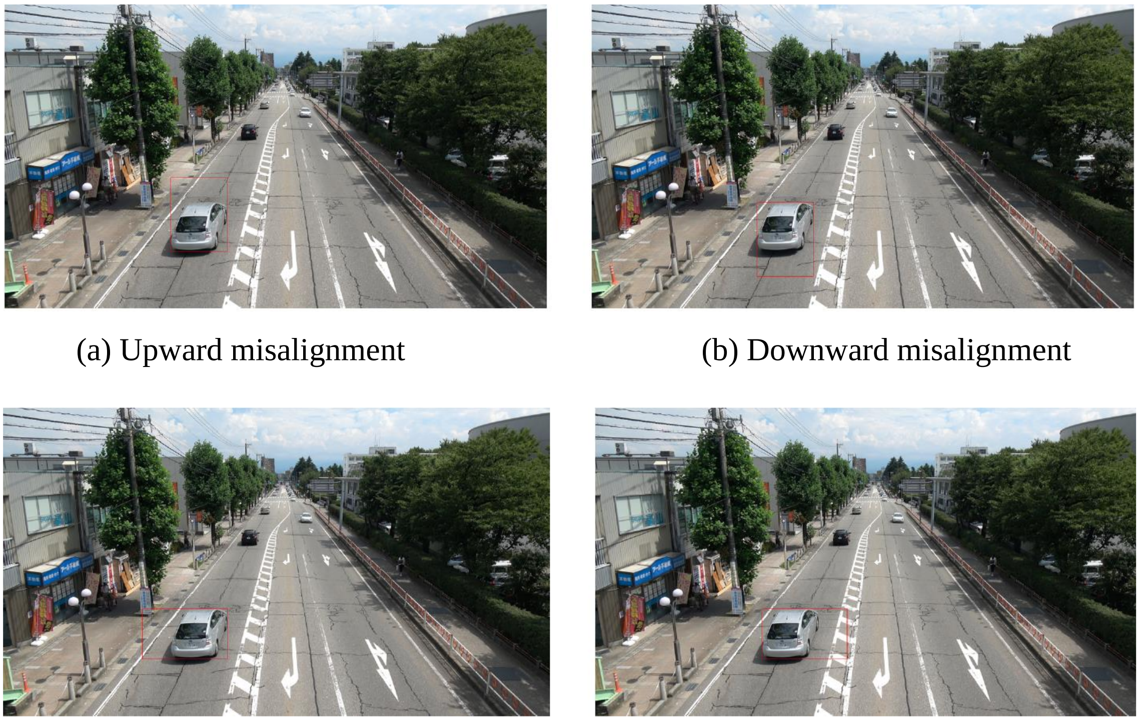


Up (150%) Down (150%) Left (150) Right (150%)

Up (200%) Down (200%) Left (200) Right (200%)

Figure 4.3: Example of misalignment in one direction

Figure 4.4 shows an example of an evaluation image of misalignment in only one direction on the left vehicle. Figure 4.4 shows an example of an evaluation image in the foreground. The color of the vehicle is gray. Figures 4.4 (a) to 4.4 (d) are all misalignment rates of 150%.



(a) Upward misalignment

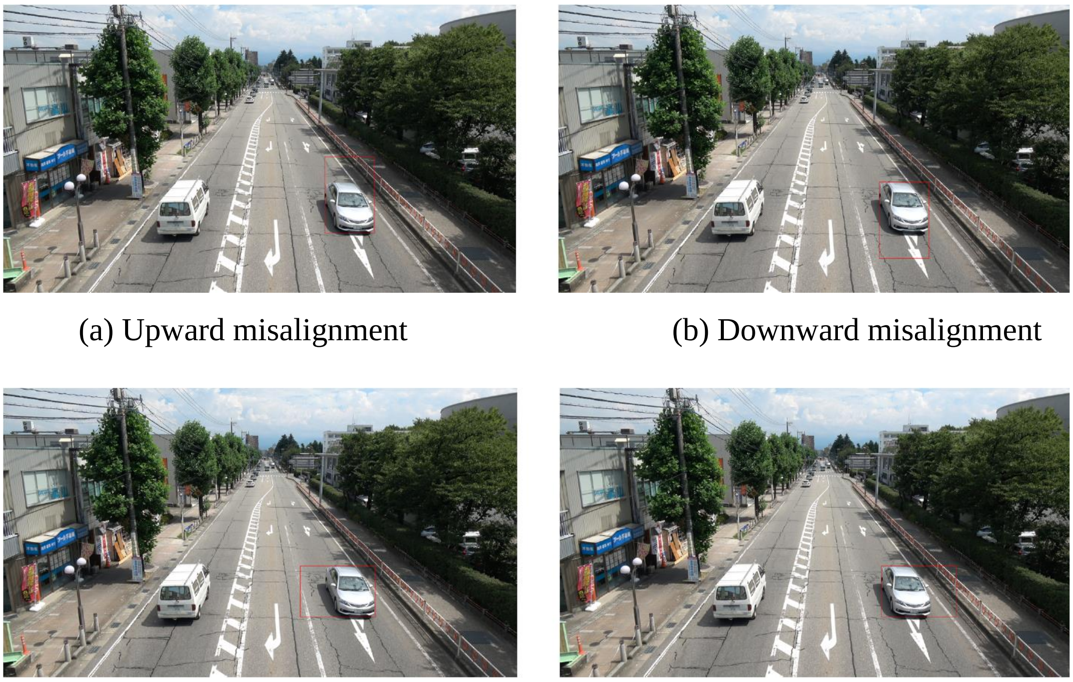
(b) Downward misalignment

(a) Left misalignment

(b) Right misalignment

Figure 4.4: Example of misalignment in only one direction on the left vehicle (ratio of misalignment 150%)

Figure 4.5 shows an example of an evaluation image for the displacement in one direction on the right-hand vehicle. Figure 4.5 shows the case of a close-up view, and the color of the vehicle is gray. Figures 4.5 (a) to 4.5 (d) all show a displacement of 150%. The subject subjectively evaluates the vehicle detection accuracy for the vehicle surrounded by the red rectangle.



(a) Upward misalignment

(b) Downward misalignment

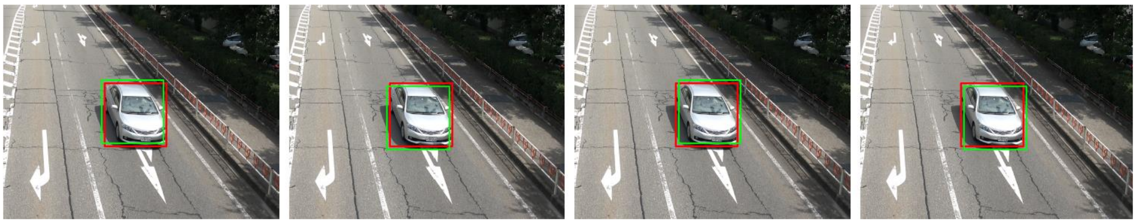
(a) Left misalignment

(b) Right misalignment

Figure 4.5: Example of misalignment in only one direction on the right vehicle (ratio of misalignment 150%)

4.3.1.2 Misalignment in two directions with respect to the vertical and horizontal directions

The misalignment in two directions is a misalignment that is misaligned in two directions with respect to the vertical and horizontal directions from the area of the vehicle without misalignment. The displacement rate was 5% to 20%. The 5% misalignment in the two directions is 5% misalignment in the left / right direction from the vehicle area, and in the up / down direction. There are four types of misalignment in the two directions, the upper left, the lower left, the upper right, and the lower right, for the region without misalignment. The following figure shows an example of misalignment in two directions with respect to the vertical and horizontal directions.

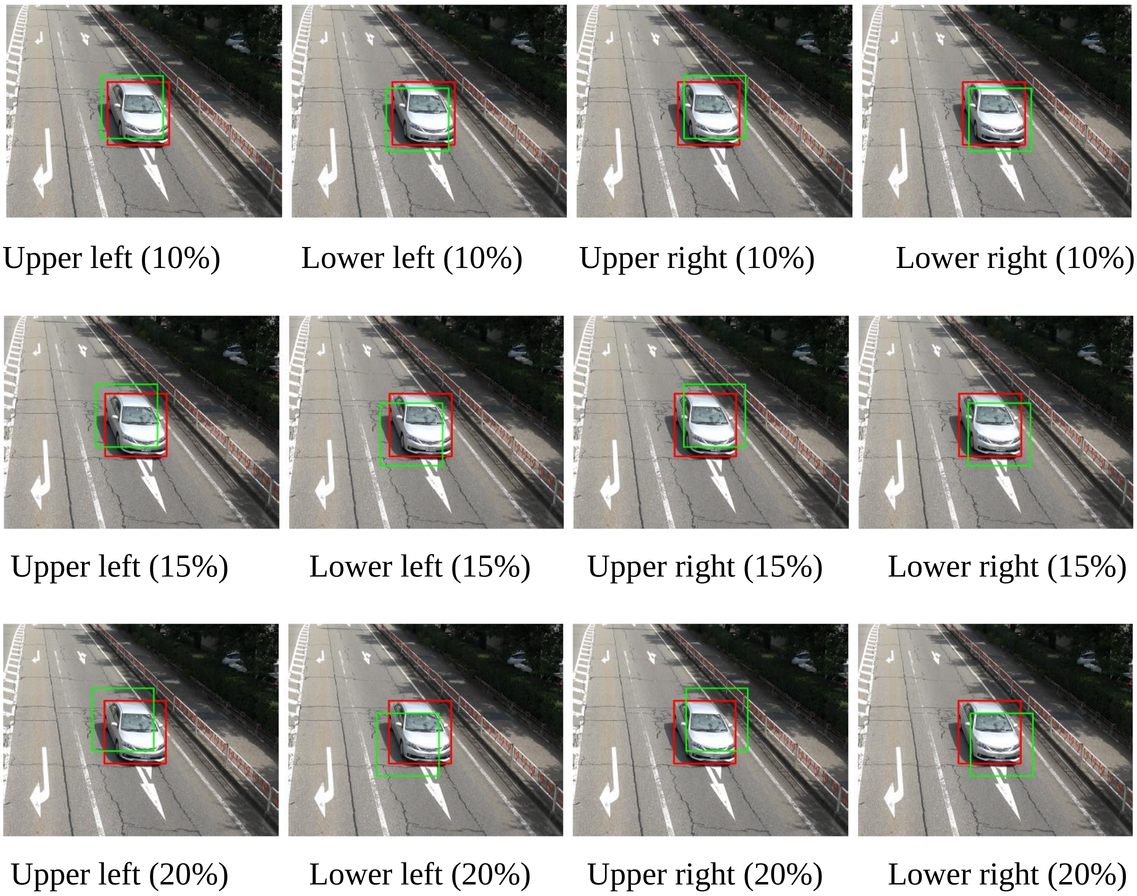


Upper left (5%)

Lower left (5%)

Upper right (5%)

Lower right (5%)



Upper left (10%) Lower left (10%) Upper right (10%) Lower right (10%)

Upper left (15%) Lower left (15%) Upper right (15%) Lower right (15%)

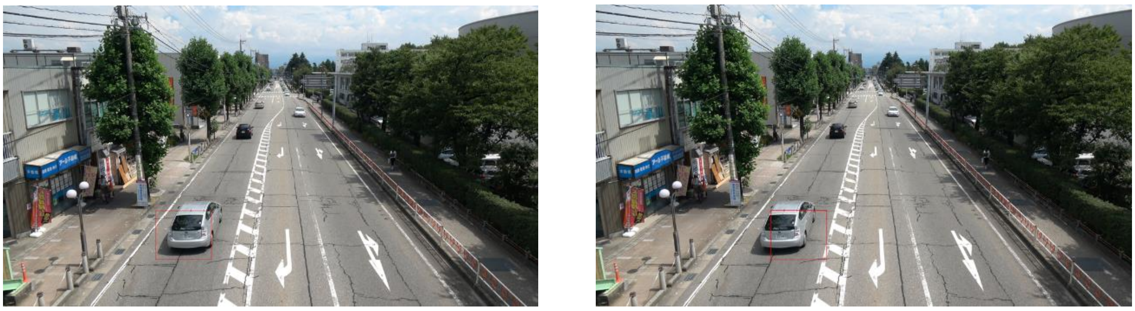
Upper left (20%) Lower left (20%) Upper right (20%) Lower right (20%)

Figure 4.6: Example of misalignment with shifting in two directions.

Figure 4.7 shows an example of an evaluation image of the misalignment in two directions on the left vehicle. Figure 4.7 shows the evaluation image in the foreground, and the color of the vehicle is gray. The misalignment rate in Fig. 4.7 is an example of an evaluation image with 20% misalignment in all directions.

(a) Upper left

(b) Upper right

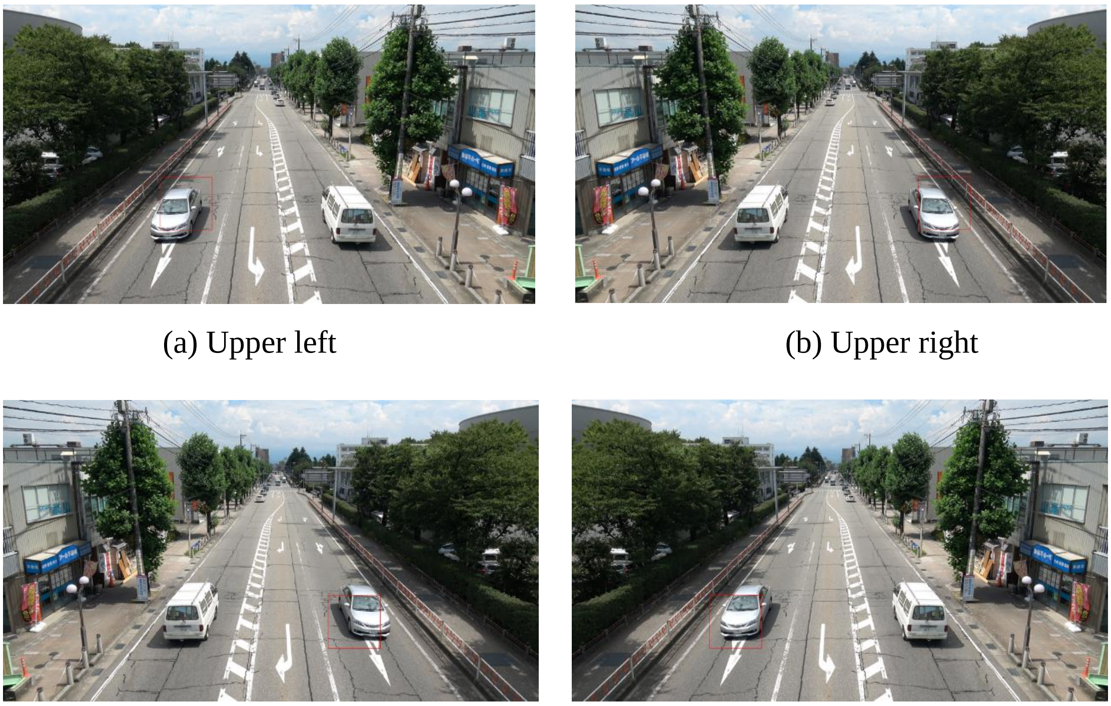


(c) Lower left

(d) Lower right

Figure 4.7: Example of evaluation image of misalignment in two directions on the left vehicle (displacement rate 20%)

Figure 4.8 shows an example of the evaluation image of the misalignment in the two directions on the right vehicle. Figure 3.8 shows an evaluation image in the foreground and the vehicle color is gray. The misalignment rate in Fig. 4.8 is an example of an evaluation image with 20% misalignment in all directions. The subject subjectively evaluates the detection accuracy of the vehicle enclosed in the red rectangle.



(a) Upper left

(b) Upper right

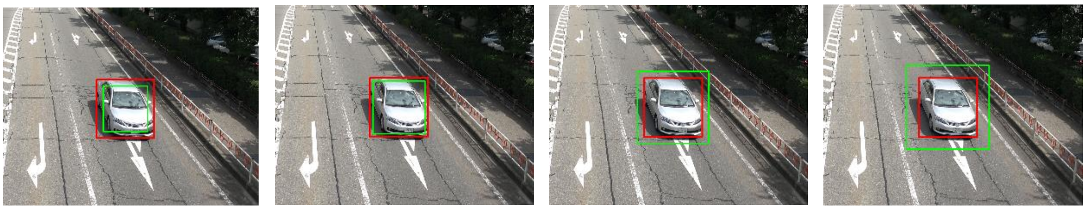
(c) Lower left

(d) Lower right

Figure 4.8: Example of evaluation image of misalignment in two directions on the right vehicle (displacement rate 20%)

4.3.1.3 Misalignment for enlargement / reduction

Positional displacement that causes enlargement / reduction refers to an area where the area of the vehicle that is regarded as having no displacement is 100%, and the positional deviation that is reduced is 60% and 80% of the area to be evaluated. The positional deviation was evaluated for the area of 150% and 200%. The four types of misalignment between 60% and 200% were used for evaluation. Figure 4.9 shows an example of misalignment.

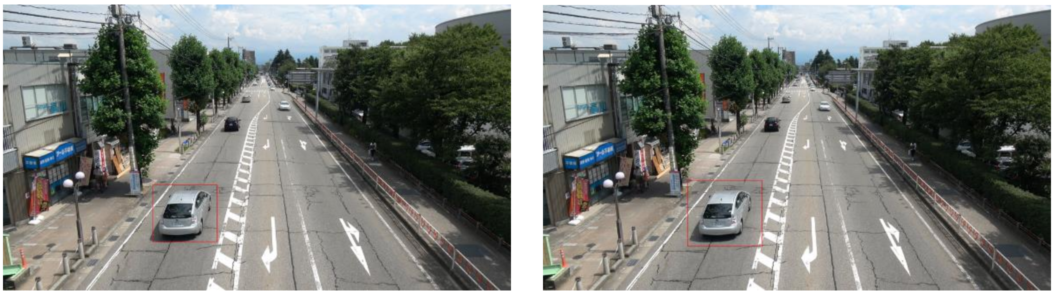


60% 80% 150% 200%

Figure 4.9: Example of misalignment for magnification

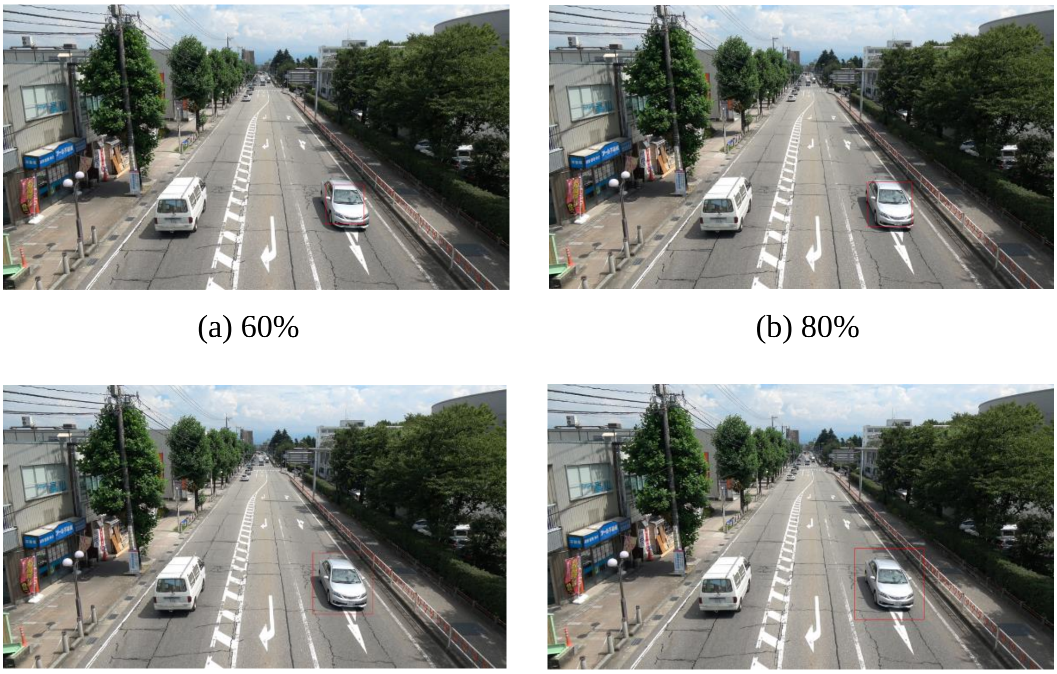
Figure 4.10 shows an example of an evaluation image of misalignment in enlargement / reduction for the left vehicle. Figure 4.10 shows the evaluation image in the foreground, and the color of the vehicle is gray. Figure 4.11 shows an example of an evaluation image of misalignment in enlargement / reduction in the vehicle on the right. Figure 4.11 shows the case of a close-up view, and the color of the vehicle is gray. The subject subjectively evaluates the detection accuracy of the vehicle surrounded by a red rectangle.

(a) 60% (b) 80%



(c) 150% (d) 200%

Figure 4.10: Misalignment when zooming in and out on the left vehicle.



(a) 60% (b) 80%

(c) 150% (d) 200%

Figure 4.11: Misalignment when zooming in and out on the right vehicle.

4.3.2 Summary of misalignment of evaluation images

Table 4.1 shows the type of misalignment, direction of misalignment, and total number. There are two types of misalignment rates (150% to 200%) in one direction of misalignment, and there are a total of eight patterns of misalignment directions: four types (left, right, top, bottom). There are 4 types (5% to 20%) of displacement rates in the two-direction misalignment, and there are 16 types of displacement directions (upper left, lower left, upper right, lower right). For enlargement / reduction position displacement, there are 4 types of displacement rate (60% to 200%), and the displacement direction is the displacement in all directions in the vertical and horizontal directions, for a total of 4 patterns. For each evaluation vehicle, 29 types of evaluation are performed including the area of the vehicle that is regarded as having no displacement.

Table 4.1: Summery of misalignment or deviation

Type of Misalignment	Variety of Misalignment rate	Direction	Total Number
One direction (Figure 4.3)	2 (150%, 200%)	Left, Right, Upper, Lower	8
Two direction (Figure 4.6)	4 (5%-20%)	Left upper, Left lower, Right upper, Right lower	16
Magnification (Figure 4.9)	4 (60%-200%)	All direction	4

4.4 Experimental conditions

As experimental conditions for conducting subjective evaluation experiments to evaluate vehicle detection accuracy, the subjects were 70 men and women. The subject evaluates the vehicle detection performance in a questionnaire format. Table 4.2 summarizes conditions of the subjective evaluation experiment.

Table 4.2: Subjective evaluation experiment conditions

Subject	70 men and women
Subjective evaluation method	7- level evaluation
Evaluation time	Unlimited
Evaluation images	580

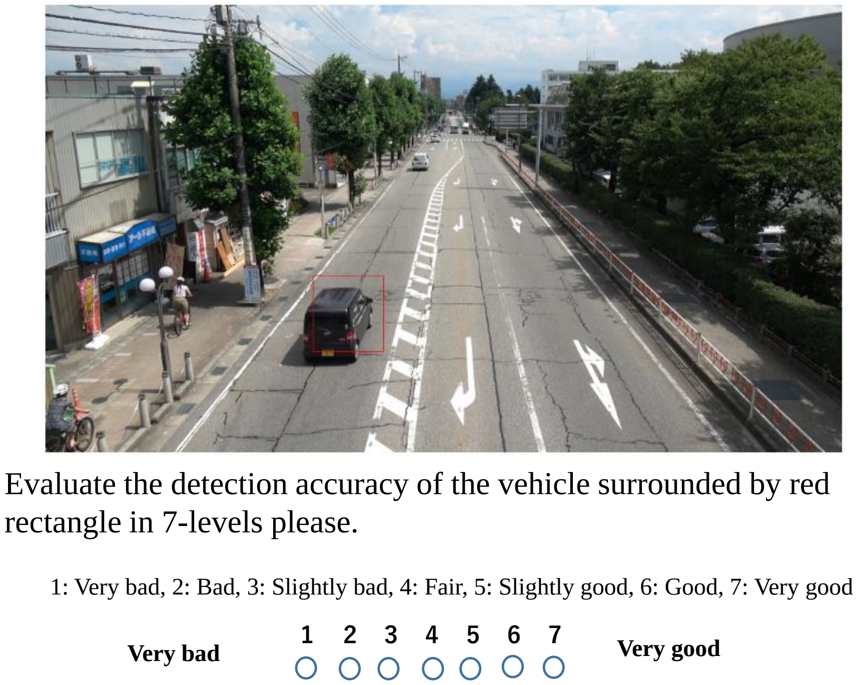
Table 4.3: Seven grade scale

Score	Rating
1	Very bad
2	Bad

3	Slightly bad
4	Fair
5	Slightly good
6	good
7	Very good

Table 4.3 shows the seven-level rating scale. There are scores from 1 to 7, and rating words from "very bad" to "very good" are supported. The evaluation time is unlimited, and it takes an average of 1 hour to complete all experiments. There are 580 evaluation images, 290 for the vehicle on the left and 290 for the vehicle on the right. Since the vehicle on the left side and the vehicle on the right side have different driving directions, the left and right side vehicles are separated from each other. First, the left vehicle evaluation images are randomly divided into six groups. Subjects conduct questionnaires for all groups. Similarly, in the evaluation image of the vehicle on the right, 290 evaluation images are randomly divided into six groups. In the subjective evaluation experiment, the right vehicle is evaluated 4 weeks after all the left vehicles have been evaluated. The subject evaluates the detection accuracy of the vehicle with a PC.

Figure 4.12 shows an example of a questionnaire evaluation screen on a PC. On the screen of the PC, an evaluation image and radio buttons corresponding to seven levels of evaluation words are displayed. The subject evaluates the detection accuracy of the vehicle enclosed in a red rectangle in seven stages with respect to the displayed evaluation image. After the evaluation is completed for each evaluation image, the next evaluation image is displayed.



Evaluate the detection accuracy of the vehicle surrounded by red rectangle in 7-levels please.

1: Very bad, 2: Bad, 3: Slightly bad, 4: Fair, 5: Slightly good, 6: Good, 7: Very good

Very bad 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ **Very good**

Figure 4.12: Sample questionnaire evaluation screen

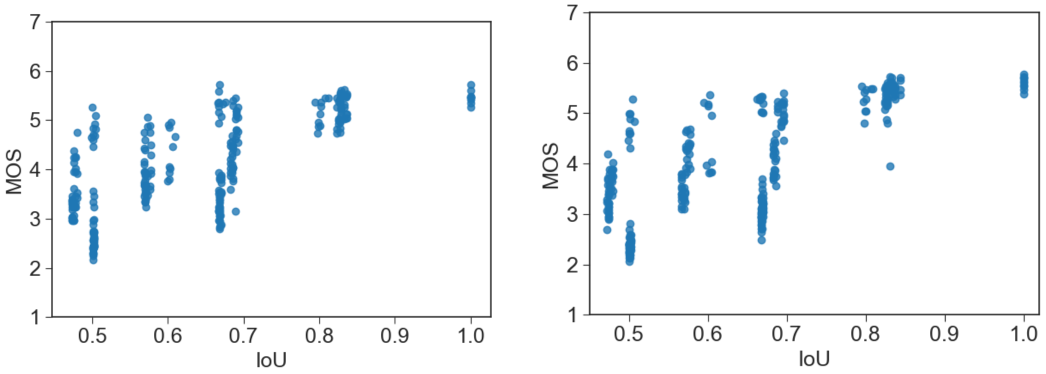
4.5 Experimental results

In this section, we describe the results of subjective evaluation experiments and considerations on vehicle detection accuracy in the left and right vehicles.

4.5.1 Relationship between IoU and subjective evaluation values

The subjective evaluation value is calculated from the questionnaire results obtained from the subjective evaluation experiment, and this is used as the subjective evaluation value. The relationship between the subjective evaluation value and IoU is shown below.

Figure 4.13 is a scatter plot of subjective evaluation values and IoU obtained from the experimental results. The vertical axis in Fig. 4.13 is the subjective evaluation value, and the horizontal axis is the evaluation value calculated using IoU. In order to approximate the relationship between the subjective evaluation value and IoU from the scatter diagram in Figure 4.13, an approximate expression of the cubic function was calculated from the linear function in this study.

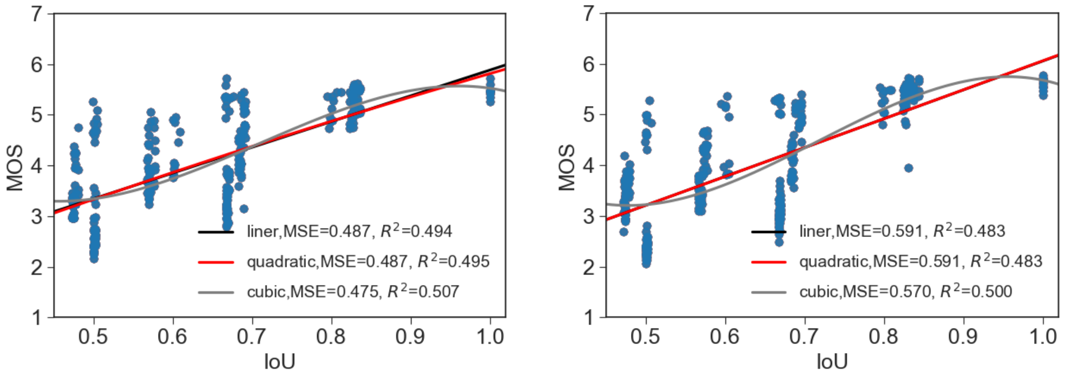


(a) Vehicle on the left

(b) Vehicle on the right

Figure 4.13: Scatter plot of subjective assessment values and IoU.

Figure 4.14 shows the linear and nonlinear regression equations calculated from the relationship between subjective evaluation values and IoU for the left and right vehicles, superimposed on the scatter plot in Figure 4.13. Figure 4.14 (a) is an approximation in the scatter diagram of the vehicle on the left, and the black straight line in the figure is an approximate line using a linear function. The red curve in Figure 4.14 (a) is an approximate curve using a quadratic function, and the gray curve is an approximate curve using a cubic function.



(a) Vehicle on the left

(b) Vehicle on the right

Figure 4.14: Approximate formulas calculated from subjective assessment values and IoU.

The mean square error (MSE) and the coefficient of determination (R2) were calculated from each approximate expression. In the same way, the approximate equation was calculated for the vehicle on the right, and the mean square error and the coefficient of determination were calculated (Fig. 4.14 (b)). Table 4.4 and Table 4.5 show the approximation formulas, coefficient of determination, and mean square error for each order for the left and right vehicles.

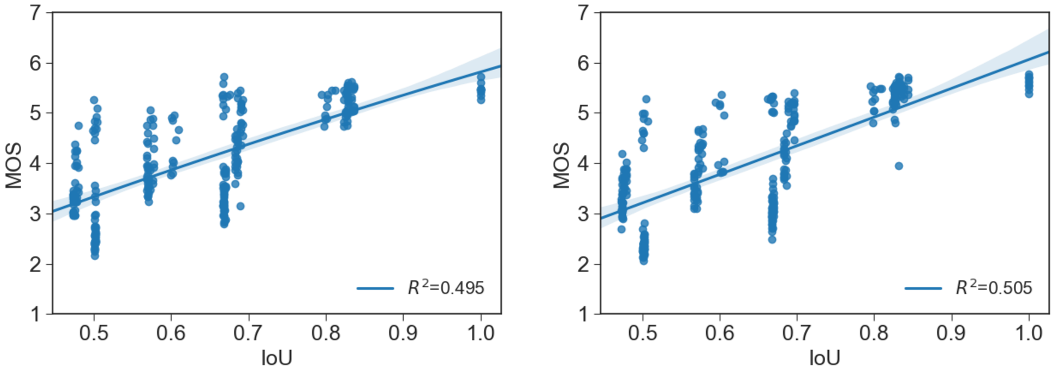
Table 4.4: Approximate Formulas Calculated from Relationship between Subjective Evaluation Value and IoU (Left Vehicle)

Degree	**Formula (y: MOS, x: IoU)**	R^2	**MSE**
1st order	$y = 5.082x + 0.8006$	0.494	0.487
2nd order	$y = 0.8838x^2 + 6.291x + 0.4053$	0.495	0.487
3rd order	$y = -35.43x^3 + 75.36x^2 - 46.65x + 12.26$	0.507	0.475

Table 4.5: Approximate Formulas Calculated from Relationship between Subjective Evaluation Value and IoU (right Vehicle)

Degree	**Formula (y: MOS, x: IoU)**	R^2	**MSE**
1st order	$y = 5.693x + 0.361$	0.505	0.5913
2nd order	$y = 0.05991x^2 + 5.611x + 0.3878$	0.505	0.5913
3rd order	$y = -46.48x^3 + 100x^2 - 63.79x + 15.92$	0.522	0.5703

As shown in Figure 4.14, there is almost no improvement in approximation due to an increase in dimensions, so an approximate curve using a quadratic function was used in this paper. Figure 4.15 shows a scatter plot of subjective evaluation values and IoU for the left and right vehicles, and approximate curves based on the respective quadratic functions.



(a) Vehicle on the left

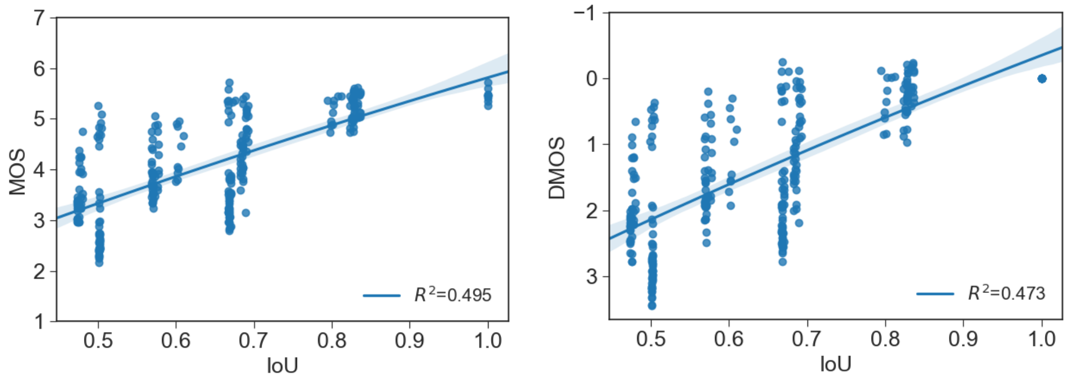
(b) Vehicle on the right

Figure 4.15: Relationship between subjective evaluation values and IoU.

Figure 4.15 (a) shows the experimental results for the left vehicle. The maximum subjective evaluation value was 5.728. Figure 4.15 (b) shows the experimental results for the vehicle on the right. The maximum subjective evaluation value was 5.772. For the left and right vehicles, the evaluation value corresponding to very good subjective evaluation values has not been obtained. Therefore, in this experiment, we calculated the difference (Diff erential MOS: DMOS) between the score of the vehicle area (Ground truth) and the score of the evaluation image, which are regarded as having no displacement.

DMOS is obtained from Eq. (4.1), and the difference between the scores of the evaluation images including misalignment is calculated based on the scores of the image of the vehicle area without misalignment defined in this experiment. Considering the score in the evaluation image without misalignment, it can be assumed that the evaluation for the evaluation image without misalignment is “7: Very good”. Since the evaluation for the evaluation image without misalignment is “7: very good”, it can be considered an ideal result in this experiment, so DMOS was calculated as the experimental result.

Figure 4.16 (a) shows the relationship between the average subjective evaluation value and the IoU evaluation value for the left vehicle.

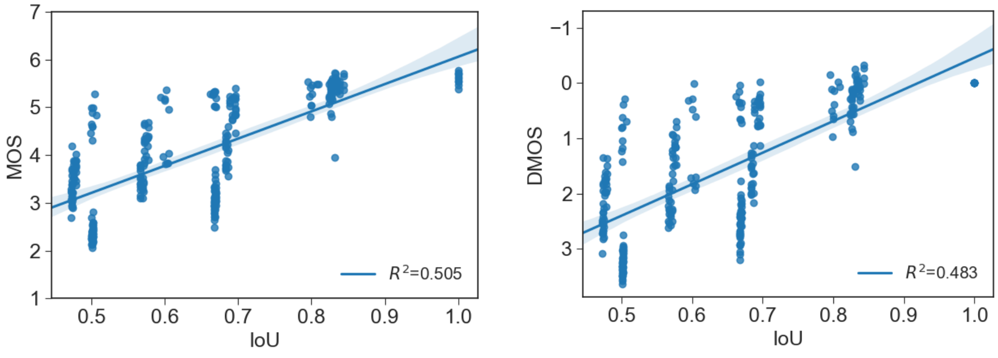


(a) (b)

Figure 4.16: Relationship IoU with MOS and DMOS (vehicle on left side)

The curve in the figure is an approximate curve with a quadratic function. In the figure, the vertical axis is the average of the evaluation values subjectively, and the horizontal axis is the evaluation value of IoU. From the experimental results, the maximum MOS value for the left vehicle was 5.728, and the minimum value was 2.157. The minimum value of IoU in the evaluation image for the left vehicle is 0.472, where 1 is the area of the vehicle that is regarded as having no displacement. In Fig. 4.16 (b), the horizontal axis is the IoU evaluation value, and the vertical axis is DMOS. The minimum value of DMOS was -0.242 and the maximum value was 3.44.

Figure 4.17 shows the relationship between the subjective evaluation value and IoU for the right vehicle.

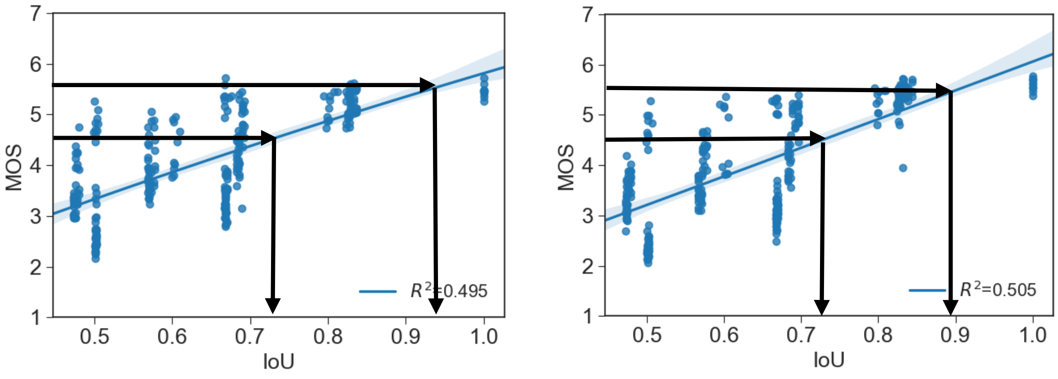


(a) (b)

Figure 4.17: Relationship IoU with MOS and DMOS (vehicle on right side)

The curve in the figure is an approximate curve with a quadratic function. The maximum MOS value for the vehicle on the right was 5.772, and the minimum value was 2.060. In DMOS, the maximum value was 3.636 and the minimum value was -0.318.

From the experimental results, there was no difference between the left and right vehicles in the relationship between IoU and subjective evaluation values.



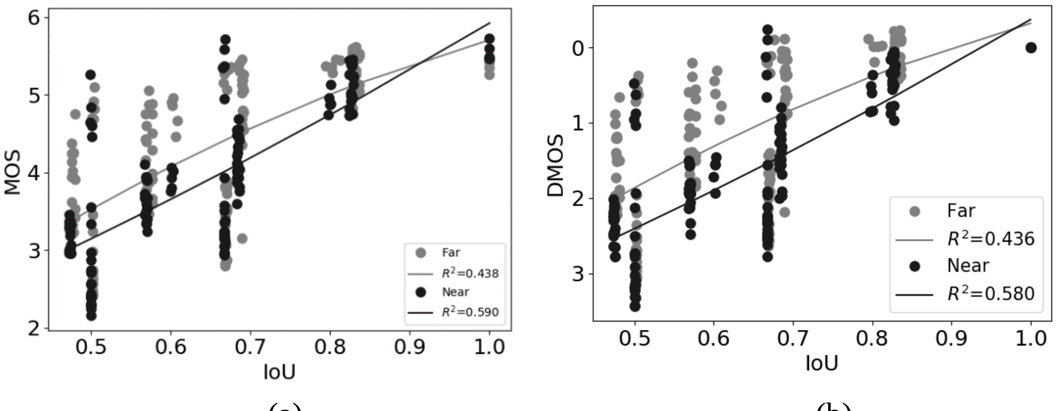
(a) Left vehicle (b) Right vehicle

Figure 4.18: Focusing on subjective score in the relationship between subjective evaluation values and IoU.

Figure 4.18 focuses on the subjective evaluation of the relationship between the subjective evaluation value and IoU, and the IoU evaluation value corresponding to the evaluation word is indicated by a black arrow. Considering 4.5, which includes “4: Normal” and “5: Slightly good” in the subjective evaluation value as the “Slightly good” threshold, the IoU corresponding to “Slightly good” was around 0.75. In addition, when the subjective evaluation value 5.5, which includes “5: Slightly good” and “6: Good”, is included as a good threshold, the IoU corresponding to “good” was 0.9 or more. Based on the above, 0.5 or 0.7 is used as the evaluation value of IoU in the past, but it is considered that 0.75 or more is necessary from the experimental results.

4.5.2 MOS and DMOS in near view and far view

Figure 4.19 shows the relationship between IoU, MOS, and DMOS when focusing on foreground and background. In the evaluation image, the target vehicle that is close to the camera is taken as the foreground, and the target vehicle that is far from the camera is taken as the far view. The horizontal axis in Figure 4.19 (a) is the IoU evaluation value, and the vertical axis is MOS. In Figure 4.19 (a), the gray point indicates the distant view, and the black point is the result obtained from the foreground evaluation image. The horizontal axis in Figure 4.19 (b) is the evaluation value in IoU, and the vertical axis is DMOS. Similarly, in Figure 4.19 (b), the gray point is the target vehicle in the distant view, and the black point is the result obtained from the evaluation image of the target vehicle in the close view. An approximate curve was drawn using the least-squares method at each point in the foreground and background. The gray curve in Figure 4.19 (a) is for a distant view, and the black curve is for a close view. The decision factors for each are shown in Figure 4.19 (a). In Figure 4.19 (b), similar to Figure 4.19 (a), approximate curves were drawn for the foreground and foreground, and the coefficient of determination was calculated. Table 4.6 shows the approximate equations for the vehicle on the left.



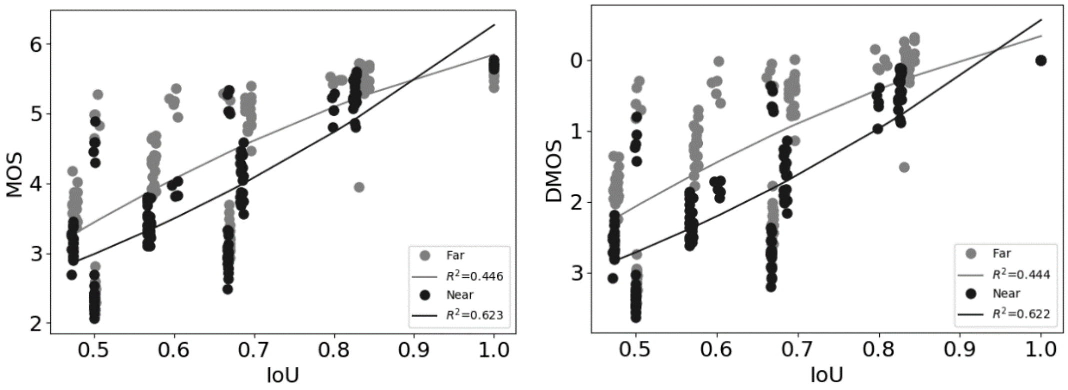
(a) (b)

Figure 4.19: Relationship between (a) IoU versus MOS, and (b) IoU versus DMOS in near and far view (Left side vehicle).

Table 4.6: Approximate formulas in near and far view (Left side vehicle)

Evaluation method	Distance	Formula(y: MOS, x: IoU)	R^2	MSE
MOS	Far	$\mathrm{y} = -2.843\mathrm{x}^2 + 8.633\mathrm{x} - 0.08518$	0.438	0.537
	Near	$\mathrm{y} = 1.162\mathrm{x}^2 + 3.811\mathrm{x} + 0.9475$	0.590	0.372
DMOS	Far	$\mathrm{y} = 2.828\mathrm{x}^2 - 8.611\mathrm{x} + 5.464$	0.436	0.543
	Near	$\mathrm{y} = -1.167\mathrm{x}^2 - 8.611\mathrm{x} + 5.464$	0.436	0.543

Figure 4.20 shows the results for the vehicle on the right. In the same way, approximate curves were calculated for the foreground and foreground for the vehicle on the right. The vertical axis in Figure 4.20 (a) is the subjective evaluation value, and the horizontal axis is IoU. The vertical axis in Figure 4.20 (b) is DMOS and the horizontal axis is IoU. Table 4.7 shows approximate equations for the vehicle on the right. From the experimental results, the foreground evaluation value of the left vehicle tended to be lower than that of the distant view. Similarly, the evaluation value of the foreground also tended to be lower than that of the distant view for the right-side vehicle. From the approximate curves of the left and right vehicles, a significant difference test was performed between the foreground and background. From the significance test, the p-value was 0.05 or less for the left and right vehicles, and there was a significant difference between the foreground and the background. Based on the above, the near view is sensitive to displacement, and the distant view is unlikely to be noticed.



(a) (b)

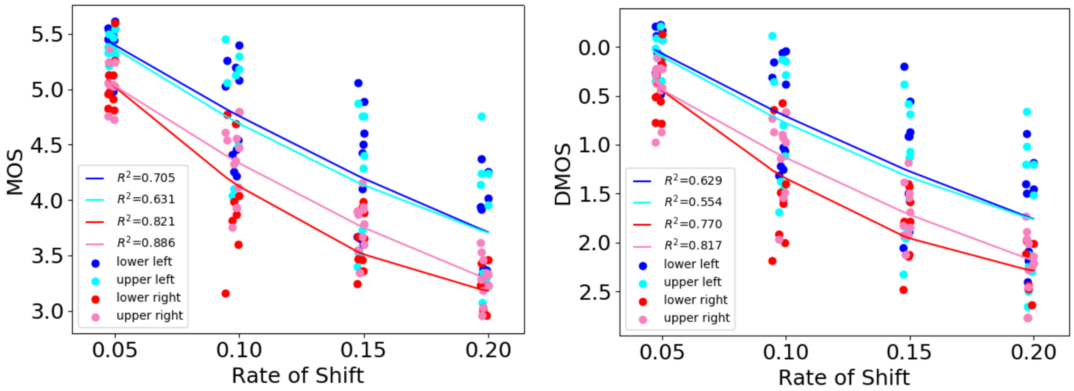
Figure 4.20: Relationship between (a) IoU versus MOS, and (b) IoU versus DMOS in near and far view (Right side vehicle)

Table 4.7: Approximate formulas in near and far view (Right side vehicle)

Evaluation method	Distance	Formula (y: MOS, x: IoU)	R^2	MSE
MOS	Far	$y = -3.515x^2 + 10.09x - 0.7277$	0.446	0.655
	Near	$y = 3.68x^2 + 1.05x + 1.539$	0.623	0.413
DMOS	Far	$y = 3.519x^2 - 10.1x + 6.241$	0.444	0.662
	Near	$y = -3.681x^2 - 1.05x + 4.164$	0.622	0.416

4.5.3 MOS and DMOS with misalignment

The experimental results focusing on the displacement in two directions with respect to the vertical and horizontal directions are shown below.



(a) (b)

Figure 4.21: Relationship between misalignment rate, MOS, and DMOS (Left side vehicle)

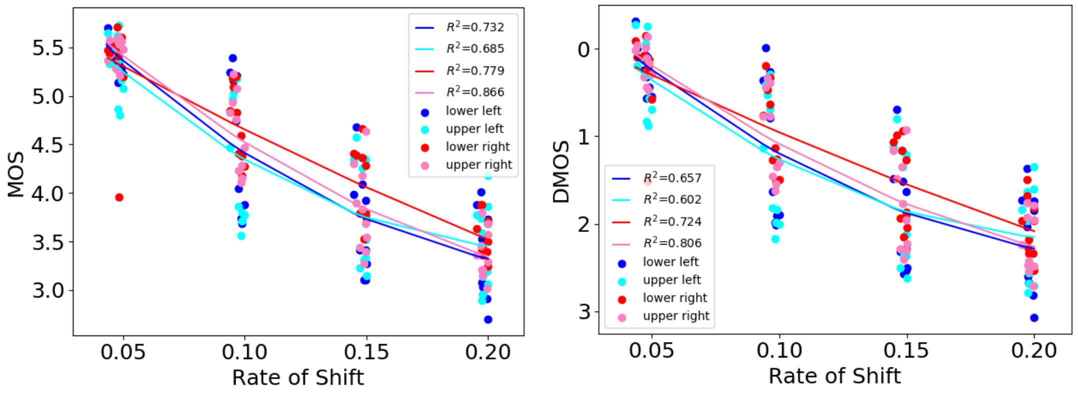
Figure 4.21 shows the experimental results focusing on misalignment in the left vehicle. In Figure 4.21 (a), the horizontal axis is the displacement rate, and the vertical axis is the average subjective evaluation value. An approximate curve was calculated for each misalignment direction. The points and curves shown in blue in Figure 4.21 (a) are when the position is shifted to the lower left. The light blue points and curves are the results when they are shifted to the upper left, the red colors are the results when they are shifted to the upper right, and the pink dots and curves are the results when they are shifted to the lower right. The DMOS results are shown in Figure 4.21 (b), and Table 4.8 shows the approximation formula, coefficient of determination, and mean square error when focusing on the displacement in two directions.

Table 4.8: Approximate formula when focusing on misalignment in two direction (Left side vehicle)

Evaluation method	**Deviation direction**	**Formula (y: MOS, x: IoU)**	R^2	**MSE**
MOS	Lower left	$\mathrm{y} = 15.75\mathrm{x}^2 - 15.2\mathrm{x} + 6.121$	0.705	0.167
	Upper left	$\mathrm{y} = 24.81\mathrm{x}^2 - 17.3\mathrm{x} + 6.172$	0.631	0.229
	Lower right	$\mathrm{y} = 57.23\mathrm{x}^2 - 26.58\mathrm{x} + 0.209$	0.821	0.109
	Upper right	$\mathrm{y} = 22.9\mathrm{x}^2 - 17.36\mathrm{x} + 5.841$	0.886	0.050
DMOS	Lower left	$\mathrm{y} = -15.55\mathrm{x}^2 + 15.15\mathrm{x} - 0.6502$	0.629	0.235
	Upper left	$\mathrm{y} = -24.61\mathrm{x}^2 + 17.25\mathrm{x} - 0.7015$	0.554	0.316
	Lower right	$\mathrm{y} = -57.03\mathrm{x}^2 + 26.53\mathrm{x} - 0.7384$	0.770	0.150
	Upper right	$\mathrm{y} = -22.7\mathrm{x}^2 + 17.31\mathrm{x} - 0.3701$	0.817	0.090

From the experimental results, a significant difference test was performed in the direction of displacement. As a result, there was a significant difference in the p-value of 0.05 or less for the left side displacement (upper left, lower left) and the right side displacement (upper right, lower right). . Based on the above, the left vehicle is considered to have anisotropy in the direction of displacement. In addition, since the evaluation of the right-side misalignment tends to be low, it may be sensitive to the misalignment toward the center line.

Figure 4.22 shows the experimental results for the vehicle on the right. As with the left vehicle, an approximation formula was calculated for each displacement direction for the right vehicle. Table 4.9 shows the approximation formula, the determination coefficient, and the mean square error when focusing on the misalignment in two directions.



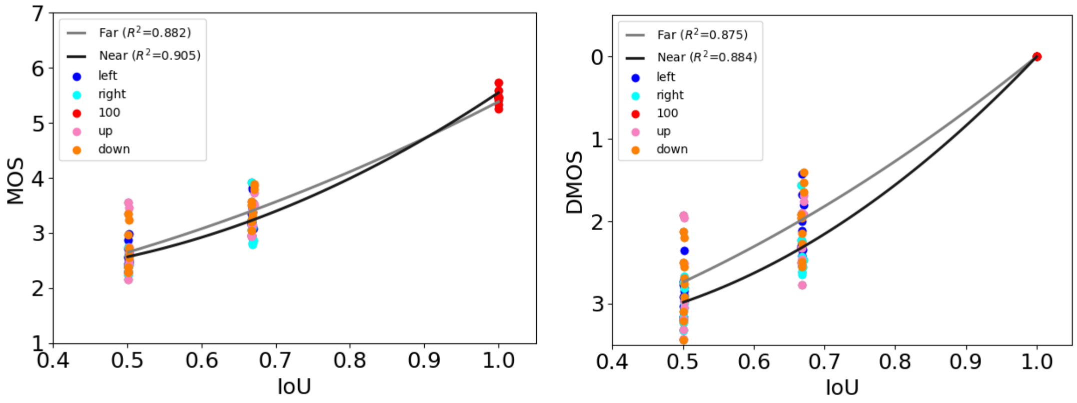
(a) (b)

Figure 4.22: Relationship between misalignment rate, MOS, and DMOS (Right side vehicle)

Table 4.9: Approximate formula when focusing on misalignment in two direction (Right side vehicle)

Evaluation method	**Deviation direction**	**Formula (y: MOS, x: IoU)**	R^2	**MSE**
MOS	Lower left	$y = 54.24x^2 - 27.16x + 6.586$	0.732	0.228
	Upper left	$y = 60.98x^2 - 27.18x + 6.453$	0.685	0.227
	Lower right	$y = 11.34x^2 - 14.72x + 6.013$	0.779	0.127
	Upper right	$y = 40.36x^2 - 23.86x + 6.505$	0.886	0.090
DMOS	Lower left	$y = -55.18x^2 + 27.43x - 0.9957$	0.657	0.327
	Upper left	$y = -61.91x^2 + 27.45x - 0.8627$	0.602	0.329
	Lower right	$y = -12.28x^2 + 14.99x - 0.4225$	0.724	0.173
	Upper right	$y = -41.3x^2 + 24.13x - 0.9145$	0.806	0.150

The experimental results focusing on the displacement in only one direction with respect to the vertical and horizontal directions are shown below.



(a) (b)

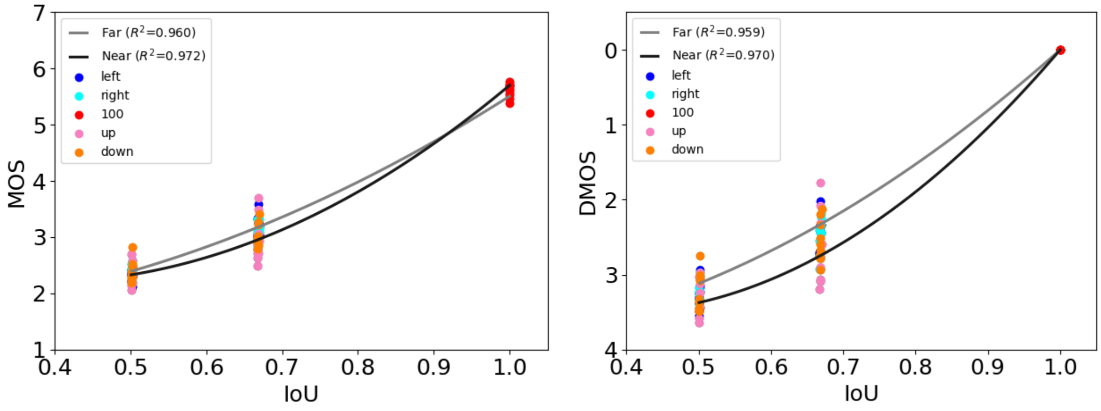
Figure 4.23: Relationship among misalignment rate in one direction, MOS, and DMOS (Left side vehicle)

Figure 4.23 shows the experimental results when the left vehicle has a positional shift only in one direction, up, down, left, or right. In Figure 4.23 (a), the horizontal axis is the evaluation value of IoU, and the vertical axis is MOS. In Figure 4.23 (b), the horizontal axis is the result when the vertical axis is DMOS. An evaluation image with an IoU of 0.5 has a displacement of 200%, and an evaluation image with an IoU of 0.67 has a displacement of 150%. In Figure 4.23, the approximate curve is drawn separately for the foreground and the background, the near view is shown in black, and the background is shown in gray. From Figure 4.23, the evaluation value became lower as the position shifted more. Table 4.10 shows the approximation formula, coefficient of determination, and mean square error when focusing on the displacement in one direction.

Table 4.10: Approximate formulas for misalignment in one direction (Left side vehicle)

Evaluation method	**Distance**	**Formula (y: MOS, x: IoU)**	$\boldsymbol{R^2}$	**MSE**
MOS	Near	$y = 6.072x^2 - 3.138x + 2.617$	0.905	0.080
	Far	$y = 2.928x^2 + 1.088x + 1.369$	0.882	0.080
DMOS	Near	$y = -6.074x^2 + 3.141x + 2.933$	0.884	0.103
	Far	$y = -2.928x^2 - 1.089x + 4.017$	0.875	0.090

Figure 4.24 shows the experimental results for the vehicle on the right. As for the vehicle on the right side, the approximate curve was calculated in the same way as the result for the vehicle on the left side. As a result, the subjective evaluation value tended to be low when the positional deviation increased. Table 4.11 shows the approximation formula, determination coefficient, and mean square error when focusing on the displacement in one direction.



(a) (b)

Figure 4.24: Relationship among misalignment rate in one direction, MOS, and DMOS (Right side vehicle)

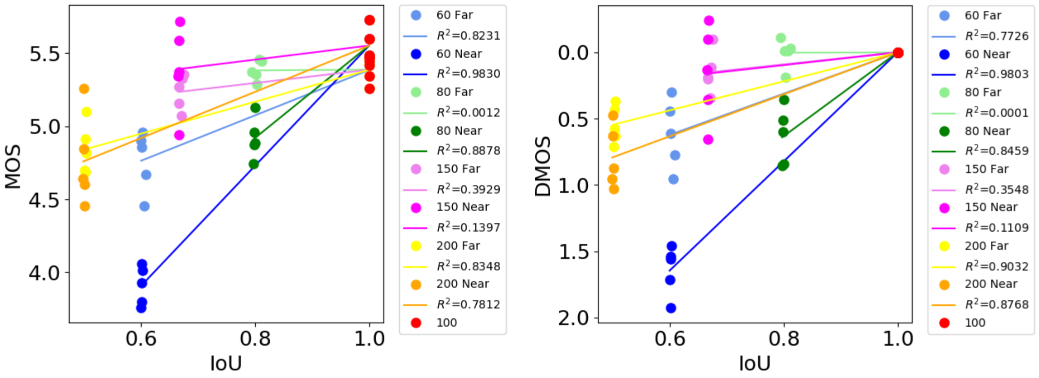
Table 4.11: Approximate formulas for misalignment in one direction (Right side vehicle)

Evaluation method	**Distance**	**Formula (y: MOS, x: IoU)**	R^2	**MSE**
MOS	Near	$y = 9.107x^2 - 6.912x + 3.508$	0.972	0.028
	Far	$y = 4.659x^2 - 0.7495x + 1.6$	0.960	0.035
DMOS	Near	$y = -9.107x^2 + 6.912x + 2.195$	0.970	0.030
	Far	$y = -4.658x^2 + 0.7489x + 3.91$	0.960	0.036

From the experimental results, the subjective evaluation value decreased as the displacement increased. In this experiment, the misregistration in only one direction evaluated was the misregistration including Ground truth, but the subjective evaluation value was “4: Normal” or less. From the above, even misalignment that includes Ground truth may be sensitive to misalignment.

4.5.4 Misalignment with ground truth for enlargement / reduction

The experimental results focusing on misalignment with respect to the ground truth are shown below.



(a) (b)

Figure 4.25: Relationship among IoU, MOS, and DMOS focusing on magnification with respect to ground truth (Left side vehicle)

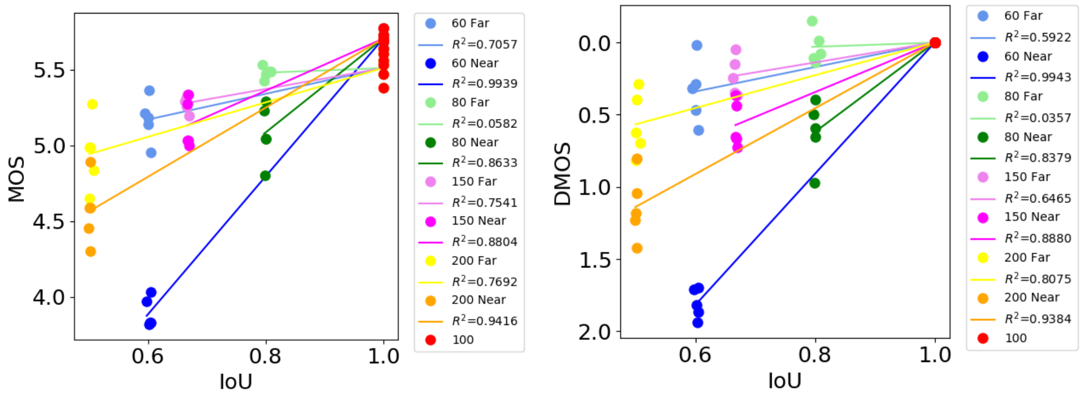
Figure 4.25 shows the results for the left vehicle. The horizontal axis in Figure 4.25 (a) is the evaluation value of IoU, and the vertical axis is MOS. For each enlargement factor, an approximate straight line was calculated using the mean square method for the near and far. The approximate straight line is a straight line that passes through the evaluation value of the defined vehicle region in order to see the change in the evaluation value with the defined vehicle region. Figure 4.25 (b) shows the results of DMOS. Table 4.12 shows the approximation formula, the determination coefficient, and the mean square error when focusing on misalignment. From Figure 4.25, the lowest evaluation value in the subjective evaluation value was in the case of 60% foreground. On the other hand, the lowest evaluation value in IoU was 200%.

Table 4.12: Approximate formulas for magnification (Left side vehicle)

Evaluation method	**Deviation rate [%]**	**Distance**	**Formula (y: MOS, x: IoU)**	R^2	**MSE**
MOS	60	Far	y = 1.555x + 3.83	0.8231	0.0203
		Near	y = 4.116x + 1.435	0.9830	0.0116
	80	Far	y = 0.02512x + 5.362	0.0012	0.0049
		Near	y = 3.166x + 2.386	0.8878	0.0127
	150	Far	y = 0.4548x + 4.932	0.3929	0.0086
		Near	y = 0.4821x + 5.070	0.1397	0.0396
	200	Far	y = 1.094x + 4.292	0.8348	0.0145
		Near	y = 1.583x + 3.968	0.7812	0.0437

DMOS	60	Far	$y = -1.554x + 1.555$	0.7726	0.0277
		Near	$y = -4.116x + 4.116$	0.9803	0.0134
	80	Far	$y = -0.0879x + 0.095$	0.0001	0.0048
		Near	$y = -3.164x + 3.164$	0.8459	0.0183
	150	Far	$y = -0.4574x + 0.456$	0.3548	0.0103
		Near	$y = -0.4821x + 0.482$	0.1109	0.0510
	200	Far	$y = -1.094x + 1.094$	0.9032	0.0079
		Near	$y = -1.584x + 1.584$	0.8768	0.0219

Figure 4.26 shows the experimental results for the vehicle on the right. The lowest objective evaluation value for the vehicle on the right was 60% of the foreground. In the IoU evaluation value, 200% was the lowest evaluation value, similar to the results for the left vehicle. Table 4.13 shows the approximation formula, the determination coefficient, and the mean square error when focusing on the misalignment. From the experimental results, the positional deviation when scaling down to 60% showed the lowest subjective evaluation value, and the foreground value tended to have a lower evaluation value than the distant view. Based on the above, it is considered that the subjective evaluation of the position shift for reduction tends to be lower than the position shift for enlargement.



(a) (b)

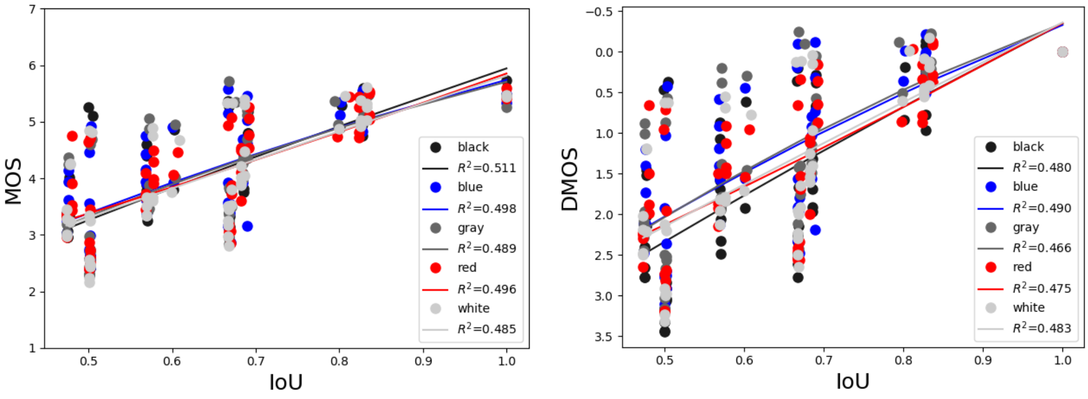
Figure 4.26: Relationship among IoU, MOS, and DMOS focusing on magnification with respect to ground truth (Right side vehicle)

Table 4.13: Approximate formulas for magnification (Right side vehicle)

Evaluation method	Deviation rate [%]	Distance	Formula (y: MOS, x: IoU)	R^2	MSE
MOS	60	Far	$y = 0.8469x + 4.662$	0.7057	0.0119
		Near	$y = 4.542x + 1.161$	0.9930	0.0040
	80	Far	$y = 0.1524x + 5.357$	0.0582	0.0036
		Near	$y = 3.095x + 2.608$	0.8633	0.0152
	150	Far	$y = 0.689x + 4.82$	0.7541	0.0043
		Near	$y = 1.716x + 3.987$	0.8804	0.0109
	200	Far	$y = 1.134x + 4.375$	0.7692	0.0238
		Near	$y = 2.277x + 3.426$	0.9416	0.0200
DMOS	60	Far	$y = -0.847x + 0.8478$	0.5923	0.0197
		Near	$y = -4.541x + 4.541$	0.9943	0.0046
	80	Far	$y = -6.143x + 0.1443$	0.0357	0.0054
		Near	$y = -3.097x + 3.097$	0.8379	0.0187
	150	Far	$y = -0.6895x + 0.689$	0.6465	0.0072
		Near	$y = -1.716x + 1.716$	0.8880	0.0102
	200	Far	$y = -1.134x + 1.134$	0.8075	0.0189
		Near	$y = -2.277x + 2.277$	0.9384	0.0211

4.5.5 Relationship among IoU, MOS, and DMOS based on vehicle color

The experimental results focusing on different vehicle color are described in this section. Figure 4.27 shows the relationship between IoU, MOS, and DMOS when focusing on the color of the vehicle. The vertical axis in Figure 4.27 (a) is MOS, and the horizontal axis is the evaluation value in IoU. In Figure 4.27 (a), black is a black vehicle, blue is a blue vehicle, gray is a gray vehicle, red is a red vehicle, and white is a white vehicle. An approximate expression was calculated for each color of each vehicle, and the coefficient of determination for each was obtained. In Figure 4.27 (a), the coefficient of determination of the approximate expression for a black vehicle was 0.511. Similarly, approximate curves for blue, gray, red, and white and their determination coefficients are shown in Figure 4.27 (a). Figure 4.27 (b) shows the DMOS results. In the DMOS results, an approximate expression was calculated for each vehicle color, and its coefficient of determination was calculated. Table 4.14 shows the approximate formula, determination coefficient, and mean square error when focusing on the color of the vehicle. The experimental results for the vehicle on the right are shown below.



(a) (b)

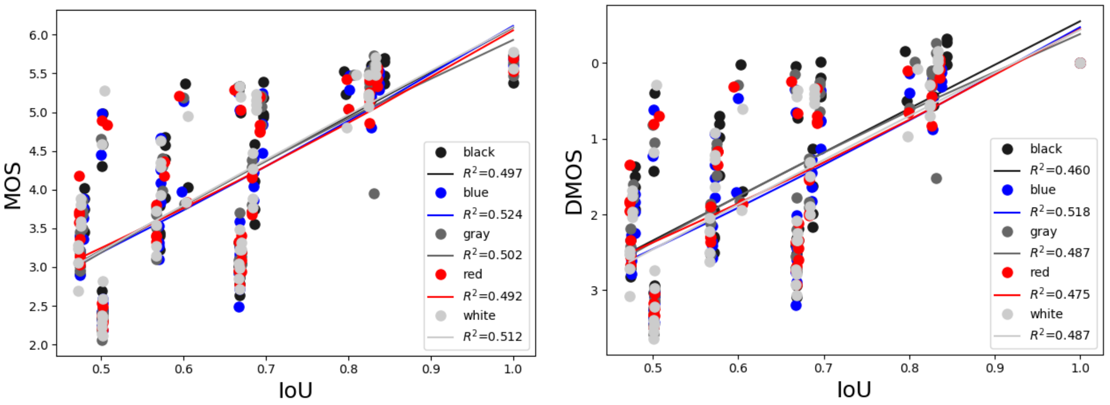
Figure 4.27: Relationship among IoU, MOS and DMOS focusing on vehicle color (Left side vehicle).

Table 4.14: Approximate formulas focusing on vehicle color (Left side vehicle)

Evaluation method	**Color**	**Formula (y: MOS, x: IoU)**	R^2	**MSE**
MOS	Black	$y = -0.7852x^2 + 6.547x + 0.1827$	0.511	0.527
	Blue	$y = -1.7x^2 + 7.268x + 0.1729$	0.498	0.455
	Gray	$y = -2.263x^2 + 8.16x - 0.1882$	0.489	0.500
	Red	$y = 0.695x^2 + 3.911x + 1.248$	0.496	0.443
	White	$y = -0.3599x^2 + 5.561x + 0.6145$	0.485	0.501
DMOS	Black	$y = 0.8136x^2 - 6.591x + 5.433$	0.480	0.598
	Blue	$y = 1.711x^2 - 7.287x + 5.248$	0.490	0.471
	Gray	$y = 2.271x^2 - 8.178x + 5.561$	0.466	0.550
	Red	$y = -0.6784x^2 - 3.951x + 4278$	0.475	0.486
	White	$y = 0.3649x^2 - 5.57x + 4.846$	0.483	0.506

Figure 4.28 shows the relationship between MOS, DMOS, and IoU in the vehicle on the right, and approximate curves were calculated for each vehicle color. Table 4.15 shows the approximation formula, coefficient of determination, and mean square error when

focusing on the color of the vehicle. From the experimental results, there was no difference between the subjective evaluation value and the trend of IoU for each vehicle color in the right vehicle. From the approximate curve in the experimental results, there was no significant difference for each color of the vehicle. From the above, we think that there is no difference in subjective evaluation depending on the color of the vehicle.



(a) (b)

Figure 4.28: Relationship among IoU, MOS and DMOS focusing on vehicle color (Right side vehicle).

Table 4.15: Approximate formulas focusing on vehicle color (Right side vehicle)

Evaluation method	**Color**	**Formula (y: MOS, x: IoU)**	R^2	**MSE**
MOS	Black	$y = -0.1019x^2 + 5.911x + 0.2825$	0.497	0.630
	Blue	$y = 0.9691x^2 + 4.395x + 0.7484$	0.525	0.554
	Gray	$y = -1.401x^2 + 7.591x - 0.2592$	0.502	0.592
	Red	$y = 1.175x^2 + 3.858x + 1.024$	0.492	0.573
	White	$y = -0.3553x^2 + 6.315x + 0.1385$	0.512	0.599
DMOS	Black	$y = 0.08438x^2 - 5.912x + 5.279$	0.460	0.738
	Blue	$y = -0.9706x^2 - 4.397x + 4.898$	0.518	0.570
	Gray	$y = 1.418x^2 - 7.624x + 5.826$	0.487	0.631
	Red	$y = -1.18x^2 - 3.858x + 4.593$	0.475	0.616
	White	$y = 0.3564x^2 - 6.329x + 5.536$	0.487	0.668

4.6 Conclusion

In this chapter, we focused on IoU, an evaluation standard widely used in vehicle detection technology, and conducted a subjective evaluation experiment on detection accuracy for manually created misalignment. We calculated the linear and nonlinear regression equations using linear, quadratic, and cubic polynomial function to analysis the subjective evaluation scores.

Chapter 5

HUMAN PERCEPTION AND IOU BASED ON VARIATION OF ANGLES AND DISTANCE

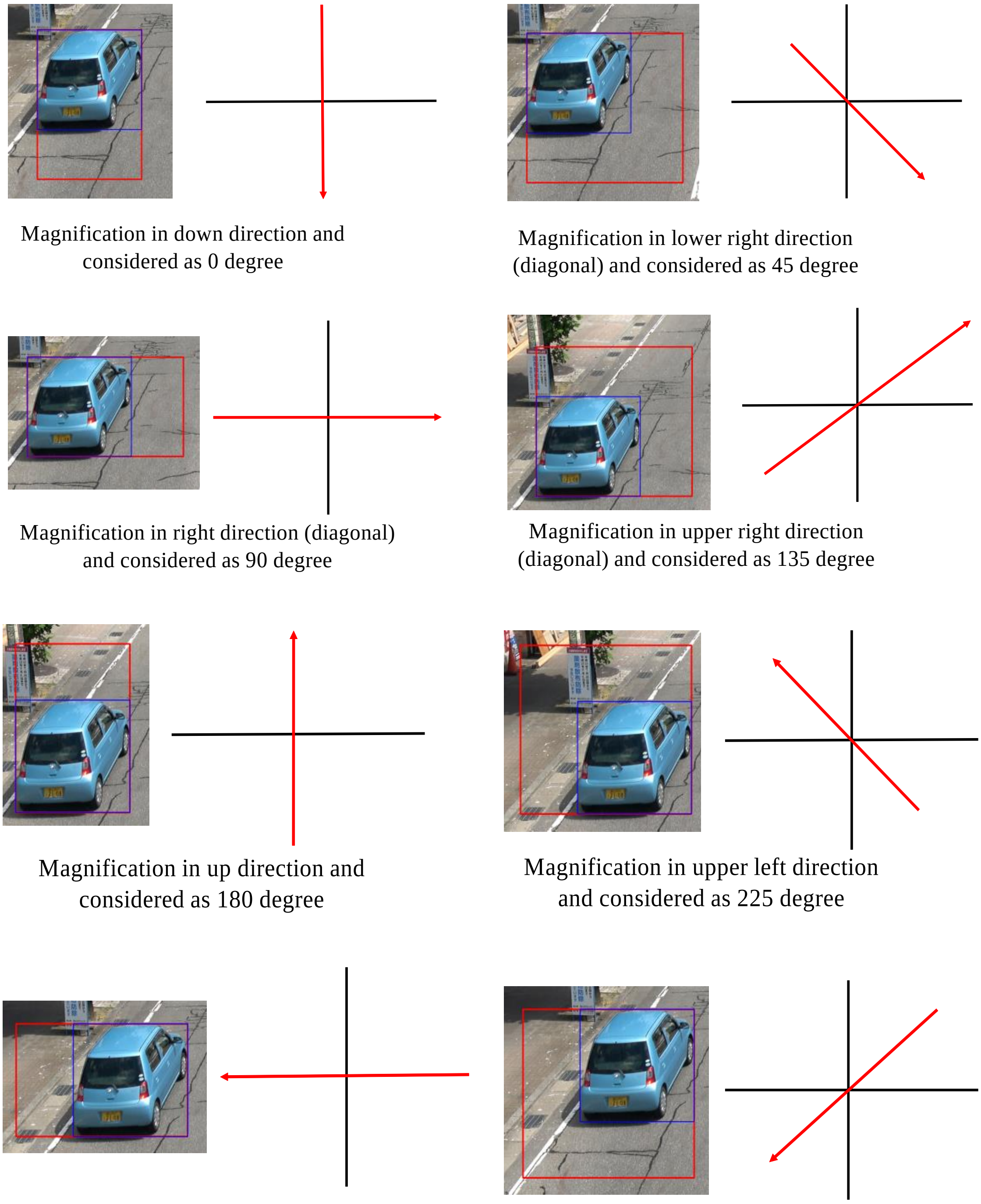
In this chapter, we compared human perception and Intersection over Union (IoU) for detection based on variation of angles and distances. Experimental results show the relationship between IoU and subjective evaluation values, which reflect different behaviour between human perception based evaluation and IoU.

5.1 Data preparation

In this experiment, we also prepared magnified rectangles in eight directions for making a relationship between IoU and degree. The images are separated based on far and near view. There eight magnified rectangles of the every selected image are generated using ground truth. For magnification, a magnification of 2X, 1.5X, 0.8X, and 0.6X against ground truth was applied, and then eight magnified rectangles in eight directions for every magnified rectangle were prepared. Figure 5.1 and Figure 5.2 show the magnified and shifted rectangles and their direction. The prepared shifted rectangles and magnified rectangles in the far and near view are used to make relationship between IoU and degree. The blue rectangles represent ground truth and the red rectangles represent magnified and shifted rectangles. The eight direction and their corresponding angles are summarized in Table 5.1.

5.2 Subjective assessment test

We conducted a subjective assessment of the accuracy of vehicle detection. Subjects evaluated the detection accuracy of vehicles surrounded by red rectangles using seven categories from 1 (very bad) to 7 (excellent). In our experiment, 70 men and women are participated as the subjects as described in the previous chapter. IoU is used to evaluate shifted images objectively.



Magnification in down direction and considered as 0 degree

Magnification in lower right direction (diagonal) and considered as 45 degree

Magnification in right direction (diagonal) and considered as 90 degree

Magnification in upper right direction (diagonal) and considered as 135 degree

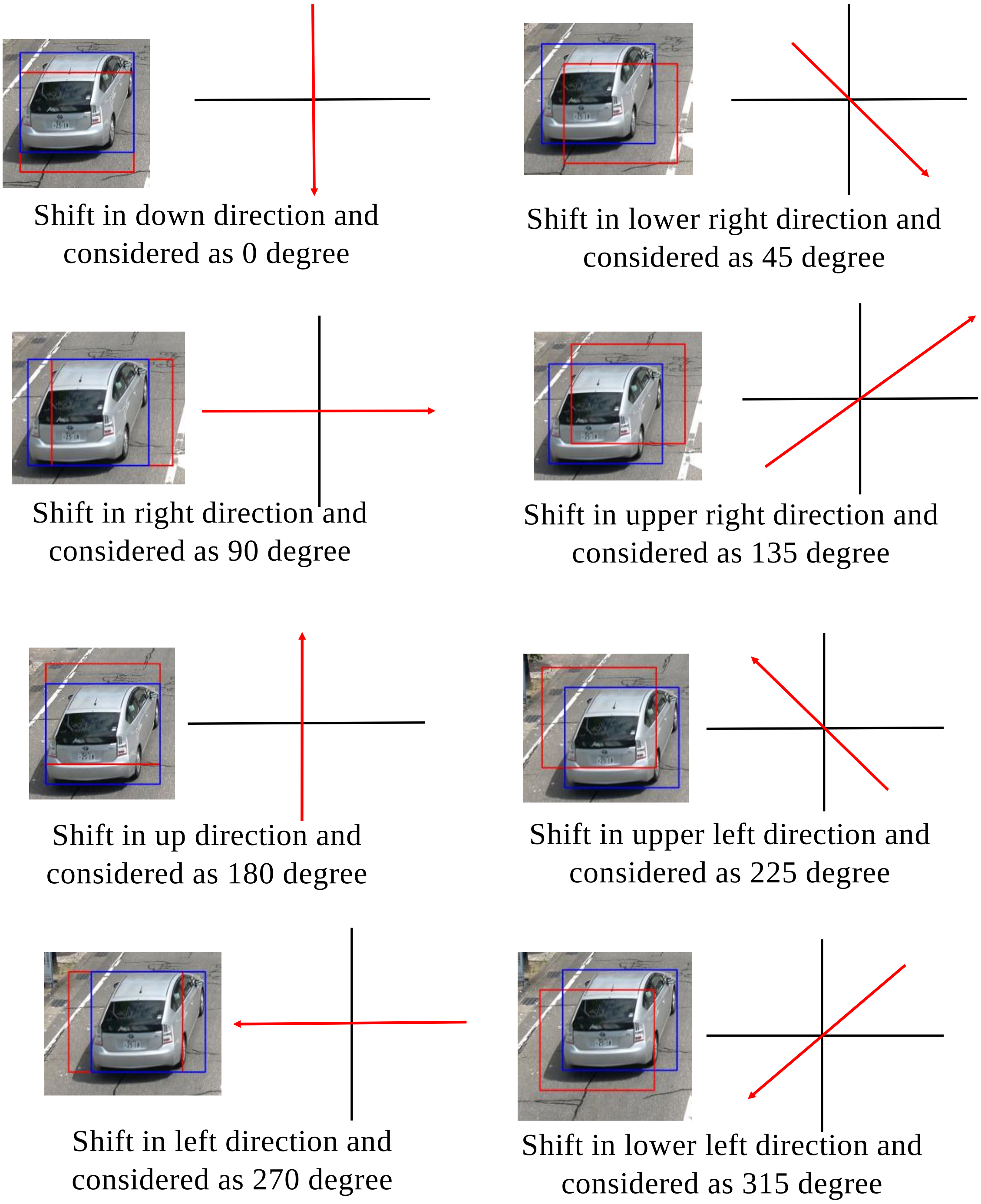
Magnification in up direction and considered as 180 degree

Magnification in upper left direction and considered as 225 degree

Magnification in left direction and considered as 270 degree

Magnification in lower left direction and considered as 315 degree

Figure 5.1: Magnification in eight direction.



Shift in down direction and considered as 0 degree

Shift in lower right direction and considered as 45 degree

Shift in right direction and considered as 90 degree

Shift in upper right direction and considered as 135 degree

Shift in up direction and considered as 180 degree

Shift in upper left direction and considered as 225 degree

Shift in left direction and considered as 270 degree

Shift in lower left direction and considered as 315 degree

Figure 5.2: Shifting in eight direction.

Table 5.1: Misalignment in eight direction

Shifting Direction	Angles (Degree)
Lower (Down)	0
Lower Right (Diagonal)	45
Right	90
Upper Right (Diagonal)	135
Upper	180
Upper Left (Diagonal)	225
Left	270
Lower Left (Diagonal)	315

5.3 Analysis of the experimental results

To analysis the evaluation results based on human perception and IoU, the relationships between IoU and shifting direction are presented in Figure 5.3 and Figure 5.4 for far and near view, respectively. The relationships between magnification direction and IoU for far and near view are shown in Figure 5.5-5.6.

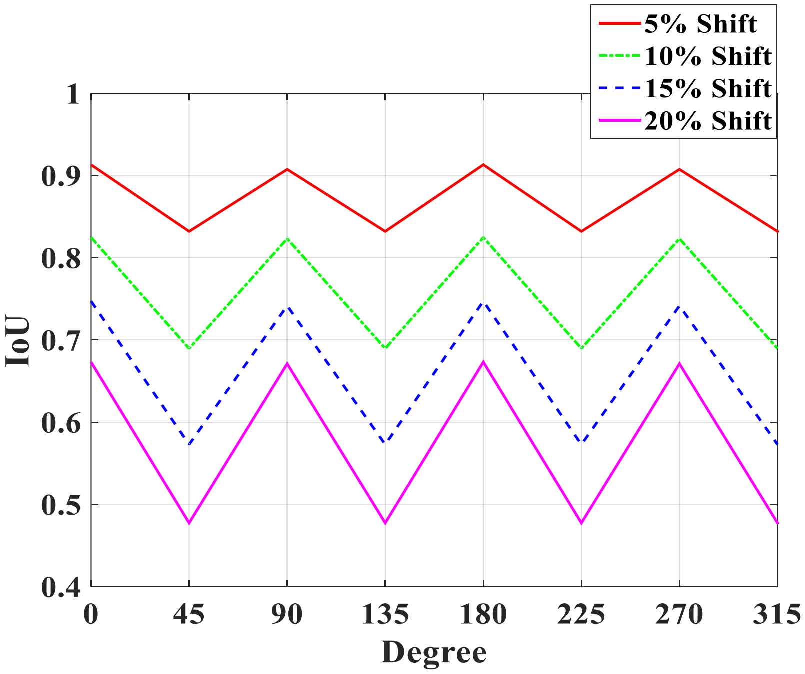


Figure 5.3: Degree versus IoU in far view after misalignment with shifting.

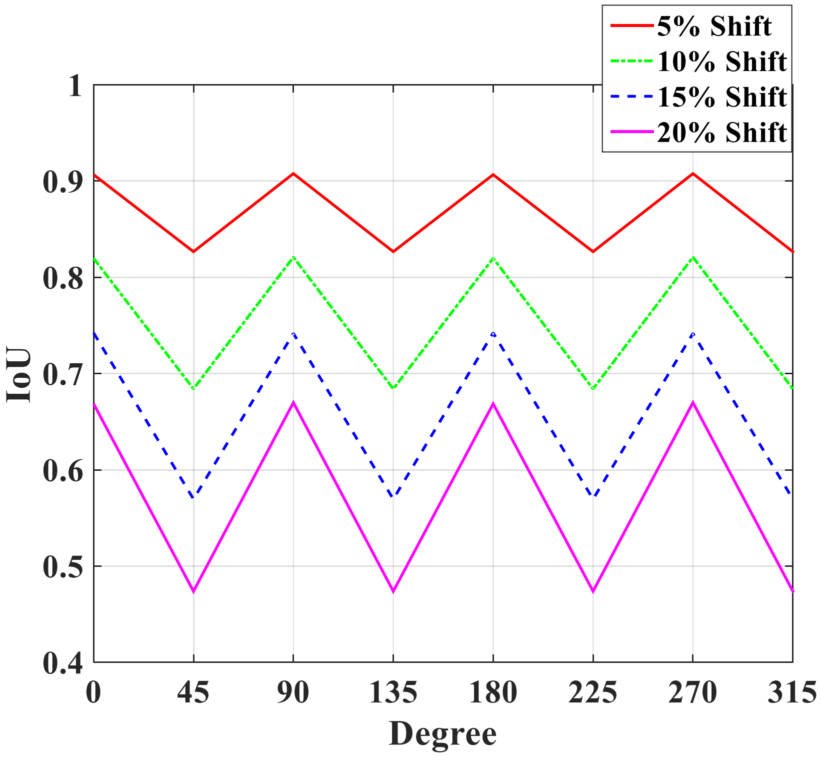


Figure 5.4: Degree versus IoU in near view after misalignment with shifting.

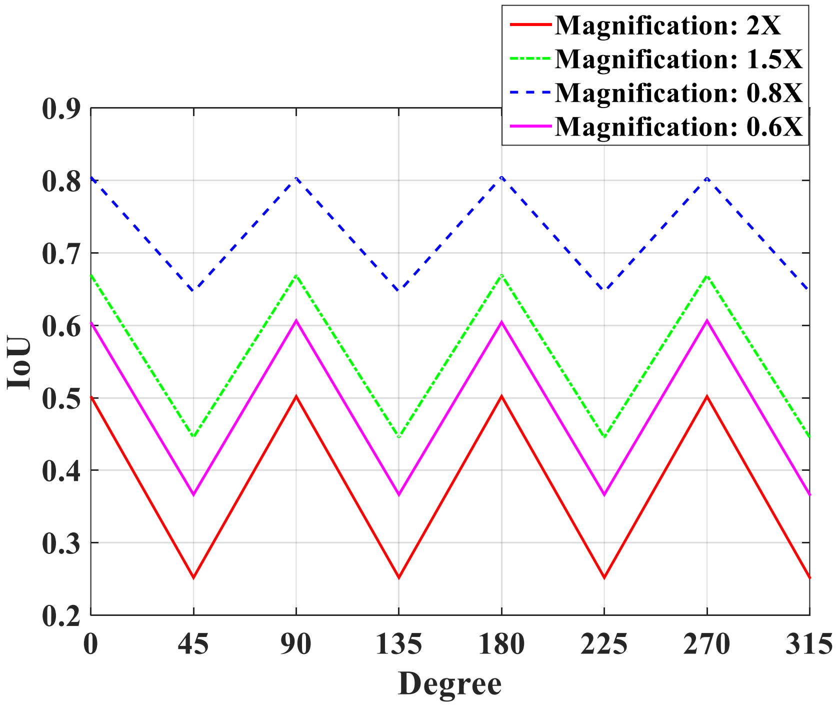


Figure 5.5: Degree versus IoU in far view after misalignment with magnification.

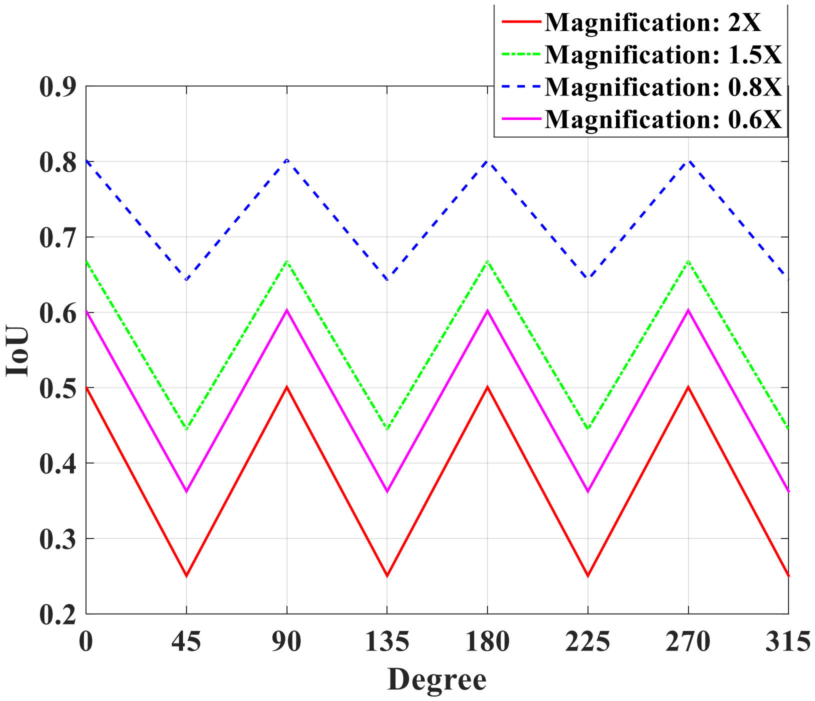


Figure 5.6: Degree versus IoU in near view after misalignment with magnification.

Figure 5.3-5.6 indicate that there is no significance difference of the performance evaluation between far and near view using IoU. On the other hand, there have significance difference using human perception based evaluation as we explained in our previous chapter. Now we explained more based on far and near view, using one and two-way ANOVA test. We analysed the subjective evaluation scores in three different ways. In the first way, the subjective evaluation scores are divided into three groups. The three groups are 0.8X magnification, 0.6X magnification, and without magnification. After, we applied one way ANOVA on this arrangement of the scores. From the results of the analysis, we found that there is a significance difference between groups, which are shown in Table 5.2-5.5 for MOS and DMOS.

Table 5.2 One-way ANOVA in far view for MOS

Source of Variation	Fluctuation	Degree of freedom	Dispersion	Dispersion ratio	P-value	F Critical
Between Groups	1.263701	2	0.63185	35.06304	9.73E-06	3.885294
Within Group	0.216245	12	0.01802			

Table 5.3 One-way ANOVA in near view for MOS

Source of Variation	Fluctuation	Degree of freedom	Dispersion	Dispersion ratio	P-value	F Critical
Between Groups	6.838966	2	3.419483	206.2634	5.1E-10	3.885294
Within Group	0.198939	12	0.016578			

Table 5.4 One-way ANOVA in far view for DMOS

Source of Variation	Fluctuation	Degree of freedom	Dispersion	Dispersion ratio	P-value	F Critical
Between Groups	1.263701	2	0.63185	23.76106	6.71E-05	3.885294
Within Group	0.319102	12	0.026592			

Table 5.5 One-way ANOVA in near view for DMOS

Source of Variation	Fluctuation	Degree of freedom	Dispersion	Dispersion ratio	P-value	F Critical
Between Groups	6.838966	2	3.419483	126.6156	8.58E-09	3.885294
Within Group	0.324082	12	0.027007			

In the second way, the subjective evaluation scores are divided into five groups based on direction of the 1.5X and 2X magnification. The five groups are named as ‘All Direction’, ‘Upper’, ‘Lower’, ‘Left’, and ‘Right’. Table 5.6-5.9 show the results of two-way ANOVA that investigate main effect and interaction of five direction and two magnifications.

Table 5.6 Two-way ANOVA in far view for MOS

Source of Variation	**Fluctuation**	**Degree of freedom**	**Dispersion**	**Dispersion ratio**	**P-value**	**F critical**
Direction	33.51607	4	8.379016	126.2521	4.02E-22	2.605975
Magnification	5.828673	1	5.828673	87.82442	1.21E-11	4.084746
Interaction	0.562041	4	0.14051	2.117159	0.09653	2.605975

Table 5.7 Two-way ANOVA in near view for MOS

Source of Variation	**Fluctuation**	**Degree of freedom**	**Dispersion**	**Dispersion ratio**	**P-value**	**F critical**
Direction	38.17202	4	9.543006	96.70699	5.22E-20	2.605975
Magnification	5.360473	1	5.360473	54.32201	5.75E-09	4.084746
Interaction	0.211649	4	0.052912	0.536203	0.709908	2.605975

Table 5.8 Two-way ANOVA in far view for DMOS

Source of Variation	**Fluctuation**	**Degree of freedom**	**Dispersion**	**Dispersion ratio**	**P-value**	**F critical**
Direction	33.51607	4	8.379016	112.5625	3.31E-21	2.605975
Magnification	5.828673	1	5.828673	78.30158	5.81E-11	4.084746
Interaction	0.562041	4	0.14051	1.887594	0.131478	2.605975

Table 5.9 Two-way ANOVA in near view for DMOS

Source of Variation	**Fluctuation**	**Degree of freedom**	**Dispersion**	**Dispersion ratio**	**P-value**	**F critical**
Direction	38.17202	4	9.543006	77.09677	2.96E-18	2.605975
Magnification	5.360473	1	5.360473	43.3066	7.22E-08	4.084746
Interaction	0.211649	4	0.052912	0.427471	0.787891	2.605975

From the table, we found significance difference among different directions, and in magnifications. However, there is no significance difference in interaction.

In the third way, the subjective evaluation scores are divided into four groups based on direction of the misalignment for shifting. The four groups are named as ‘Lower Right’, ‘Upper Right’, ‘Lower Left’, and ‘Upper Left’. Table 5.10-5.13 show the results of two-way ANOVA that investigate main effect and interaction on the four directions or angles with shifting operations. From the table, we found significance difference among different directions, and shifting. However, there is no significance difference in interaction.

Table 5.10 Two-way ANOVA in far view for MOS

Source of Variation	**Fluctuation**	**Degree of freedom**	**Dispersion**	**Dispersion ratio**	**P-value**	**F critical**
Direction	9.779694	3	3.259898	48.99377	1.43E-16	2.748191
Shifting	27.02039	3	9.006796	135.3652	1.17E-27	2.748191
Interaction	1.090122	9	0.121125	1.820412	0.081531	2.029792

Table 5.11 Two-way ANOVA in near view for MOS

Source of Variation	**Fluctuation**	**Degree of freedom**	**Dispersion**	**Dispersion ratio**	**P-value**	**F critical**
Direction	1.556265	3	0.518755	12.86305	1.12E-06	2.748191

Shifting	43.6928	3	14.56427	361.1356	4.84E-40	2.748191
Interaction	0.307592	9	0.034177	0.84745	0.575855	2.029792

Table 5.12 Two-way ANOVA in far view for DMOS

Source of Variation	Fluctuation	Degree of freedom	Dispersion	Dispersion ratio	P-value	F critical
Direction	9.779694	3	3.259898	45.27877	8.04E-16	2.748191
Shifting	27.02039	3	9.006796	125.101	1.01E-26	2.748191
Interaction	1.090122	9	0.121125	1.682377	0.111698	2.029792

Table 5.13 Two-way ANOVA in near view for DMOS

Source of Variation	Fluctuation	Degree of freedom	Dispersion	Dispersion ratio	P-value	F critical
Direction	1.556265	3	0.518755	8.598575	7.05E-05	2.748191
Shifting	43.6928	3	14.56427	241.4086	7.89E-35	2.748191
Interaction	0.307592	9	0.034177	0.566495	0.819382	2.029792

Experimental results showed that there was a significant difference in subjective evaluation values between far and near view with misalignment in different angles or directions. Based on the experimental results in this work, the following indices are proposed based on the relationship between subjective evaluation values and IoU:

$$Very\ good: if\ IoU > 0.90$$
$$Good: if\ 0.75\ < IoU\ \leq 0.90$$

Conventionally, the threshold of IoU used in vehicle detection is 0.5 or 0.7, but it is considered that 0.75 or more is necessary from the experimental results. Based on the above, IoU's evaluation criteria widely used in vehicle detection technology can use an index that considers subjective evaluation. From the experimental results, the evaluation using IoU can be expected to take into account subjective evaluation by using 0.75 or more.

Chapter 6

CONCLUSIONS AND FUTURE DIRECTION

It is indubitable that vehicle detection and evaluation is very important research topic nowadays. In this study, vehicle detection and performance analysis in compressed domain are carried out using a novel biologically inspired approach in combination with local image features. The proposed model of spatial saliency (visual attention) combined with image features significantly accelerates detection performance as compared with other saliency based methods, from complex background as demonstrated by the experimental results. Furthermore, the one-to-one symmetric search helps to detect the overlapping objects, where previous methods fail to detect overlapping objects in the context of spatial saliency based methods. Although, deep learning based methods [24-26] showed usefulness for vehicle detection, our method is completely different concept from deep learning based method. Deep learning based algorithms expensive to train due to complex data models. Moreover, deep learning requires expensive GPUs and hundreds of machines. This increases cost to the users. In our case, we use a database but there is no need to train like deep learning based algorithm. The proposed method is compared with conventional saliency methods due to utilizing saliency information and local features. The performance of the conventional saliency methods is not good, therefore we combined with SIFT, Harris, and saliency information for detection, which performs better than conventional methods. However, we will compare in future with the deep learning based methods. In this study, We use 4K video due to their high resolution and good image quality. Experimental results show that the propose method is able to detect desired object such as vehicles, from 4K video. Combination of SIFT and Harris features provide large number of features resulting in a good coverage of the vehicle region, which lead to improve detection performance. One more advantage point of the SIFT and Harris features is that they are less sensitive on shadow as compared to motion feature. Our main goal is to improve detection performance by taking advantages of 4K image sequences and develop video quality model. For facilitating real time road monitoring in future, this analysis results of the detection performance in compressed domain will lead us to develop video quality model for detection by transmitting reasonable high-quality video.

In this research, subjective evaluation experiments were also conducted to investigate the relationship between IoU and subjective evaluation values. Experimental results showed that there was a significant difference in subjective evaluation values between left and right misalignment, far and near view, and among colors of the vehicles. Based on the experimental results in this paper, we proposed evaluation model based on the relationship between subjective evaluation values and IoU for overcoming the limitations of the widely used evaluation method namely IoU.

In the present model, we used nearest neighbor search algorithm, which is not suitable for real time road monitoring. Therefore, we will enhance at this part of our model in our future work. The proposed subjective evaluation model can be used in future for evaluation of the detection in industrial application due to limitations of IoU.

List of Publications

1. **Most Shelina Aktar**, Yuukou Horita, “Performance Analysis of Vehicle Detection Based on Spatial Saliency and Local Image Features in H.265 (HEVC) 4K video for Developing a Relationship between IoU and Subjective Evaluation Value”, IEEJ Transactions on Electrical and Electronic Engineering. **(Accepted)**

2. Naho ITO, **Most Shelina Aktar,** Yuukou Horita, “Construction of Subjective Vehicle Detection Evaluation Model Considering Shift from Ground Truth Position”, *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, Vol. E102-A, No. 9, 2019.

3. ***Most Shelina Aktar***, K M Ibrahim Khalilullah, Kautaro Katayama, Yuukou Horita, “Spatial Saliency and Local Image Features Based Vehicle Detection from 4K Image Sequences”, *The Ninth International Workshop on Image Media Quality and its Applications (IMQA2018),* September, 2018, Kobe, Japan.

REFERENCES

[1] G. Bhosle, et al., “Vehicle Tracking Using Image Processing,” IJRASET, vol. 6, no. 1, pp. 1235–38, 2018.

[2]. Ji, X.; Wei, Z.; and Feng, Y. (2006). Effective vehicle detection techniques for traffic surveillance systems. *Journal of Visual Communication and Image Representation*, 17(3), 647-658.

[3]. Lozano, A.; Manfredi, G.; and Nieddu, L. (2009). An algorithm for the recognition of levels of congestion in road traffic problems. *Mathematics and Computers in Simulation*, 79(6), 1926-1934.

[4]. Zhou, J.; Gao, D.; and Zhang, D. (2007). Moving vehicle detection for automatic traffic monitoring. *IEEE Transactions on Vehicle Technology*, 56(1), 51-59.

[5]. Niu, X. (2006). A semi-automatic framework for highway extraction and vehicle detection based on a geometric deformable model. *ISPRS Journal of Photography and Remote Sensing*, 61(3-4), 170-186.

[6]. Zhang, W.; Fang, X.Z.; and Yang, X. (2006). Moving vehicles segmentation based on Bayesian framework for Gaussian motion model. *Pattern Recognition Letters*, 27(1), 956-967.

[7]. Li, X.; Liu, Z.Q.; and Leung, K.M. (2002). Detection of vehicles from traffic scenes using fuzzy integrals. *Pattern Recognition*, 35(4), 967-980.

[8] H. Chung-Lin and L. Wen-Chieh, "A vision-based vehicle identification system," in Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on, 2004, pp. 364- 367 Vol.4.

[9] Z. Wei, et al., "Multilevel Framework to Detect and Handle Vehicle Occlusion," Intelligent Transportation Systems, IEEE Transactions on, vol. 9, pp. 161-174, 2008.

[10] N. K. Kanhere and S. T. Birchfield, "Real-Time Incremental Segmentation and Tracking of Vehicles at Low Camera Angles Using Stable Features," Intelligent Transportation Systems, IEEE Transactions on, vol. 9, pp. 148-160, 2008.

[11] N. K. Kanhere, "Vision-based detection, tracking and classification of vehicles using stable features with automatic camera calibration," ed, 2008, p. 105.

[12] Itti L, Koch C, Niebur E. 1998. A model of saliency-based visual attention for rapid scene analysis. IEEE Trans. Pattern Anal. Mach. Intell. 20, 1254–1259.

[13] Rahtu E, Kannala J., Salo M. and Heikkil J (2010) Segmenting salient objects from images and Videos. Proc. European Conference on Computer Vision (ECCV 2010).

[14] T. N. Vikram, M. Tscherepanow, and B. Wrede, A saliency map based on sampling an image into random rectangular regions of interest, Pattern Recognition, vol. 45, issue: 9, pp. 3114-3124, 2012.

[15] N. Imamoglu,W. Lin and Y. Fang, ”A Saliency Detection Model Using Low-Level Features Based on Wavelet Transform,” in IEEE Transactions on Multimedia, vol. 15, no. 1, pp. 96-105, Jan. 2013.

[16] Weining Wang, Dong Cai, Xiangmin Xu, Alan Wee- Chung Liew, Visual saliency detection based on region descriptors and prior knowledge, Signal Process.: Image Communication, 29 (3) (2014) 424433.

[17] D.B. Walther, C. Koch, Attention in hierarchical models of object recognition, Progress in Brain Research 165 (2007).

[18] C. Guo, L. Zhang, A novel multiresolution spatiotemporal saliency detection model and its applications n image and video compression, IEEE Transactions on Image Processing 19 (2010) pp. 185-198.

[19] N.D.B. Bruce, J.K. Tsotsos, Saliency based on information maximization, Advances in Neural Information Processing Systems (2005) pp. 155-162.

[20] D. Gao, V. Mahadevan, N. Vasconcelos, The discriminant center-surround hypothesis for bottom-up saliency, Advances in Neural Information Processing Systems (2008) pp. 497-504.

[21] W. Kienzle, F.A. Wichmann, B. Scholkopf, M.O. Franz, A nonparametric approach to bottom-up visual saliency, Advances in Neural Information Processing Systems (2007) pp. 689-696.

[22] H. J. Seo, and P. Milanfar, Visual saliency for automatic target detection, boundary detection, and image quality assessment, IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP), pp. 5578-5581, 2010.

[23] J. Harel, C. Koch, P. Perona, Graph-based visual saliency, Advances in Neural Information Processing Systems (2007) pp.545–552.

[24] Ming-Ming Cheng, Niloy J Mitra, Xiaolei Huang, Philip HS Torr, and Shi-Min Hu. Global contrast based salient region detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(3):569-582, 2015.

[25] Ali Borji, Ming-Ming Cheng, Huaizu Jiang, and Jia Li. Salient object detection: A benchmark. *IEEE transactions on image processing*, 24(12):5706-5722, 2015.

[26] Andreas Geiger, Philip Lenz, and Raquel Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. pp. 3354-3361, 2012.

[27] Andreas Geiger, Philip Lenz, Christoph Stiller, and Raquel Urtasun. Vision meets robotics: The kitti dataset. *The International Journal of RoboticsResearch*, 32(11):1231-1237, 2013.

[28] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The cityscapes dataset for semantic urban scene understanding. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3213-3223, 2016.

[29] Mark Everingham, SM Ali Eslami, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. The pascal visual object classes challenge: A retrospective. *International journal of computer vision*, 111(1):98-136, 2015.

[30] Koch, C. and Ullman, S. Shifts in selective visual attention: Towards the underlying neural circuitry. Hum. Neurobiol. **1985**, 4, 219–227.

[31] Niebur E, Koch C. 1996. Control of selective visual attention: modeling the ‘where’ pathway. In Advances in Neural *Information Processing Systems 8 (NIPS 1995). Advances in Neural Information Processing Systems*, no. 8. pp. 802–808. Cambridge, MA: MIT Press.

[32] W. Reichardt, “Evaluation of optical motion information by movement detectors,” Journal of Comparative Physiology A, 161(4), pp. 533–547, 1987.

[33] T. Tuytelaars, and K. Mikolajczyk, “Local invariant feature detectors: A survey,” Foundations and Trends in Computer Graphics and Vision, 3 (3), pp. 177-280, 2008.

[34] C. Harris and M.J. Stephens. A combined corner and edge detector. In Alvey Vision Conference, pp. 147–152, 1988.

[35] B. Schauerte, and R. Stiefelhagen, “How the Distribution of Salient Objects in Images Influences Salient Object detection,” In Proceedings of the 20th International Conference on Image Proceesing (ICIP), The Technical Writer’s Handbook. Mill Valley, CA: University Science, 1989.