Development of a Machine Learning System for Safety of Operators in Agricultural Farm Environments Using Machine Vision

著者（英）	Yan ZHANG
内容記述	この博士論文は内容の要約のみの公開（または一部非公開）になっています
year	2019
その他のタイトル	マシンビジョンを用いた農業圃場における操縦者安全のための機械学習システムの開発
学位授与大学	筑波大学 (University of Tsukuba)
学位授与年度	2018
報告番号	12102甲第9065号
URL	http://hdl.handle.net/2241/00156386

Development of a Machine Learning System for Safety of Operators in Agricultural Farm Environments Using Machine Vision

January 2019

Yan ZHANG

Development of a Machine Learning System for Safety of Operators in Agricultural Farm Environments Using Machine Vision

A Dissertation Submitted to
the Graduate School of Life and Environmental Sciences,
the University of Tsukuba
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy in Bioresource Engineering
(Doctoral Program in Appropriate Technology and Sciences for Sustainable Development)

Yan ZHANG

Abstract

The most important points to ensure the safety of agricultural machinery operators are correctly understand the driver action or behavior and external driving environments. Worldwide aging farmer increases which causes higher risk in driving in rural road structures. A safety system is required with adopting in common datasets for farmer's inattention while driving and surrounding environments including road conditions, obstacles, and free spaces in rural environments. Therefore, the objective of this research is to develop a safety system through recognition of driving environments in rural road structures and driver's inattention or behavior while driving agricultural vehicles. On the other hand, a machine learning approach is required to deal with the large datasets for driving environment and driver's inattention behavior.

To achieve the objective of this research, vision-based sensors has the potential as a non-contact sensor to the driver. In this regard, first, establish an internal and external monitoring system, sensor layout included Kinect sensor, thermal camera and monocular camera for developing a vision-based rescue system. The Kinect sensor and the thermal camera were used to collect the sequence of images for the operator's action of driving; the monocular camera was used for road images to realize the driving environments.

Second, the machine learning was introduced for the analysis of images and videos that incorporated the sub space method. The mutual subspace method was extended for nonlinear analysis using a kernel called KMSM (KMSM: Kernel Mutual Subspace Method). The developed KMSM-based machine learning system was used to establish a

large datasets system for driving environment and driver action monitoring. At the same time, KMSM-based data processing system ensured low computational complexity and high accuracy recognition algorithm to satisfy real-time application and emergency response system for rescue of agricultural machinery operator.

Third, KMSM was implemented to classify road type and developed a Hankel-based KMSM to recognize driver inattention actions while driving a tractor. To reduce the computational complexity for fulfilling the requirements of real-time recognition, high resolution of raw images were resized to low-dimensional images. The resized low-dimensional vectors were used to generate low-dimensional block Hankel matrices as representations for input road images and action sequences. A sliding window was designed both for road type and action classifiers. By using the sliding window, stability and efficiency was improved through generation of sub sequences.

Fourth, the performance of the proposed KMSM and Hankel-based KMSM was evaluated. A road type dataset was established included three categories: straight road, curve road, and cross road under structured road and unstructured road classifiers, respectively. Typical roads included agricultural roads with grasses, without grasses, publicly available agricultural roads in Japan for the establishment of road type dataset. A driver action dataset was established that included 10 subjects and 5 classifiers of inattention actions for a tractor, respectively. Inattention actions with high possibility lead to an accident was discussed and selected for the driver inattention action dataset, which included: talking on the phone, look aside, rubbing eye, nodding and yawning. These inattention actions were collected and discussed in this research. The driver inattention actions were categorized into three danger levels, and the corresponding countermeasures for each danger level's actions were similarly classified. Referring to the driver danger level, and MRM (MRM: Minimal Risk Maneuver) was defined as a system response under different road conditions. According to driver action as danger level and road condition, countermeasures and MRM included warning the driver for speed down, drive to roadside or stop, or proceed to parking. If the action level were recognized as higher danger level such as fatal accident or sudden comma or unable to drive, the rescue system could able to contact to the emergency center using android-based mobile communication.

Fifth, designed off-line and on-line experiments using KMSM and Hankel based KMSM algorithms was used to evaluate recognition performance of classifiers. For road condition classification, similar and different road driving conditions were trained and tested. For driver action recognition, similar subjects (volunteer drivers) and different subjects (volunteer drivers) were conducted to evaluate the inattention action recognition performance. For road classification using KMSM, the off-line classification accuracy rates were 97.7%, 98.1% and 95.4% for curve road, straight road and cross road under structured road and unstructured road, respectively. The on-line classification accuracy rates were 100%, 85.5% and 91.55%, curve road, straight road and cross road under structured road and unstructured road classifiers, respectively. The average computation time was 0.03s, proved the system with low calculation time to be implemented as a real-time system. The developed Hankel-based KMSM was used in the off-line recognition system for driver inattention action classification. The average recognition accuracy of the classifier were 91.18% and 86.18% for similar subjects training and testing and different subjects testing, respectively for Kinect sensor using RGB images. On the other

92.2% and 47.52% for similar subjects training and testing and different subjects testing, respectively, for thermal camera using thermal image. The on-line classification accuracy rates were 87.02% and 79.97% for similar subjects training and testing and different subjects testing, respectively in case of RGB image; and with accuracy of 83.24% and 42.75% for similar subjects testing and different subjects testing, respectively in case of thermal image. The average computation time was 0.07 s and 0.08 for RGB image and thermal image, respectively. Which proved the system with low calculation time can be implemented in the real-time application for a rescue system in presently uses agricultural machinery.

In conclusion, the core of this research work contributed to developing the architecture of the driving rescue system, driver action inattention behavior datasets that included driver danger status monitoring with inattention actions and fatigue detection; road type classifies and minimal risk maneuvers strategy. As a conclusion, the proposed method of the KMSM and Hankel-based KMSM could satisfy the real-time application with higher accuracy requirements and minimal type of recognition for classifiers of rural road type and driver inattention action behavior to develop a driver rescue system for agricultural vehicles.

Table of Contents

Nomenclature

MSM: Mutual subspace method

KMSM: Kernel mutual subspace method

PCA: Principal component analysis

KPCA: Kernel principal component analysis

d : Dimensionality of the subspace used for recognition

P : Reference subspace that were registered as the dictionary

Q : Input subspace that were obtained from the input data

u_i : Eigenvectors in the reference subspace P

v_i : Eigenvectors in the input subspace Q

R^n : N-dimensional feature space

x_i : Pattern vector

C : Covariance matrix from k feature vectors $\vec{x}_a$

λ : Eigenvalues of the eigenvectors

X : N-dimensional input patterns

F : High- or infinite-dimensional feature space

$\emptyset$: Nonlinear map

α_i : Enumeration of coupling coefficients

θ : Similarities between the reference subspace P and input subspace Q

Kij : Gaussian Kernel function

σ : A free parameter of Gaussian Kernel

I : Sequential images of an action

f_i : Image vector

H_I : Block Hankel Matrix for sequential images I

hj : Column of the block Hankel Matrix H_I

r : Hankel block parameter for Hankel matrix

S : Set of the sub-actions that were generated with sliding window

s_i : Sub-action that includes r action sequence images in a sliding window

G: Dimensional of vector, where $G = d \times r$

M: Number of action classes

N: Number of action sequence

J: Number of testing sequence

NSW: Width of sliding window

NAR: Recognized times of inattention action

TRB: Training data sets

TEB: Testing data sets

List of Tables

List of Figures

Chapter 1
Introduction

With the development of autonomous driving technology, agricultural vehicles for bio-production systems have obtained significant intelligence improvements. Technology of smart agricultural machinery is creating a chance to changes the traditional agricultural and transform to an intelligent agricultural operation method. Autonomous agricultural production could ensure precise operation, increase productivity, saving workforce, and improving production. In these systems, autonomous applications have recently gained high adoption potential through the improvements to sensors, positioning and navigation performance. And now, autonomous tractors with advanced sensing technologies, control theories, have been developed.

However, most of the previous research focused on the navigation of tractors and autonomous agricultural task operation. Under current conditions, there is in fact a strong need to ensure driving safety in agricultural operation. In these regards, the datasets are not also available to speed up the autonomous agricultural operation and farmers safety perspectives. From a safety standpoint, many barriers to the implementation of autonomous driving systems and agricultural vehicle operation exist because in contrast to passenger car control in agriculture, autonomous driving must not only adapt to uneven farmland roads and terrain but also conform to the requirements of a variety of tasks. Most of the researches have been contributed to passenger car control systems. Agricultural driving monitoring especially rural roads and terrain environmental detests are very much important. Once the datasets are established artificial intelligence can be extended for the autonomous operations.

1.1 Rescue System

To ensure safety during driving, some of the researches have been performed, and many types of machine learning algorithm and sensor systems have been applied to agricultural navigation. The advancements may possibly lead to the increased safety of agricultural vehicles for autonomous applications. Researches worked for driving safety and efforts can be divided into three categories or approaches: first, safety and correctly driving strategy making based on external environment sensing; second, vehicle control based on its posture detection relay on internal sensors; and third, monitoring a driver's condition to ensure safety by alarming or trigger MRM (minimal risk maneuver). In researches of driving safety, the position of machine learning is becoming more and more important. As the machine leaning is possible to provide solutions to handling complex models and conditions, with advantage to establish a robust and flexible system then traditional model approach.

Systems that sense the external environment and detect driver posture inarguably improve the safety of autonomous driving and advance driver assistance; monitoring a driver’s condition is both effective and important in ensuring driving safety. Statistics of accidents caused by driver fatigue and distraction have been reported by many countries. One important factor in these accidents, especially on rural roads, is inattention caused by driver fatigue or monotony. Moreover, with the aging population and the rapid reduction in the agricultural labor force, the average age of farmers worldwide has risen to approximately 60. One study reported that by 2020, more than 60% of the people engaged in farming in Japan will be older than 65, and 45% will be older than 70 (Zhang et.al, 2018). The poorer physical condition of older farmers causes them to be easily distracted and fatigued, which poses higher health risks while driving. Another challenge for aged farmers to drive a tractor under uneven farmland conditions, for their slow response and incorrectly understanding to the surrounding environment and road condition. Thus, a rescue system that monitor environmental and a driver's condition while driving agricultural vehicles in farmland is essential to ensure driving safety.

1.2 Environmental Monitoring

Environment of an agricultural vehicle that is possible to drive and conduct tasks in most of the features of irregularity farmland ground, varied topography, and terrible road condition. Agricultural vehicles have to confront conditions such as slippery land cause accumulated odometry error; dust and flog would lead optic sensors detection noise; and strong sunshine would even lead to blind of cameras. For driving a vehicle, the most important surround environment information is driving able area, means roads and free space. Detection of road, boundary, obstacles or holes on the road to support the autonomous driving focused by some of the researches. However, it is difficult to support correct detection under different land and road condition using one kind of sensor or algorithm. Approach that robust to environment variousness is required to adapt different ground conditions. Machine vision and machine learning algorithm by dealing with raw images is possible to meet for the above requirement. In this research, the machine learning approach-based machine vision was designed to recognize agricultural environment. The machine vision has the potential to provide rich information and also low-cost involvement in automation. There are several machine learning approaches and this becomes increasing important due to demand of artificial intelligence in agriculture. Deep learning, Neural Network has been implemented in the machine system. However, some of the machine learning system is hard to explore as the computation complexity and black box issues. In this regard, the machine learning algorithm of Mutual Subspace Method (MSM) could be possible solution to interpret and developed accordingly to reduce environment noises and meet with requirement of environment variance.

1.3 Driver's Action Monitoring

Monitoring drivers' conditions by measuring physiological characteristics such as their brain waves, heart rate, skin conductivity and pulse rate yields the maximum detection accuracy. However, there are disadvantages when using physiological sensors that require physical contact with drivers that include causing annoyance, noise from the electrical signals obtained from such sensors, and lack of a realistic environmental perception. Less intrusive techniques, including computer vision techniques such as eye- or gaze-tracking have provided good results in judging whether the driver's view deviated from the driving direction. At the same time, such techniques provide more information concerning the driver's face, head and body movements that can be even more meaningful than eye motion. Driver video images could be used to monitor driver behaviors such as eyelid movements, eye closure percentage, nodding, yawning, gazing, sluggish facial expressions, and sagging posture. However, the existing solutions lack sufficient accuracy and speed to function reliably under real vehicle conditions. Sequential images of human actions contain rich information on body postures and motion. However, the images are easily affected by different viewpoints, illumination conditions and individual characteristics. Although many approaches have been proposed to recognize human actions, how to effectively handle the rich body of gesture information and capture the temporal information from human action sequences is the most important problem need to conquer. It is possible to address the former by modelling the distribution of the image patterns using the machine learning approach of nonlinear subspace. Very recently, the Kinect device and thermal camera has been identified to have higher the potential compare to stereovision images due to depth of RGB images and infrared images. Thermal image has the advantages in uses of monitoring during low illumination condition. The machine-learning algorithm of Kernel Mutual Subspace Method (KMSM) has the advantage to modeling high nonlinear human actions.

1.4 Objectives

Therefore, this study intent to address the issues of driving safety in two ways: One approach is by monitoring driving environmental of road condition and road type to identify structured road and unstructured road, straight roads, curve roads and cross roads. Another approach is by monitoring driver condition and action to identify inattention action and fatigue status. Based on the driving environment and driver condition recognition results, a minimal risk policy is designed to ensure driving safety. Machine vision system and machine learning algorithm with accuracy and stable recognition and low computational is designed to complete the rescue system from road driving and field operation.

Therefore, the objectives of this thesis are listed according to reach the goal:

1. To develop a rescue system for farmland and uneven road driving, with road type monitor, driver condition and action monitoring, minimal risk policy, communication network, and emergency center.

2. To develop dataset of road types with different environment, dataset of driver action through including general dangerous behaviors or status as health risk.
3. To development high accuracy and low computational complexity road types recognize system and driver action/status recognize system which is fulfill real-time recognize.
4. To development a minimal risk policy considering road condition and driver status, to realize minimal risk and ensure safety.

1.5 Organization of this Thesis

The research works in this thesis describes the progress of using machine vision for the road type monitoring and driver action and status monitoring in the outdoor environment. In this chapter, discussed the nature of the problems: rescue ensure for agricultural vehicle driving in section 1.1; the scope of this research and highlights the application of environment and driver monitoring in approaching is discussed in the section 1.2 and 1.3; section 1.4 covers the objectives of this research; section 1.5 presents the structure of the thesis; The rest of the thesis is composed as follows:

Chapter 2 covers the major work contributed in research of driving rescue, machine learning approaches for action recognition and environment recognition. The uses of machine learning algorithm and sensors to develop the rescue systems are reviewed in this chapter.

Chapter 3 introduces the subspace method of MSM, KMSM, Block Hankel Matrix and the designed Hankel based KMSM. Where the KMSM is used for recognition of road type and the Hankel based KMSM is used for recognition of driver action and status.

Chapter 4 discusses the structure of hardware system used in the research. The tractor and vehicle used data collection, the vision sensors selected for providing images is introduced, which are used to contribute the rescue system. Also, installation of sensors in the experimental tractor/vehicle and the sensors specifications are described.

Chapter 5 describes the analysis of agricultural environments especially the agricultural road conditions. Based on the result, establishment of dataset for road types cover major road feature under different environments are discussed. And the implement of KMSM is introduced, which included the training and recognize processes. The recognized result with developed KMSM algorithm is addressed.

Chapter 6 describes the analyzation of driver actions and status potentially lead to accidents. Establishment of datasets included the defined general danger driver actions and implement of designed Hankel based KMSM algorithm is introduced, in which the Kinect sensor and thermal camera are used. Recognized results of driver action recognize system by utilizing two different vision sensors are evaluated and compared, applied the developed Hankel based KMSM algorithm.

Chapter 7 concludes the thesis with the summary of contributions to achieve the research target and discussed for future research.

Chapter 2
Literature Reviews

With the development and improvement of traditional method and appear of innovated approach, performance of recognize complex application such as action recognition was greatly improved. And the improved sensors device promoted machine learning technology to understand human action and surround environment. However, compare with indoor application, driver action recognize could be affected by illumination condition, vehicle vibration and driver posture change caused by uneven ground. Also, different with the autonomous vehicles working indoor or running on highway public roads, the autonomous agricultural machines request higher performance of sensing system and robust controller. As a rescue system to recognize agricultural vehicle driver ab-normal actions and land road condition, the system has to conquest worse condition under agricultural environment at first and further ensure accuracy recognize. In this chapter, the follower sections discuss the contributions regarding agricultural vehicle rescue system.

2.1 Machine Learning for Action Recognition and Environmental Recognition

Machine learning is an important study area of computer science, which target to establish model or algorithm to learn from data. Generally, the machine learning algorithms can be divided to two main types: supervised learning and unsupervised learning. The supervised learning algorithms aim to minimize some error criterion based on the difference between the targets and the outputs. On the contrary, unsupervised learning algorithms exploit similarities between inputs to cluster those inputs that are similar together. Typical approach for action recognizes and environment machine learning algorithm such as Hidden Markov Models, Linear Dynamic Systems, Support Vector Machine, Principal Component Analysis / Mutual Subspace Method and Deep Learning is wildly researched in the world.

2.1.1 Hidden Markov Models

Hidden Markov Models (HMM) are popular approach to model and describe state transitions between actions. Since the HMM can easily model the temporal evolution of a single or a set of numeric features extracted from the data. An HMM include a number of states and each of which is indicate a probability of transition from one state to another state. And state transitions occur stochastically with time. Same as Markov models that states at any time depend on the state at current time. Symbol yielded from each of the HMM states according to the probabilities assigned to the states. HMM can be observed only through a sequence of observed symbols, as states are not directly observable. The selection of the feature set and the related emission probability function are the key issues to be defined. In particular, if the training set is not sufficiently large, a manual or automatic feature selection and reduction is mandatory. The main challenge to design a method that the developed system is allow to cope with different types of action, even the actions are very similar to each other and also in the case of cluttered and complex scenarios. Occlusions, shadows and noise are the main problems to be faced. In video surveillance applications the actions should usually be recognized by means of an image stream coming from a single camera. (Yamato et.al, 1992) proposes a human action recognition method based on the HMM. A feature-based bottom up approach with HMM is characterized by its learning capability and time-scale invariability. One set of time-sequential images is transformed into an image feature vector sequence, and the sequence is converted into a symbol sequence by vector quantization. In learning human action categories, the parameters of the HMMs, one per category, are optimized to best describe the training sequenced from the category. To recognize an observed sequence, the HMM which best matches the sequence is chosen. Experimental results of real time-sequential images of sports scenes showed recognition rates higher than 90%. Lee et.al: researched for mapping the driver's sight line and the driving lane path using two cameras by capture the driver's image and the front-road image. Two correlation coefficients among the driver's sight line, the driving lane path and the car heading direction are calculated in the global coordinate to monitor the driving status such as a safe driving status, a risky driving status and a dangerous driving status. By implement the HMM, four driving patterns including the driving in a straight lane, the driving in a curve lane, the driving of changing lanes, and the driving of making a turn was successes for recognition.

2.1.2 Linear Dynamic Systems

Linear Dynamic Systems is a powerful tool to represent temporally data. LDSs were wildly utilized in engineering, controls, economics and physical and social sciences. The need for comparing LDSs also arising in computer vision, for example to effectively describe or model videos of human motion or activity. As compared with traditional methods, more dynamic information in action exhibits and suitable for being analyzed by LDSs. By fitting an LDS to a sequence image, different actions conducted by same people or same action conduct by different people can be identified. The distance between LDSs of modeled action is useful for recognizing an action. Learning a dynamical system from data involves finding the parameters and the distribution to maximize likelihood of observed data. The maximum likelihood solution for these parameters can be found through iterative techniques such as expectation maximization (EM), gradient descent and least squares on a state sequence. A hierarchy of dynamic medial axis structures at several spatial-temporal scales was modeled using a set of Linear Dynamical Systems (LDSs) (Chaudhry et.al, 2013). Then propose novel discriminative metrics for comparing these sets of LDSs for the task of human activity recognition. Combined with simple classification frameworks, the proposed features and corresponding hierarchical dynamical models provide high human activity recognition rates by testing on several skeletal datasets. Wang et.al: Using LDSs to describe the dynamic texture which exhibits certain stationarity properties in time. They are adopted to model the spatiotemporal patches which are extracted from the video sequence, as the reason that, rather than the video sequence the spatiotemporal patch is more analogous to a linear time invariant system. By adopt a kernel principal angle to measure the similarity between LDSs, and using a multiclass spectral clustering to generate the codebook for the bag of features representation. And the method was proved by testing UCF Sports and Feature Films datasets in realistic complex scenarios.

2.1.3 Support Vector Machine

The Support Vector Machine algorithm (SVM) is also a popular classifier and wildly used for computer vision tasks as its high accuracy and efficiency. The SVM method has vastly demonstrated its performance in two-class discriminant classification. It is used to find the straight line that leads to the optimal separation (maximum margin) between the classes or clusters in the input space. A considerable advantage of SVM over the traditional neural networks is its better generalization performance even with a small dataset. (Li et.al, 2013) trained an SVM classifier and using eighty FFT and wavelet-based features that are extracted from 1-min HRV signals from four subjects. The averaged leave-one-out (LOO) classification obtained higer performance using wavelet-based feature better than the FFT-based results.

A simple Linear Iterative Clustering based Support Vector Machine (SLIC-SVM) is proposed to solve the problem that the SVM can only output a single terrain label and fails to identify the mixed terrain (Zhu et.al, 2018). The Simple Linear Iterative Clustering based SegNet (SLIC-SegNet) single-input multi-output terrain classification model is derived to improve the applicability of the terrain classifier. Series of experiments on regular terrain, irregular terrain and mixed terrain were conducted and both super pixel segmentation based synthetic classification methods can supply reliable mixed terrain classification result with clear boundary information.

2.1.4 Principal Component Analysis / Mutual Subspace Method

Principal component analysis (PCA) is used to reduce redundancy and obtain a compact eigen joints for each frame. PCA is one of the most successful approaches to the problem of creating a low dimensional data representation and interpretation, it got good result for face recognize. Multi-dimensional data is projected onto the singular vectors corresponding to a few of its largest singular values. It using an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables. The main components are obtained in decreasing order of importance and a number of them collect most of the information. The first eigenvector explains the main part of the total variance. Human action is represented with temporal displacement of joints, relative joint positions and joints offset concerns with the initial frame. The PCA is applied to the classification of human action. The Mutual Subspace Method (MSM) is methods for object recognition based on the PCA and eigenvectors of subspace obtained by PCA on the entered images. (Fukui et.al, 2003) Utilized the MSM for face recognition, using multiple face patterns obtained in various views for robot vision. As a face pattern may change dramatically due to changes in the relation between the positions of a robot, a subject and light source. Constrained mutual subspace method (CMSM) using multi viewpoint face patterns was designed to overcome problem that a single face pattern is not capable of dealing with the variations of face pattern. And effectiveness of the CMSM method for robot vision is demonstrated, face recognition was robust against variations caused by the changes. (Eduardo et.al, 2015) Designed a PCA for path recognition and obstacle detection for an autonomous mobile vehicle. The Path Extraction Algorithm (PEA) recognizes drivable paths on the road by image processing. The Environment Extraction Algorithm (EEA) provides the special pose of the mobile vehicle and obstacle detection by the data processing of the 2D laser scanner. The PCA was used to classify road patterns by the use of trained Artificial Neural Networks. The Navigation-Path Extraction Algorithm (NPEA) is comprised of these three sub-systems. Reliable and robust for Navigation-Path Extraction System (NPES) is evaluated, and the system was thought with possibility to be implementable on a mobile vehicle to achieve self-driving.

2.1.5 Deep Learning

Deep learning is widely used to resolve difficult problems which cannot be handled properly using conventional methods. One of the most popular and successful architectures in research of deep learning is convolutional neural networks (CNNs),

CNNs showed a great potential especially for tasks such as image classification, object detection, emotion recognition, scene segmentation tasks. CNNs composed layers as convolutional layers, pooling layers, a dropout layer, a flatten layer, and a set of fully-connected neuron layers. Multiple layers possible to highlight different features or characteristics of the input and CNN models typically have millions of parameters. As a result, large amounts of training data are needed and the collection, pre-processing, and annotation of datasets require a lot of human and computational resource. (Zhang et.al, 2010) infrared videos and CNN network are utilized for detecting and propose an eye state recognition. The designed system has high recognition accuracy of state of eyes when wearing glasses and can detect the fatigue effectively. (Lundahl et.al, 2014) propose an approach to detect drivable road area in monocular images. The system automatically generates training road annotations for images using OpenStreetMap through vehicle pose estimation sensors, and camera parameters. CNN was trained for road detection using these annotations. Reasonably accurate training annotations was showed in KITTI data-set.

2.2 Sensors

2.2.1 Biomedical Sensor

Physiological signal detected by biomedical sensors can provide most direct information to describe driver physiological condition. Potential of physiological signals such as electromyography (EMG), electrocardiogram (ECG), electroencephalogram (EEG), electro-oculogram (EOG), and respiration signals (RESP) are wildly discussed. Physiological features such as complexity, approximate entropy (ApEn), sample entropy (SampEn), heart rate variability (HRV), medium frequency (MF), integrated electromyography (IEMG) are researched to identify the river fatigue condition. (Masala et.al, 2014) Proposed to using single-channel EEG device (TGAM-based chip) to monitor changes in mental state (from alertness to fatigue). Fifteen drivers performed a 2-h simulated driving task while we recorded, simultaneously, their prefrontal brain activity and saccadic velocity. Saccadic velocity was used as the reference index of fatigue. Subjective ratings of alertness and fatigue was also collected during driving. Results suggest that the TGAM-based chip EEG device is able to detect changes in mental state during driving. Roman et.al: Analysis the physiological signals (EEG, ECG EOG) of drowsy and alert drivers described to identify the fatigue level of the drivers. After extracted signal from EEG spectrum, blinking frequency and fractal properties were utilized to recognize the driver alert/drowsy states.

2.2.2 Vision Sensors

Vision based intelligent system has been deeply researched for scene analysis. The camera information provides un-contact solution for driver condition recognize, and for drive environment understanding. The vision-based driver status monitor and sensing system for ADAS solution has become a standard equipment for vehicles. (Cyganek et.al, 2014) Present a hybrid visual system for monitoring driver's states of fatigue, sleepiness and inattention based on driver's eye recognition. Two cameras operating in the visible and near infra-red spectra are used in car conditions and processing in daily and night conditions. Two classifiers are used, one for detection of eye regions based on the proposed eye models specific to each spectrum and the another for eye verification. High recognition accuracy and real-time processing performance are verified the potential of the system become a part of the advanced driver’s assisting system. A long-range vision-based sensing system for mobile robotics was designed using stereo camera (Raia et.al, 2009). Present a self-supervised learning process for long-range vision that is able to accurately classify complex terrain at distances up to the horizon, thus allowing superior strategic planning. A deep hierarchical network is trained to extract informative and meaningful features from an input image, and the features are used to train a real-time classifier to predict travers-ability. The trained classifier possible to identify obstacles and paths from 5 to more than 100 m, far beyond the maximum stereo range of 12 m, and adapts very quickly to new environments.

2.2.3 Laser Ranger Finder/Lidar

Laser Range Finder (LRF) and lidar owns the advantage for distance and angle measurement. Using the laser information, intelligent system as autonomous vehicle possible to obtain 3D surround information and with less effect by illumination condition. As the LRF with long range and high resolution many researches have been introduced and research both for indoor and outdoor for localization and navigation. HAN et.al: A road boundary detection and tracking algorithm was developed using lidar information both for structured and unstructured roads. In the research, the road features were extracted as line segments in polar coordinates relative to the lidar sensor. And a nearest neighbor filter was applied to a vehicle's local coordinates by tracking the extracted road features. The lidar-based system accurately detected the road boundaries regardless of the road type. (Giulio et.al, 2016) Proposed a LIDAR-stereo combination to detect traversable ground in outdoor applications. Two self-learning classifiers of LIDAR-based classifier and stereo vision-based classifier are integrated in the proposed system. The applied two classifiers are intended to detect the broad class of drivable ground. Each single-sensor classifier includes two main phases: phase of training and phase of classification. At training phase, the classifier learns to associate geometric of 3D data with class labels automatically. And then, the algorithm makes predictions based on past observations. By statistically combining output of two single-sensor classifiers to exploit their individual strengths. Such that the system could reach an overall better performance than achieved by single classifier..

2.3 Concluding Remarks

The HMM method required modeling of human body for accuracy action classify, however it is impossible to describe the body using skeleton information during driving. The SVM is good for two classes classification, while sub motion during driving could be considered as noise to lead the system failure. Deep learning is powerful for action recognize and scene recognize, however a large number of training data is required. In this research, the vision-based monitor system is proposed using subspace method based on PCA, which is utilized to recognize driver fatigue condition and road condition according divided by the numbered of classifiers according to the danger of action.

Chapter 3
Development of Hankel based KMSM Algorithm

3.1 Mutual Subspace Method

Mutual Subspace Method (MSM) (Maeda & Watanabe, 1985) is one of the well-known methods for object recognition based on the image sets. In the MSM, a subspace that has d-dimensional vectors is selected according to a criterion, such as the cumulative contribution rate from the eigenvectors, which are obtained by a Principal Component Analysis (PCA) on the entered images, and next, it calculates the similarities θ of the eigenvectors P= $\{u_i\}$ that were registered as the dictionary and the eigenvectors Q= $\{v_i\}$ that were obtained from the input data (Fig. 3.1).

$$\cos\theta = \max_{u_i \epsilon Q} \max_{v_i \epsilon P} u^T v \qquad 3.1$$

where $u_i^T u_i = v_i^T v_i = 1, u_i^T u_j = v_i^T v_j = 0, i \neq j$ .0< $i, j \leq d,$ and d is the dimensionality of the subspace used for recognition.

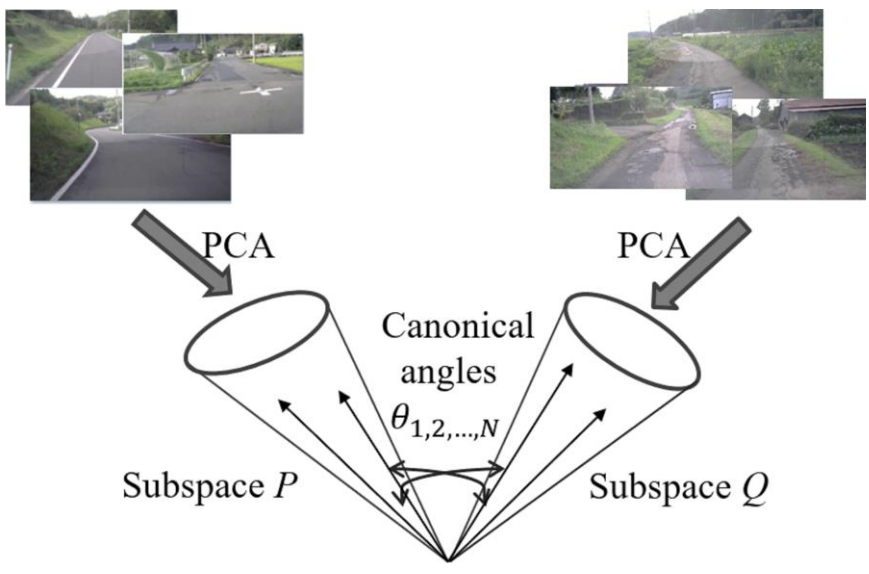


Fig. 3.1 Comparison between two sets of images using MSM

Normally, the process of calculating PCA is as follows [Sakano, Mukawa, & Nakamura, 2005]: First, calculated the covariance matrix C from k feature vectors $\vec{x}_a (a = 1, \dots, k)$ in an n-dimensional feature space R^n.

$$C = \frac{1}{k}\sum_{b=1}^{k} \vec{x}_b \vec{x}_b^T \quad 3.2$$

Then, a principal component analysis solves the characteristic equation to obtain the principal components $\vec{v_i}(i = 1, \dots, k)$ of the distribution.

$$\lambda\vec{v} = C\vec{v} \quad 3.3$$

However, it was assumed that all of the data here were calculated from the data centroid. This principal component describes the direction of the largest data variation under a linear approximation. The above characteristic equation can be transformed as follows:

$$\lambda\vec{x} = \left[\frac{1}{k}\sum_{b=1}^{k}(\overrightarrow{x_b} \cdot \vec{x}_b^T)\right]\vec{v} \quad 3.4$$

$$= \frac{1}{k}\sum_{b=1}^{k}(\vec{x}_b \vec{x}_b^T)\vec{v} = \frac{1}{k}\sum_{b=1}^{k}(\vec{x}_b \cdot \vec{v})\,\vec{x}_b \quad 3.5$$

Since $\vec{v}$ is in $\{\vec{x}_1, \dots, \vec{x}_k\}$, it can be obtained as follows:

$$\lambda(\vec{x}_m \cdot \vec{v}) = \vec{x}_m \cdot C\vec{v} \quad 3.6$$

Although MSM could achieve a high recognition rate for linear structures, the performance drops significantly when the pattern distributions were compared and had highly nonlinear structure. To address this problem, the MSM has been extended to KMSM (Sakano, Mukawa, and Nakamura, 2005; Wof & Shashua, 2003).

3.2 Kernel Mutual Subspace Method

Kernel Mutual Subspace Method (KMSM) is a nonlinear extension of the mutual subspace method (MSM) by using the kernel trick. In the KMSM, the PCA was applied on the high-dimensional space F to generate a nonlinear subspace that contained the nonlinear distribution of each set of image patterns.

Let $x \epsilon R^f$ be a vectorized image. In the KMSM, the n-dimensional input patterns $X = \{x_1, x_2, \ldots, x_n\}$ were mapped into a high or infinite-dimensional feature space F via a nonlinear map $\emptyset$, $F = \emptyset(X)$. To perform PCA on the mapped feature space F, the covariance matrix need be calculated. The covariance matrix in space F can be written as:

$$\bar{C} = \frac{1}{n}\sum_{j=1}^{n}(\emptyset(\vec{x_j})\emptyset(\vec{x_j})^T) \quad 3.7$$

Then the principal component in F denoted by $V \in F/\{0\}$,

$$\lambda\vec{V} = \bar{C}\vec{V} \quad 3.8$$

The expression of Eq. 8 be can described for R^f as:

$$\lambda(\emptyset(\vec{x_k}) \cdot V) = (\emptyset(\vec{x_k}) \cdot \bar{C}\vec{V}) \quad 3.9$$

If the principal component is a linear combination of the data, it is clear that for a certain coefficient α, the following is:

$$V = \sum_{i=1}^{n} \alpha_i \emptyset(\vec{x_i}) \quad 3.10$$

Combine Eq. 9 and Eq. 10, can obtain

$$\lambda \sum_{i=1}^{n} \alpha_i(\emptyset(\vec{x_k}) \cdot \emptyset(\vec{x_i})) = \frac{1}{n}\sum_{j=1}^{n} \alpha_i(\emptyset(\vec{x_k}) \cdot \sum_{j=1}^{n} \emptyset(\vec{x_j})\,(\emptyset(\vec{x_j}) \cdot \emptyset(\vec{x_i})) \quad 3.11$$

Because the dimension of the feature space F can be very high, so the inner products calculation is difficult. In thus, the nonlinear map $\emptyset$ is defined through a Kernel function:

$$K_{ij} = (\emptyset(\vec{x_i}) \cdot \emptyset(\vec{x_j})) \quad 3.12$$

Eq. 11 can be written as:

$$\lambda \sum_{i=1}^{n} \alpha_i K_{ji} = \frac{1}{n}\sum_{j=1}^{n} \alpha_i(\emptyset(\vec{x_k}) \cdot \sum_{j=1}^{n} \emptyset(\vec{x_j})\,(K_{ji})$$

$$= \frac{1}{n}\sum_{j=1}^{n}\sum_{j=1}^{n} \alpha_i K_{kj} K_{ji} \quad 3.13$$

Since the α_i here are enumerations of coupling coefficients, these can be treated as a vector, and Eq. 13 can be written as:

$$n\lambda\alpha = \alpha K \qquad 3.14$$

It is apparent that the "kernel trick" can solve the high dimensional inner products problem. Therefore, a commonly used kernel function is the Gaussian kernel, which is defined as follows:

$$K_{ij} = exp\left(-\frac{||x_i - x_j||^2}{\sigma^2}\right) \qquad 3.15$$

One common approach to human action recognition was to model the motion dynamics of the action using the Hidden Markov Model (HMM) (Yamato.et.al, 1992). However, the HMM required a large number of action instances samples and complicated parameter tuning to achieve high recognition performance. To approximate the dynamical system more effectively, a Hankel matrix was used to model the temporal evolution of the observed data over time (Moonen.et.al, 1994). A Hankel-based matrix was used for action recognition with a linear subspace (Li.et.al, 2011).

3.3 Block Hankel Matrix

The Hankel matrix contained the sequential images of an action that is described by I = $\{f_1, f_2, \dots, f_n\}$, where $I \in R^{d\times n}$,$f_i \in R^{d\times 1}$, n is the image number of the action sequence I, and d (d=width×heigh of image) is the dimension of each vector f_i, $i \in \{1,2,\dots,n\}$ (Fig. 3.2).

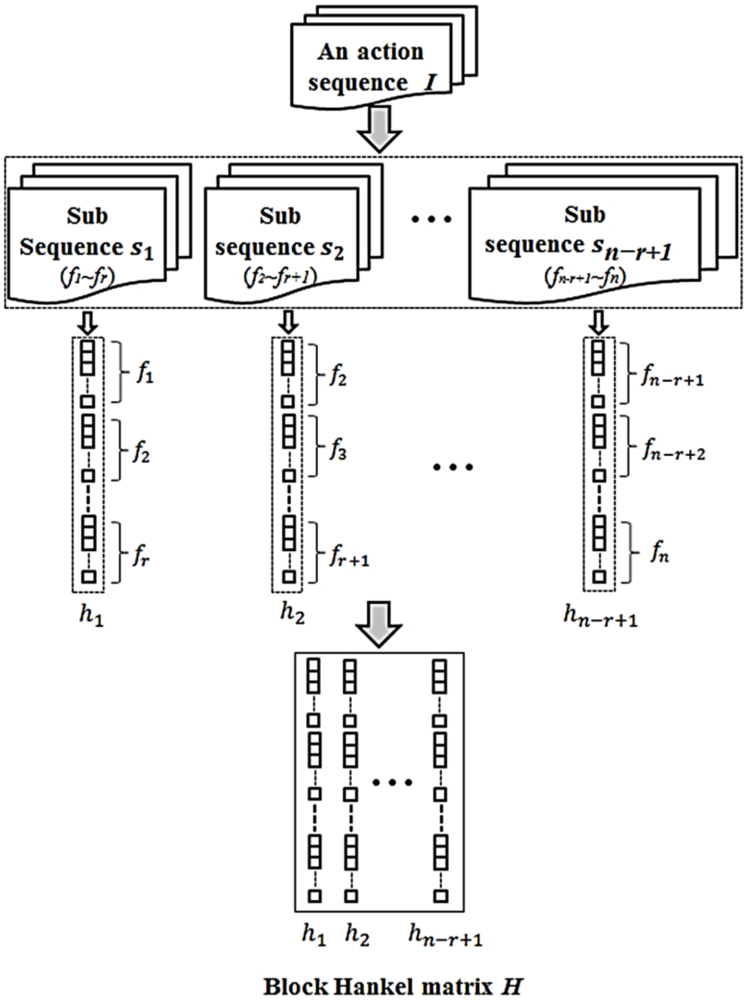


Fig. 3.2 Process of generating a block Hankel matrix from an action sequence

Corresponding sequential images in action sequence I, the block Hankel matrix H_I is defined as follows:

$$H_I = [h_1, h_2, \dots, h_{n-r+1}] = \begin{bmatrix} f_1 & f_2 & \cdots & f_{n-r+1} \\ f_2 & f_3 & \cdots & f_{n-r} \\ \cdots & \cdots & \cdots & \cdots \\ f_{r-1} & f_r & \cdots & f_{n-1} \\ f_r & f_{r+1} & \cdots & f_n \end{bmatrix} \tag{3.16}$$

where r is a Hankel block parameter for the Hankel matrix H_I, which determines the number of consecutive images that are stacked in each column vector of $h_j, j \in \{1,2,\dots,n-r+1\}$ in the Hankel matrix.

Each column h_j in H_I corresponds to each sub sequence $s_j \in S$:

$h_j = \{f_i, f_{i+1}, \cdots, f_{i+r}\}$ 3.17

Thus, the h_j would be a $G \times 1$ dimensional vector, where G= $d \times r$ and $H_I \in R^{G\times(n-r+1)}$.

3.4 Hankel-based Kernel Mutual Subspace Method

To overcome the limitations of the conventional subspace method in addressing temporal information, which is important for a driver's inattention action recognition, the Hankel-based KMSM was designed (Fig. 3.3). The procedure of comparing two action sequences based on the proposed method generate a block Hankel matrix from each action sequence, and by applying the Kernel PCA to the action patterns (sub action sequence) can generate a nonlinear subspace. The similarity between the two action sequences is defined based on the canonical angles between the two corresponding subspaces.

In the proposed method, a Hankel matrix was used to capture the temporal relations among the sequential images of an action. Note that each column vector $h_i \epsilon R^G$ in the generated Hankel matrix H contains r sub sequence image patterns, including the temporal relations. While the Hankel matrix represents the temporal information of an action, the subspace generated from the Hankel matrix can also capture sub-temporal relations of the action. The distribution of patterns includes the sub-action temporal information (Fig. 3.4).

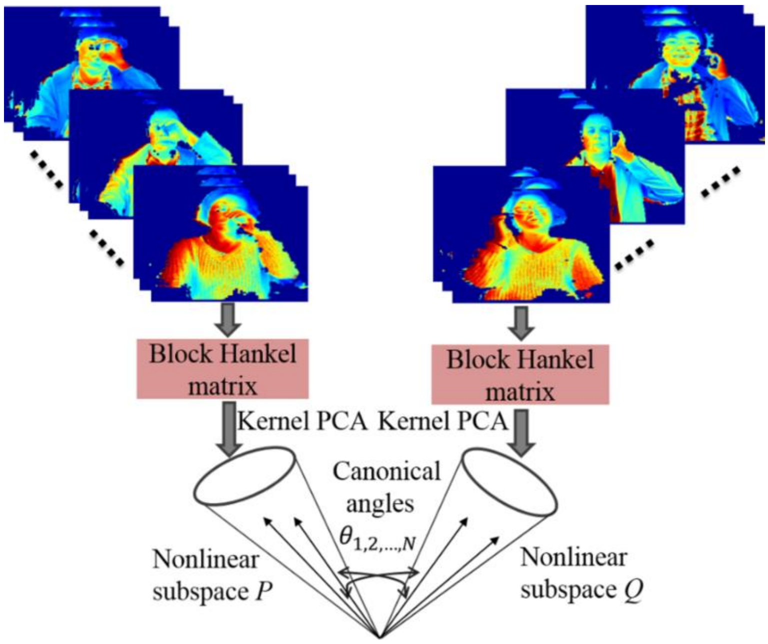


Fig. 3.3 Basic idea of the Hankel-based KMSM for action recognition

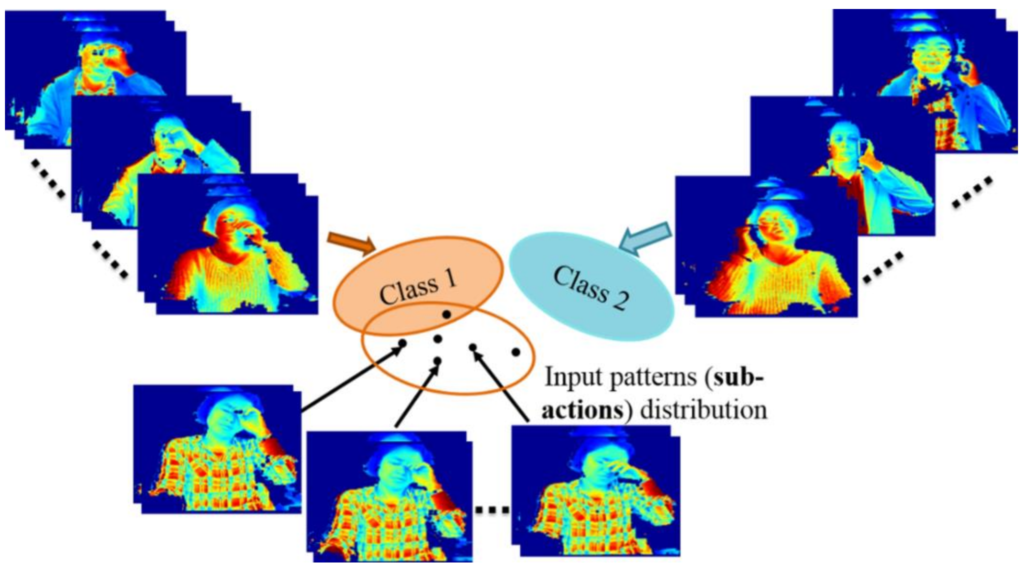


Fig. 3.4 Distributions of training patterns and input patterns for sub action

Chapter 4

Sensors Configuration and Data Acquisition from Action Recognition and Environmental Features

4.1 Sensor System Configuration

In this research, a Kinect and a thermal camera were used to recognize the driver action. The Kinect device and thermal camera was installed inside of the tractor and FOV of these two sensors were covered for driver area. An RGB camera was used to recognize road conditions and road types, which was designed to install on the tractor roof. The visible area of the RGB camera covered the road from the front of the tractor. The layout of sensor system was configurated as Fig 4.1.

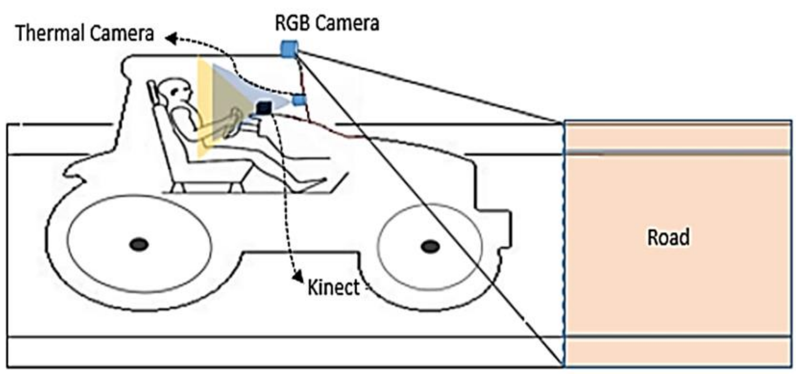


Fig. 4.1 Sensor system configuration

4.2 Basic Instrumentation

A YANMA tractor with cabin was used for the driver action recognition experiment. The Kinect device and thermal camera were installed back of the steering wheel, with direction of towards the driver (Fig 4.2). The tractor with cabin helped to install the device and also ensured the device safety. The cabin can also reduce the effect of sunshine to the Kinect and thermal camera.

For the road environment recognition experiment, a Honda life car (CBA-JB5 model, with high of 1500mm) was used for data collection. The small size of the car had the advantage to run both public roads and also narrow uneven farmland roads. An RGB camera was installed on the car roof on center line direct to the driving direction (Fig 4.3).

Fig. 4.2 The experimental tractor

Fig. 4.3 The experimental unit for running in the public road

4.3 Kinect Sensor

Compared with general RGB cameras, Kinect is a low-cost RGB-D sensor that can provide more information, which is composed of a colour camera, an infrared (IR) emitter and an IR depth sensor. The Kinect can capture and provide the RGB-image, depth image, skeleton joints and human body region (player image), and the driver area in the captured image can be automatically separated from the background and cropped. For those reasons, the Kinect was widely utilized in human detection and action recognition. In this research, the Kinect version1 was used (Fig 4.4), and it can detect object in rang of 0.5m~4.5m with FOV of 58.5×46.6 degree and 62×48.6 degree for IR camera and RGB camera, respectively. The major parameters are listed in Table 4.1.

Fig. 4.4 The Kinect device used in this experiment

Table 4.1 Major parameters of Version 1 Kinect device

Item	Parameters
Color camera resolution	640×480 30fps
IR camera resolution	320×240 pixels
Color camera FOV	62×48.6 degree
IR camera FOV	58.5×46.6 degree
Min/Max Distance	0.5~ 4.5m
Tracked joins	20

4.4 Thermal Camera

The FLIR VUE PRO R 336 thermal camera was used in this experiment to generate thermographers. The thermal camera provides high resolution and sensitive infrared image. In this research, the thermal camera was used to collect driver action and status, as compare with the Kinect RGB image recognition. Also, the thermal camera possible to provide a possibility discussion for night or low illumination driver status and environment recognition. The major parameter of the thermal camera was provided as Table 4.2.

Table 4.2 Major parameters of FLIR VUE PRO R 336 thermal camera

Item	Parameters
Lens size	13mm
FOV	25×19 degree
Resolution	336×256 30/60Hz
Pixel pitch	17um

Fig. 4.5 Thermal camera, VUE Pro FLR used in the experiment

4.5 RGB Camera

A Logitech Pro C920 web camera was equipped on the test vehicle to collect road images. It could provide 2 million effective pixels at a maximum 30 frame rate. The major parameter of Logitech Pro C920 web camera was provided as Table 4.3.

Table 4.3 Major parameters of Logitech Pro C920 camera

Item	Parameters
Lens size	13mm
FOV	78 deg
Resolution	1080/ 30fps and 720p 30fps
Focus type	Autofocus

Fig. 4.6 RGB camera,　Logitech Pro C920

4.6 Android System

For establish a communication network and alarm system, a ZTE android cellphone was used (Fig 4.6). By connect WIFI network, the android cellphone could connect with the tractor PC and emergency center with a server PC. After installation the developed app in the cellphone, the tractor PC could send an alarm message to cellphone and server PC.

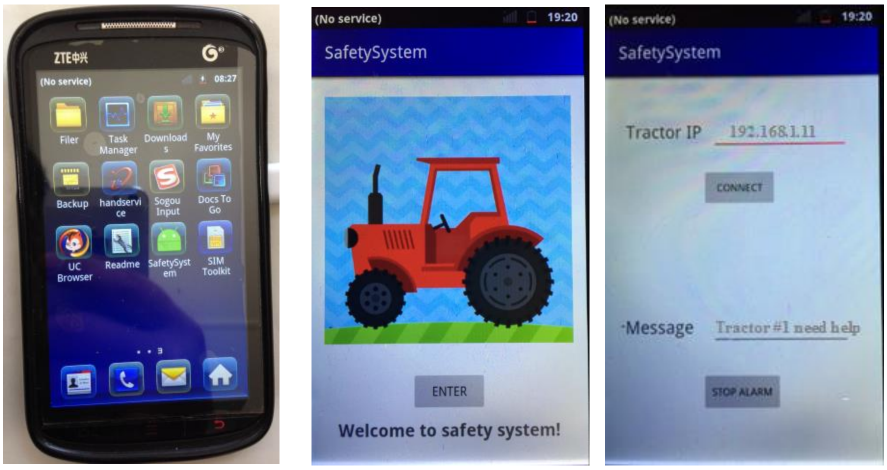


Fig. 4.7 The ZTE android cellphone system

4.7 Programming Modules

The recognition program was corded with MATLAB 2013 on a Windows platform and OpenCV library was used for image collection and image processing. The program includes image frame catch, image processing, driver action recognition process, road recognition process. The alarm system program was written by Java and Android Studio, included message sending and receiving through WIFI network and sound alarm.

4.8 Conclusions

The designed sensor layout of secure system composed by Kinect, thermal camera and RGB camera that could cover driver area and road area. The installation of Kinect and thermal camera was fully considered and adjusted without trouble normal driving. The installation of RGB camera considered the condition when install on a tractor, with similar height and direction.

Chapter 5

Agricultural Farm Environment Monitoring for Recognizing Rural Road Classification to Ensure Safety of Driving

Correctly understanding surround environment is the most directly approach to ensure driving safety. Under agricultural environment the driving environment is relative complex bad condition then public roads, such as road with grass, hole or variance obstacles, not clear boundary, and usually curve and narrow roads. Such complex condition made it difficult to understand the road condition, but on the other hand, it is essential to let the system know how to satisfy current environment. In this research, the secure system was composed by driver monitoring system and MRM (minimal risk manoeuvre) system. After health risk of tractor driver is detected by driver monitor system, the MRM system must take control of tractor to ensure safety. The MRM system includes two functions: understanding of environment and tractor action strategy making.

5.1 Agricultural Roads

By analysing the condition of agricultural road, the roads for running an agricultural vehicle has several possibilities to drive through main agricultural road, sub agricultural road and road among lands. From the view of safety, Action strategy under different environments are analysed as follow:

In the lands: Stop the tractor mostly in safety action, while driving the tractor to the land boundary. In this case emergency assistance can be added.

On unstructured roads: Road such as cultivation road (Fig 5.1 a), with bad road condition. It has a danger to control a tractor driving on unstructured road, the safest way is to stop the tractor on the road immediately.

On structured roads: Structured agricultural road includes general agricultural road (Fig 5.1 b), main agricultural road-1 (Fig 5.1 c), and main agricultural road-2 (Fig 5.1 d). Those three types of roads usually with heavier traffics, and the most safety action is control tractor stop on left road side.

Thus, in this research the roads were classified into three categories:

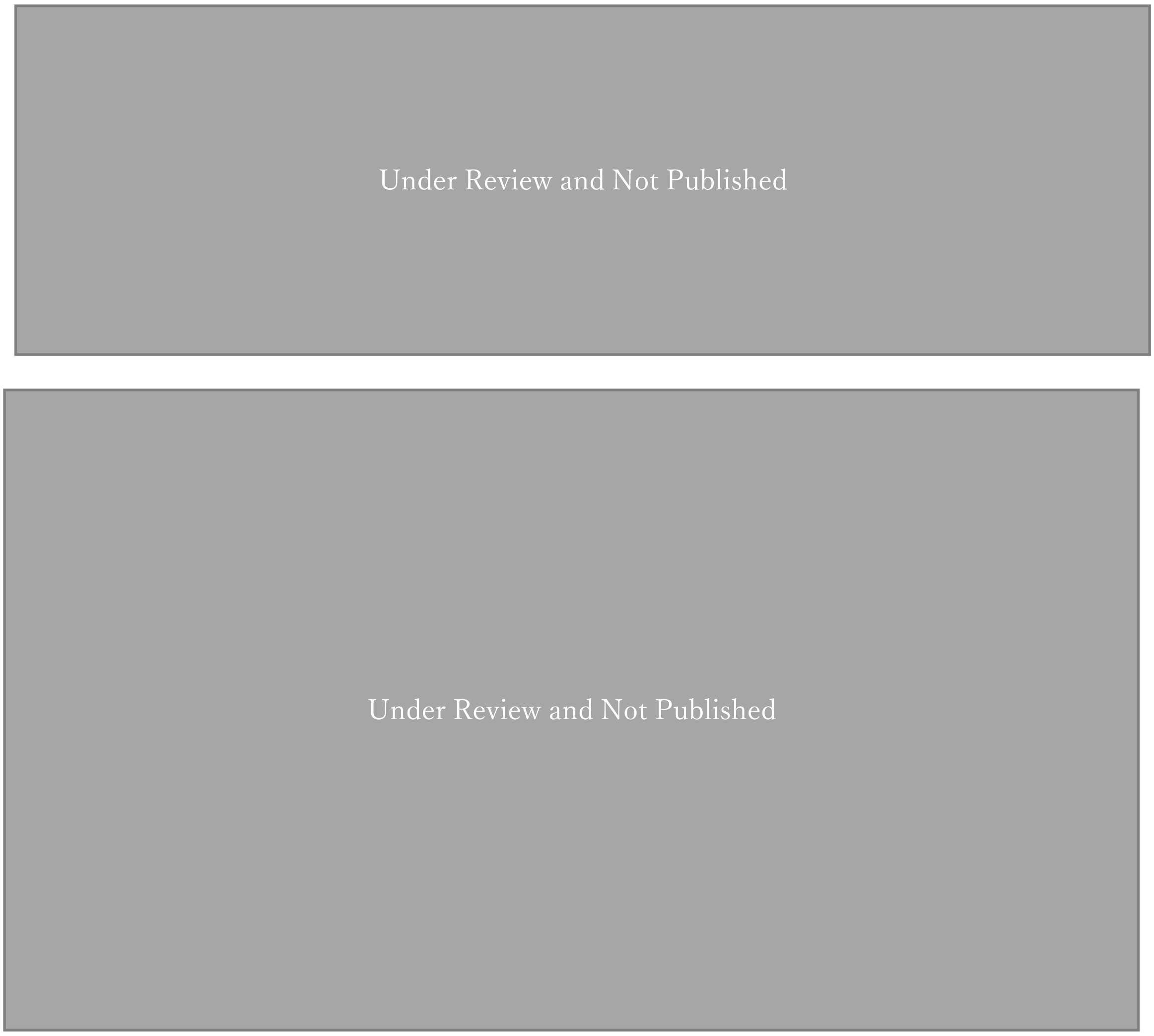


Fig. 5.1 Types of Agricultural Roads in Japan

5.2 Off-line Road Type Recognition System

The process flow of the KMSM had two components: the learning phase and the recognition phase (Fig. 5.2).

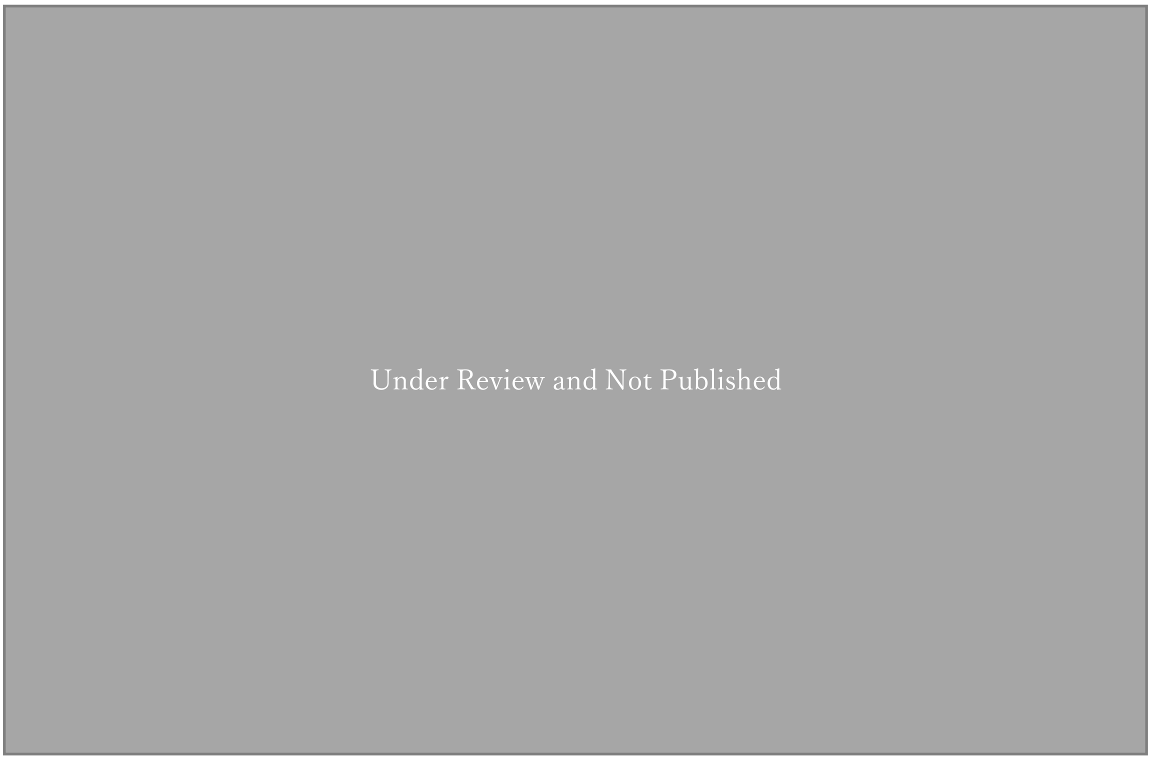


Fig. 5.2 Flow of the KMSM for Road Recognition

Algorithm 1 Learning phase:

Under Review and Not Published

Under Review and Not Published

Algorithm 2 Recognition phase:

Under Review and Not Published

5.3 On-line Road Type Recognition System

In the on-line road recognition process, the subspace patterns trained in the off-line road recognition process were utilized, and process of the on-line system included a recognition phase and an alarm phase. Considering the lower computation time that is required for a real-time system, we ensured the detection accuracy and the system stability;

(Fig. 5.3).

A user interface was designed, and the information display on the interface included the current cropped image, the testing image set, the predicted category, the recognition rate (the correct classifications were known during the test), the computational time, and the similarity plot (Fig. 5.4).

Under Review and Not Published

Fig. 5.3 Flow of the on-line operation

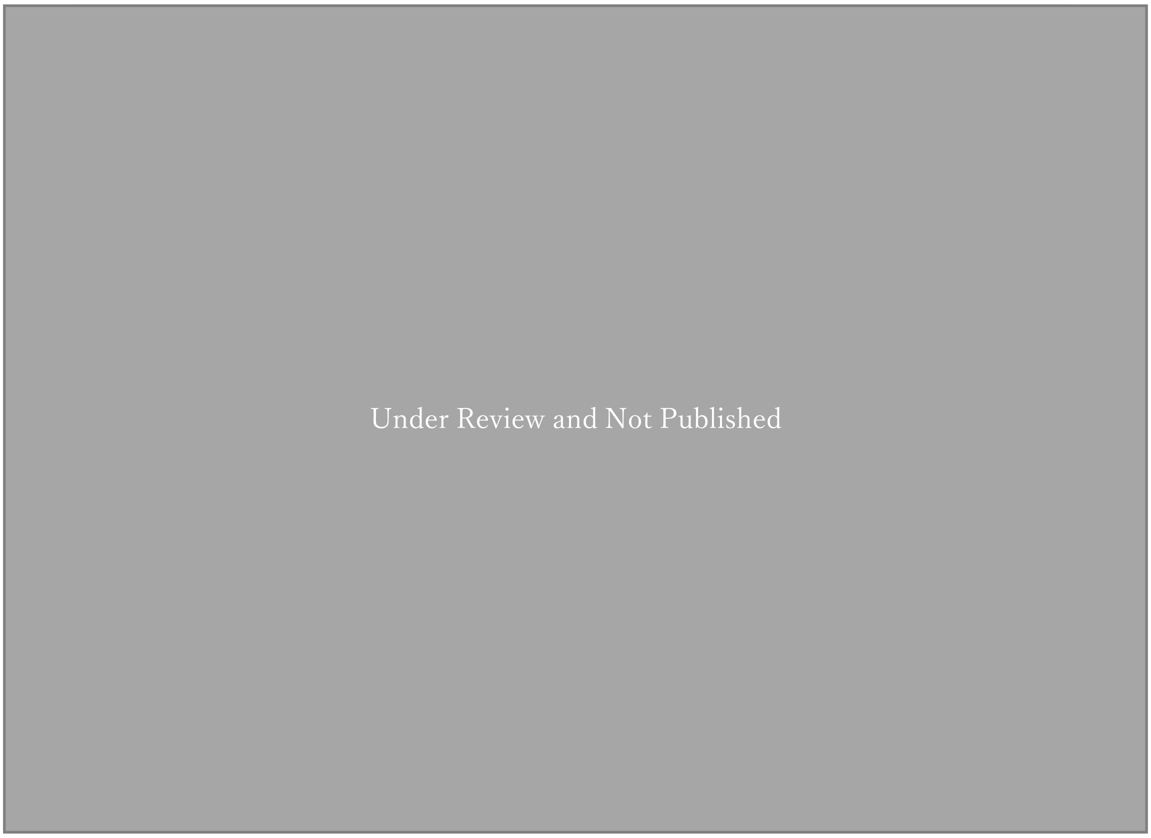


Fig. 5.4 Designed interface of the on-line operation

5.4 Field Experiment

5.4.1 Road Database Collection and Reference Subspace

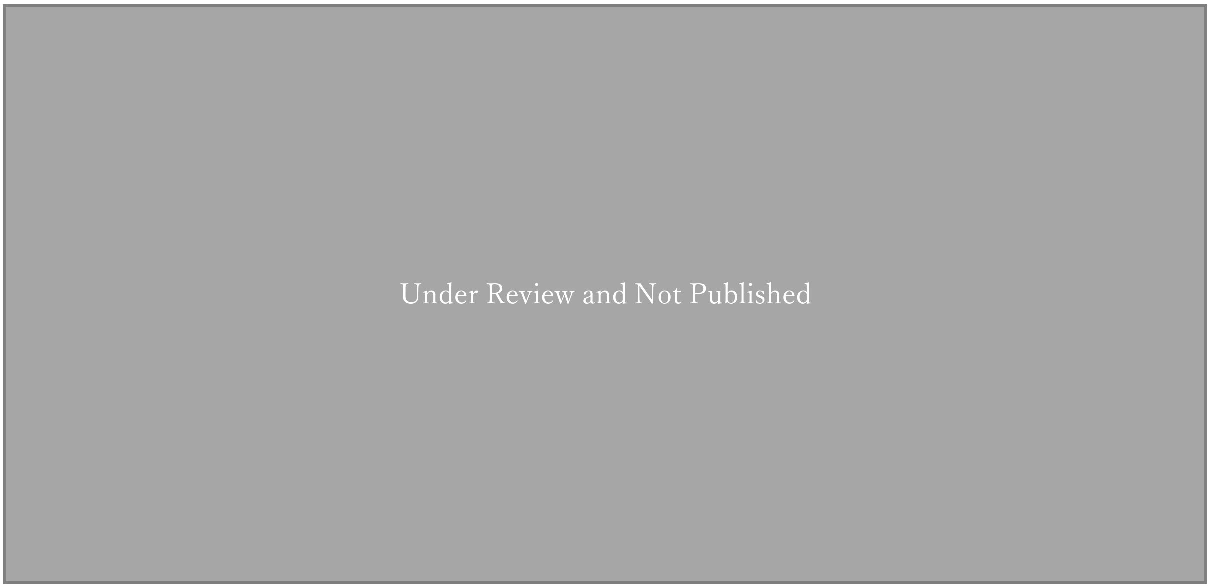
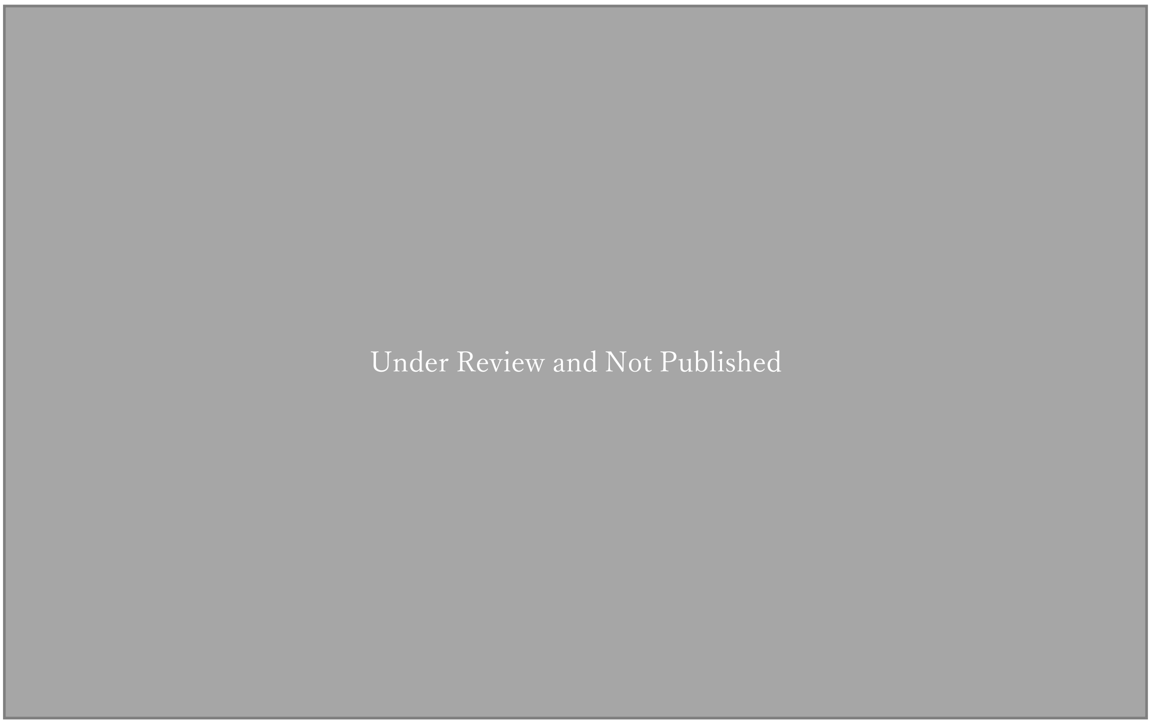


Fig. 5.5 Instance images for type of road

The three kinds of agriculture road databases are generated as follow:



For each reference subspace, odd-numbered images were selected from instance images as training data used for establish the reference subspace for offline experiment; all the instance images as training data used for establish the reference subspaces for online experiment. Each image was resized to 20×20 pixels and vectorised into a 400-dimensional vector.

5.4.2 Off-line Road Type Recognition

Under Review and Not Published

5.4.3 On-line Road Type Recognition

To simulate the real-time operation, three testing streams were constructed for the three databases. In order to improve recognition accuracy, an ideal condition was assumed and preproduced the test images by removing all the noise images which not belong to any class.

5.4.3.1 Testing Streams Generation

Test streams 1 (Fig. 5.6) and 2 (Fig. 5.7) for S road, straight road and T road under structured road and unstructured road in data base *II* and database *III* are generated as follow:

Under Review and Not Published

Under Review and Not Published

Under Review and Not Published

Fig. 5.6 Test stream 1 in database *I*

Under Review and Not Published

Fig. 5.7 Test stream 2 in database *II*

Under Review and Not Published

Fig. 5.8 Test stream 3 in database *III*

5.4.3.2 On-line Experiments Setting

Under Review and Not Published

Under Review and Not Published

Fig. 5.9 Example of test stream and sliding window

5.5 Experiment Results

5.5.1 Off-line Road Type Recognition

Under Review and Not Published

Table 5.1: Confusion matrix for recognition result for struct road and un-struct road

Under Review and Not Published

Table 5.2: Confusion matrix for recognition result for S road, Straight road and T road in un-struct road

Under Review and Not Published

Table 5.3: Confusion matrix for recognition result for S road, Straight road and T road in struct road

Under Review and Not Published

5.5.2 On-line Road Type Recognition

Under Review and Not Published

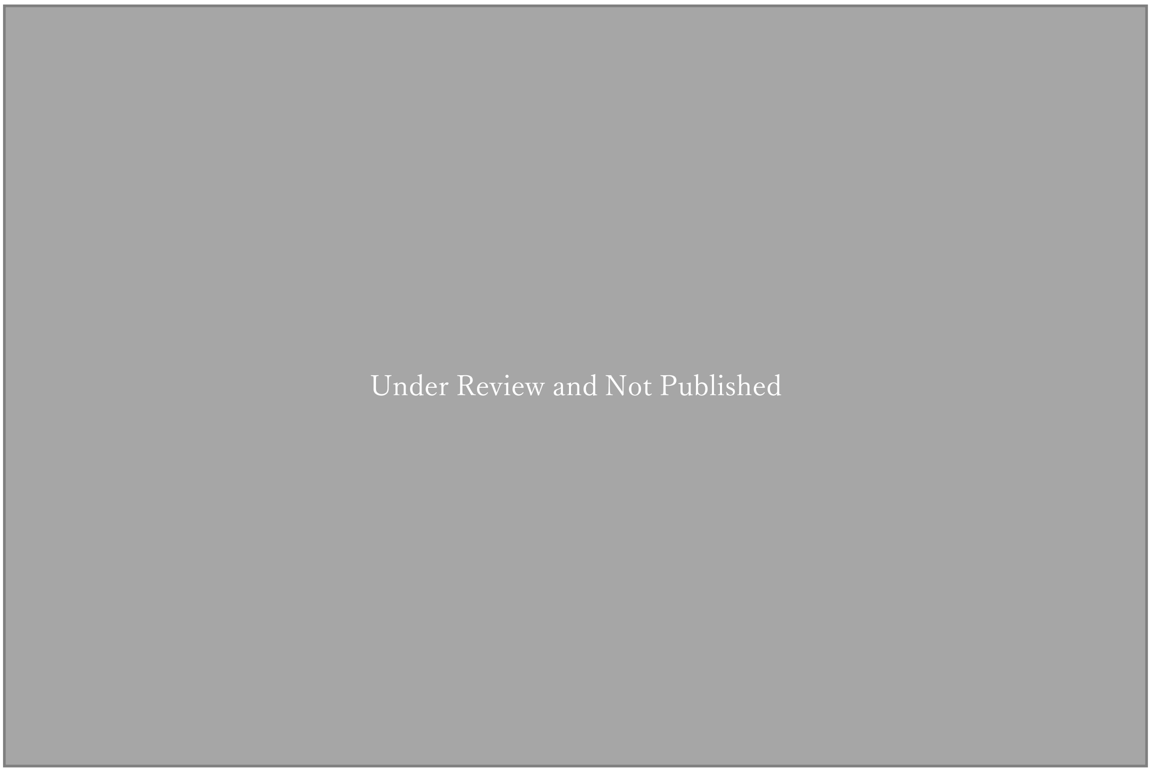


Fig. 5.10. On line results for structured road and unstructured road

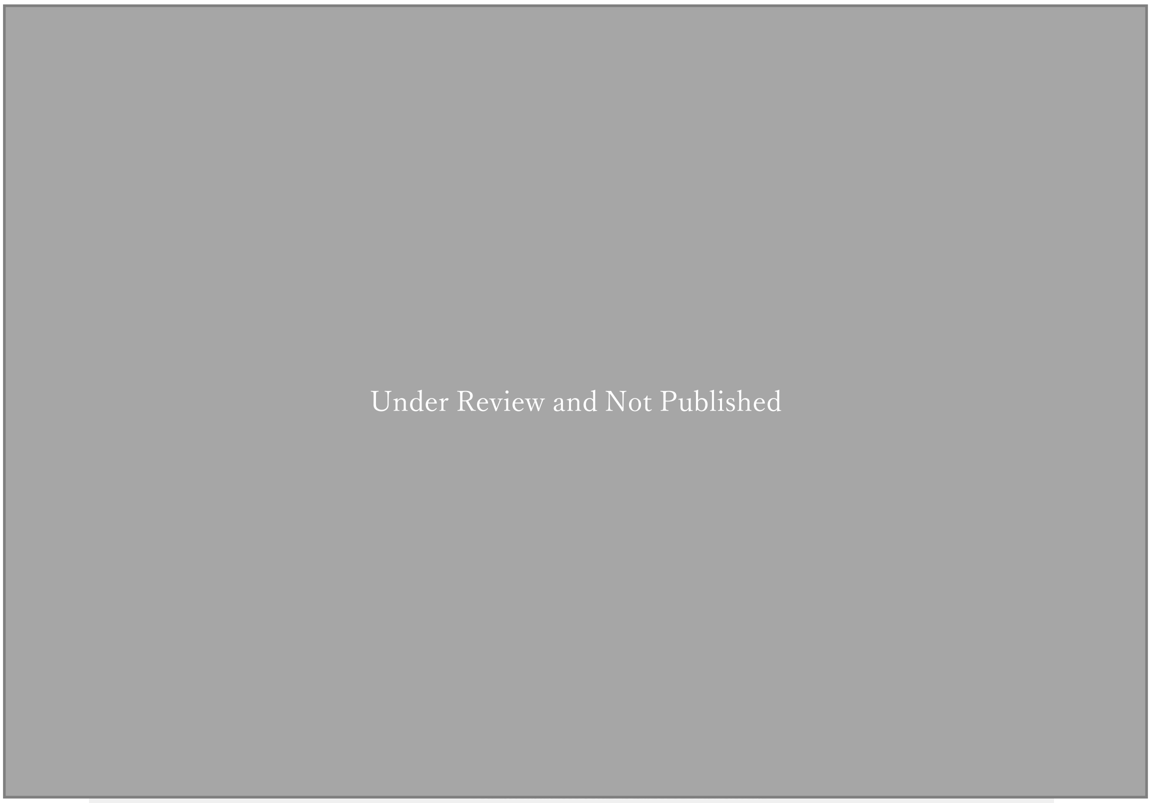


Fig. 5.11. On line results for S road, Straight road and T road in structured road

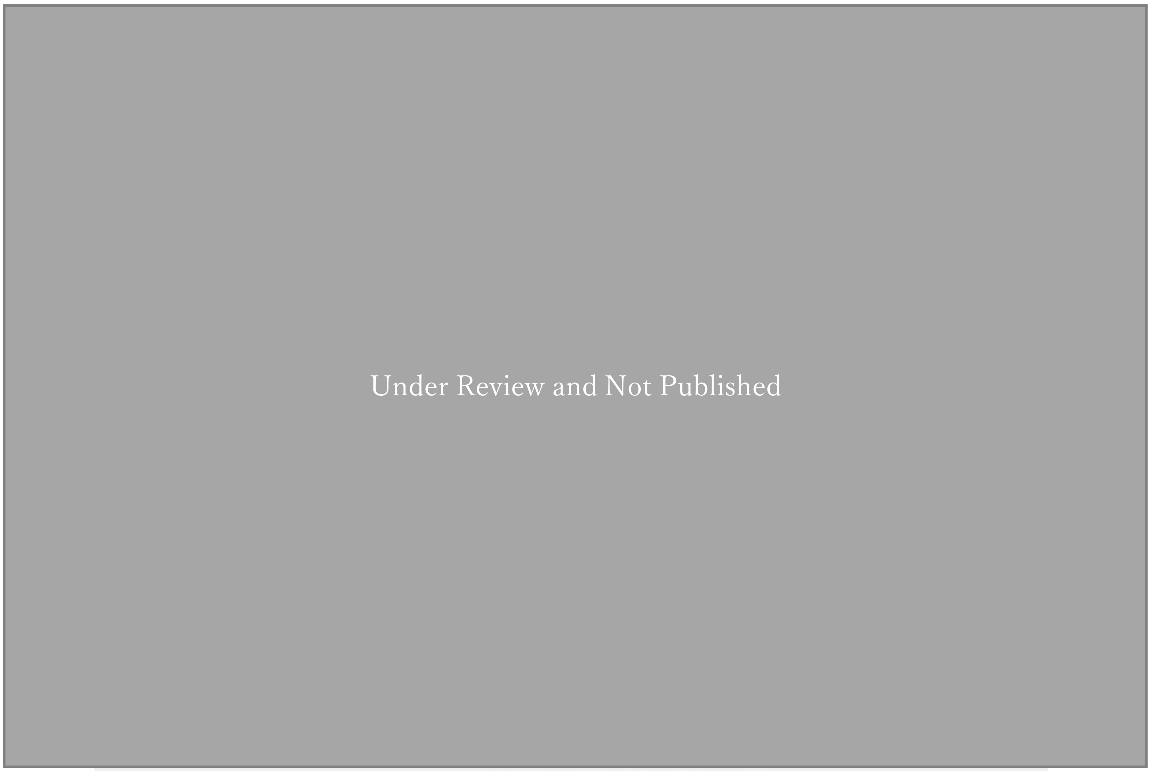


Fig. 5.12 On line result for S road, Straight road and T road in unstructured road

The similarity values between each class and testing image set for three databases were compared (Fig. 5.13, 5.14, 5.15).

Fig. 5.13 Similarity of structured and unstructured roads

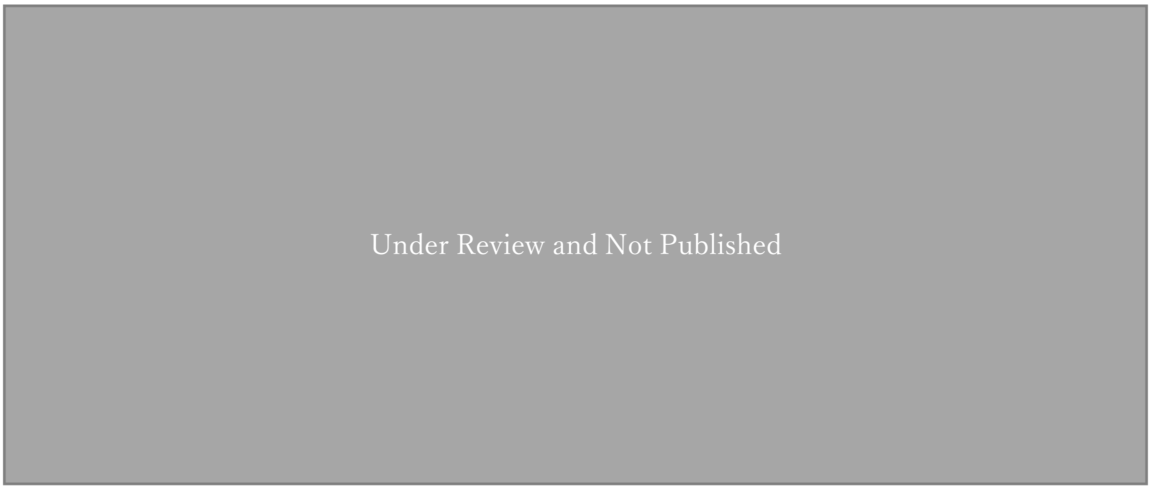


Fig. 5.14 Similarity of S road, straight road and T road in structured road

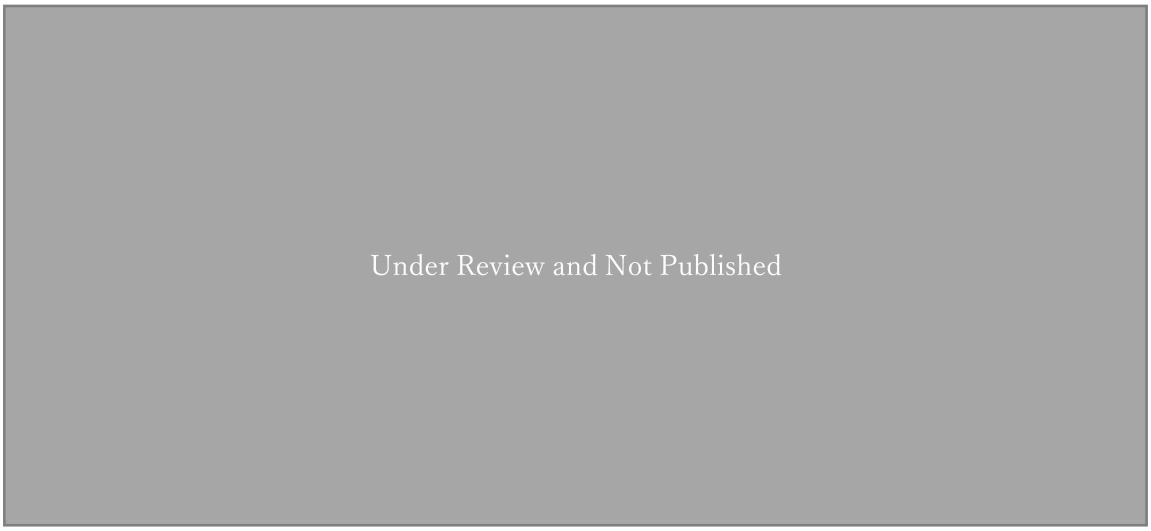


Fig. 5.15 Similarity of S road, Straight road and T road in unstructured road

From the results, it was found that that the system was able to recognize the structure road and unstructured road with very high accuracy (Fig 5.10). For the classifiers of S road, Straight road and T road, higher accuracy under unstructured road (Fig 5.12) was obtained then unstructured road (Fig 5.11). This is because, under the structure road the surround environment is not qualitative then under the unstructured road. For example, under the structure road, the road size, road mark and other vehicles on the road reduces the classifiers recognition accuracy. While under the unstructured road, there was less noise information on the road.

5.6 Conclusion

Under Review and Not Published

and accuracy requirements for enabling autonomous driving in the recognized driving environmental condition.

Chapter 6
Driver Action Monitoring in Farm Environment to Ensure Operator's Safety and Rescue System

6.1 Driver Inattention Action Dataset

The driver monitoring system was required to detect inattention actions from normal driving actions, and the driver inattention action dataset should include typical actions during driving. By analyzing the behavior patterns and actions that can possibly cause accidents, it could generalize some of the danger actions, such as looking aside, and talking on the phone. Additionally, the status of driver distraction and fatigue could finally be expressed through body motions, such as nodding and yawning. Thus, a driver inattention action dataset was established that consisted of one negative category (normal driving actions) and five inattention categories (inattention actions). The inattention categories include the following actions: looking aside, talking on the phone, nodding, rubbing eyes, and yawning (Fig. 6.1).

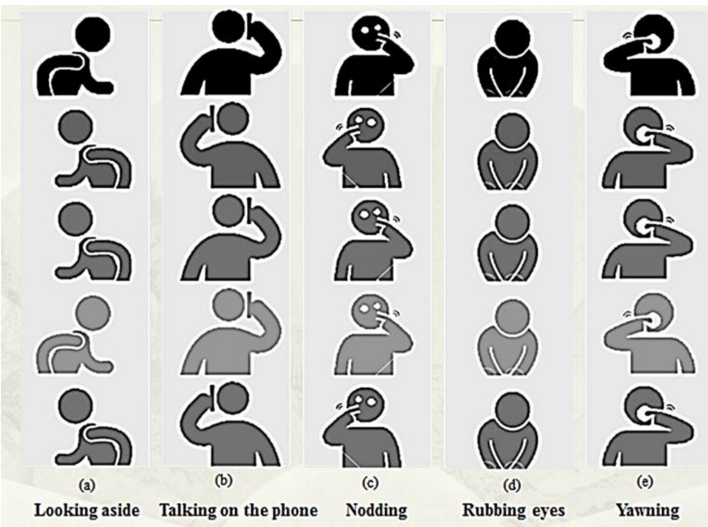


Fig. 6.1 The five inattention categories of driver

6.2 Off-line Action Recognition System

The process flow of the Hankel-based KMSM had two components: the learning phase and the recognition phase (Fig. 6.2).

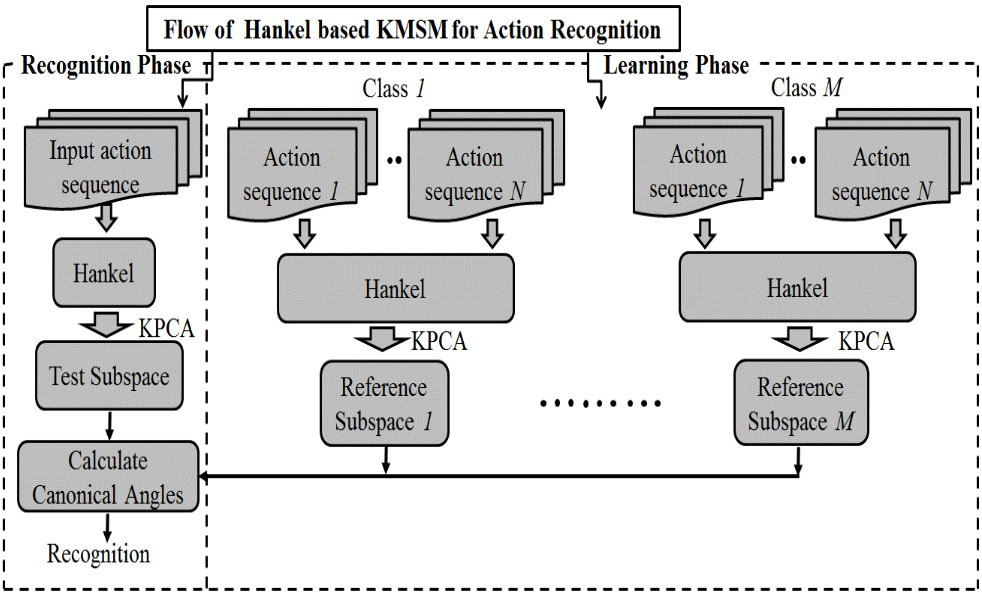


Fig. 6.2 Flow of the Hankel-based KMSM for Action Recognition

The learning phase included three stages: First, the system read the collected training action sequence images of each class $m \in \{1, \ldots, M\}$. Second, for each class a block Hankel matrix was calculated. Finally, by applying the KPCA to the generated Hankel matrix at the second step, a nonlinear subspace was established as a reference subspace for each class. The process of learning process was described by algorithm 3 as follow:

Algorithm 3 Learning Phase:

Input: Training data sets $TRB = \{\{I_j\}_{j=1}^{N}\}_{m=1}^{M}$, where M is the number of the classes.

For all $\boldsymbol{m} \in \{\mathbf{1}, \ldots, \boldsymbol{M}\}$ do

Calculate Hankel matrix H_m via (Eq. 16).

Calculate Reference Subspace P_m by applying KPCA on H_m.

End for

Save $P = \{P_m\}_{m=1}^{M}$.

The recognition phase included four stages: First, the system read the all of collected testing action sequence images of I_j, $j \in \{1, \dots, J\}$, and for each I with frames of $\{ f_1, f_2, \dots, f_n \}$. Second, for each testing action sequences a $G \times (n - r + 1)$ -dimensional block Hankel matrix can be calculated. Third, by applying the KPCA to the generated Hankel matrix at the second step, a nonlinear subspace was established as a testing subspace for each testing action sequence I_j. Finally, the canonical angles of current testing subspace and each reference subspaces θ_m was calculated and classify the current action sequence to the class that had the smallest canonical angles (the highest similarity). The process of testing process was described by algorithm 4 as follow:

Algorithm 4 Recognition Phase:

Input: Reference subspaces P; Testing data sets $TEB = \{I_j\}_{j=1}^{J}$, where J is the number of the testing sequence.

For all $j \in \{1, \dots, J\}$ do

Calculate Hankel matrix H_j via (Eq. 16).

Calculate Testing Subspace Q_j by applying KPCA on H_j.

Calculate the similarities θ_m via (3.1).

Output: class (I_j).

End for

6.3 On-line Action Recognition System

In the on-line action recognition process, the subspace patterns trained in the off-line action recognition process were utilized, and process of the on-line system included a recognition phase and an alarm phase. Considering the lower computation time that is required for a real-time system, we ensured the detection accuracy and the system stability; a sliding window with a width of N_{SW} was utilized to perform the classification. To avoid a miss-alarm, the system could trigger an alarm only when the system categorized the set of incoming images as a risk for N_{AR} times continuously.

The on-line action recognition process could be described as follows: First, the least N_{SW} continuous images were selected as a test sub action sequence, after N_{SW} continuous images were available from the Kinect. Second, we generated a block Hankel matrix using the sub action sequence. Third, we classified the set of sequential images using the KMSM classifier. Finally, we sent alarms when inattention action was detected for N_{AR} times (Fig. 6.3). A user interface was designed, and the information display on the interface included the current cropped image, the testing sub action sequence, the predicted category, the recognition rate (the correct classifications were known during the test), the computational time, and the similarity plot (Fig. 6.4).

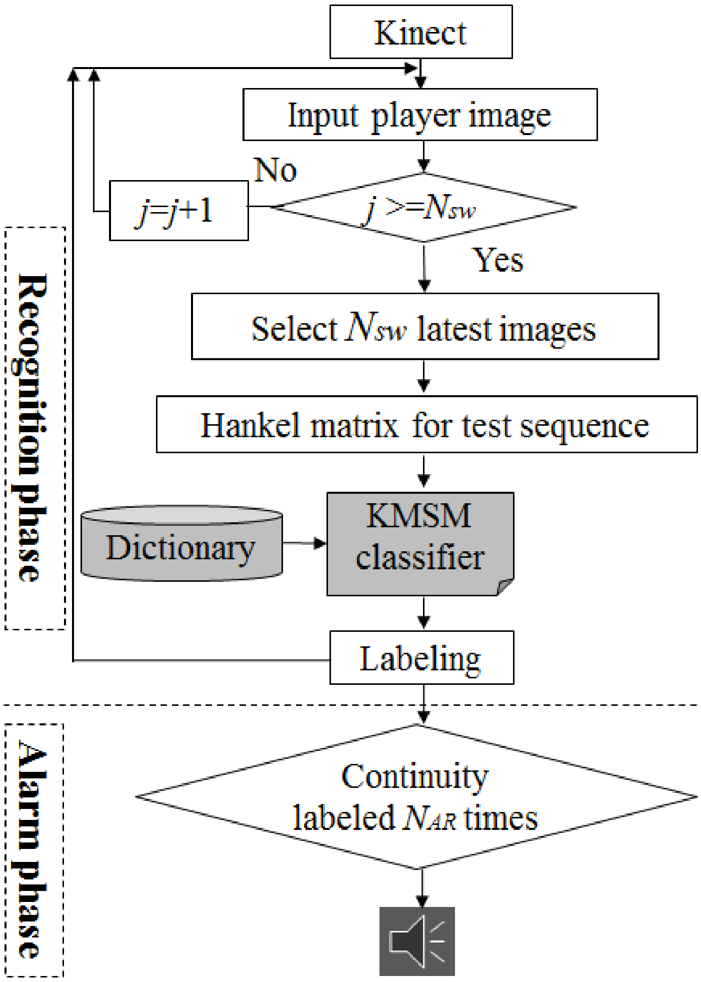


Fig. 6.3 Flow of the on-line operation

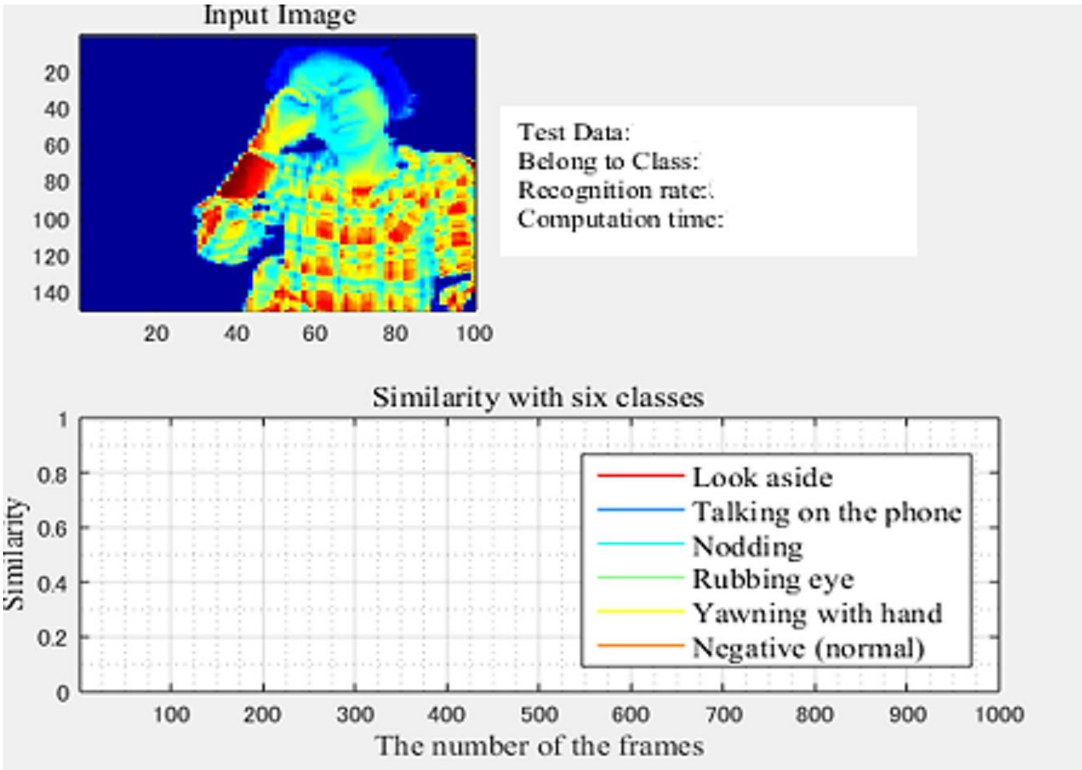


Fig. 6.4 Designed interface of the on-line operation

6.4 Field Experiment

6.4.1 Driver Inattention Action Dataset Collection and Training Dataset

The experiment data sets were collected in a tractor driver's cabin under outdoor conditions in different times. Additionally, the numbers of frames in each instance were different as each volunteer own their character and habit for actions.

6.4.1.1 Experiment Data using KINECT Sensor (KN)

Experiment KN data sets, (Fig.6.5, 6.6) was collected using a 15-fps RGB-D KINECT camera (KN) with a resolution of 320×240. Five subjects in group1 performed each category as follow: 24, 24, 23, 16, 16 action sequences for each class were performed by five subjects, respectively; Five subjects in group2 performed each category as follow: 10, 7, 7, 10, 7 action sequences for each class were performed by five subjects, respectively.

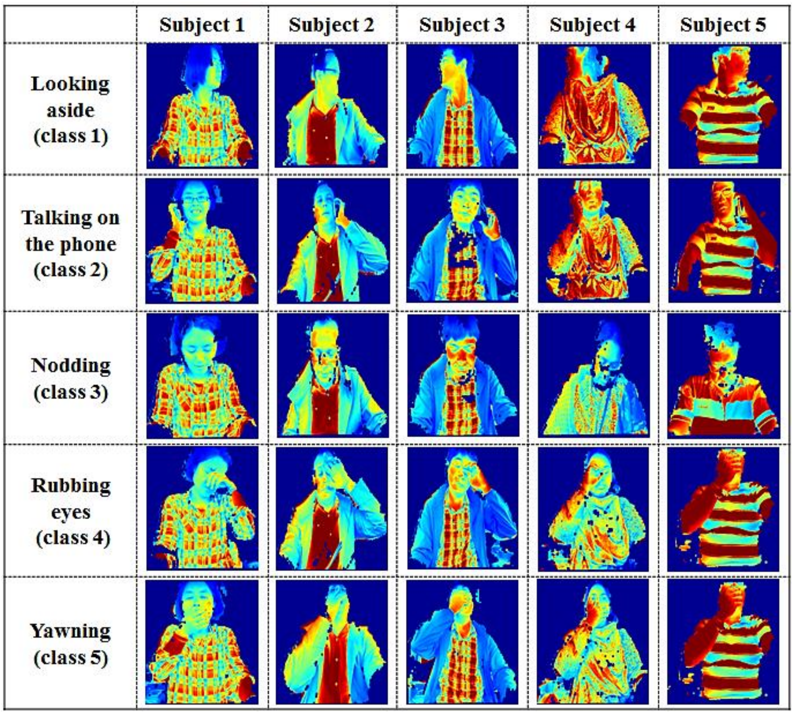


Fig. 6.5 Training and testing driver inattention actions datasets performed by group1 for similar subject recognition using KN

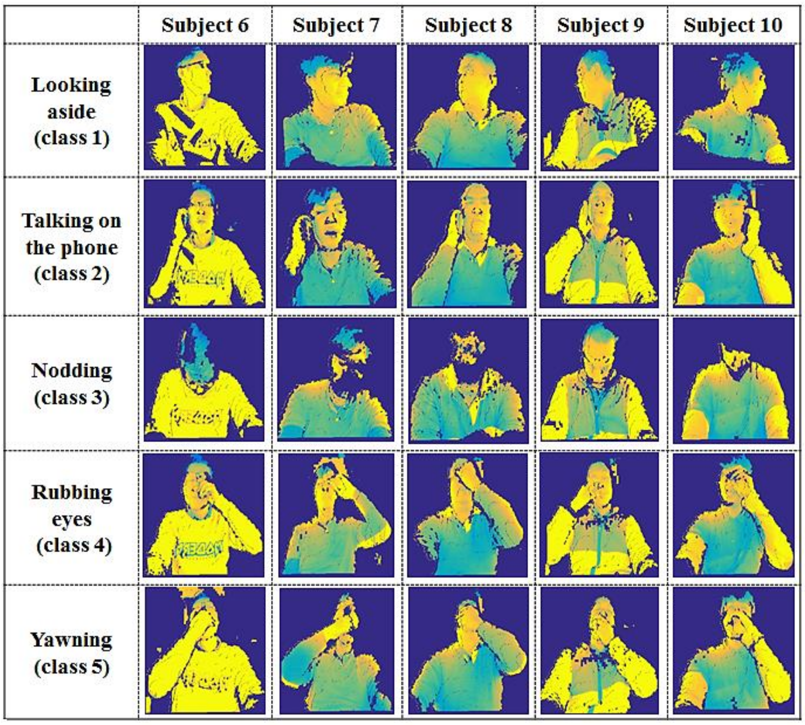


Fig. 6.6 Testing driver inattention actions datasets performed by group2 for different subject recognition using KN

Totally, 864 instance samples of depth image sequences for six categories were collected: 618 instance samples were performed by 5 subjects in group1, and 246 instance samples were performed by 5 subjects in group2. For KN training data, odd-numbered samples were selected from instance samples which performed by group1, and the number of odd-numbered samples used for establish the training dataset was 312.

6.4.1.2 Experiment Data using Thermal Sensor (TH)

Experiment TH data sets (Fig.6.7, 6.8), was collected using a 30fps THERMAL sensor (TH, Flir Vue pro 336) with a resolution of 336×256. Five subjects in group 3 performed each category as follow: 29, 23, 19, 14, 12 action sequences for each class were performed by five subjects, respectively; Five subjects in group4 performed each category as follow: 6, 11, 10, 11, 9 action sequences for each class were performed by five subjects, respectively.

Totally, 864 instance samples of depth image sequences for six categories were collected: 582 instance samples were performed by 5 subjects in group3, and 282 instance samples were performed by 5 subjects in group4. For TH training data, same with experiment data *I*, odd-numbered samples were selected from instance samples which performed by group3, and the number of odd-numbered samples used for establish the training dataset was 291.

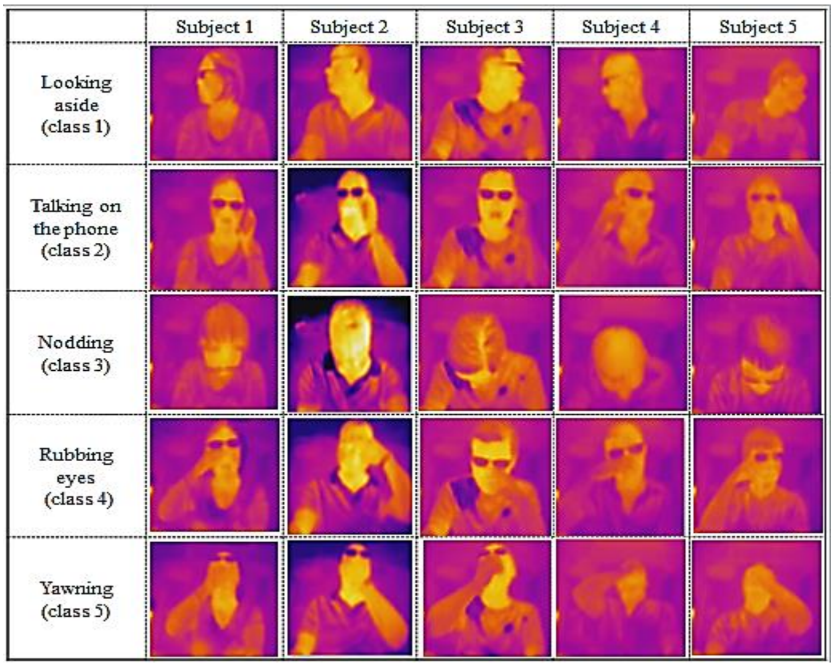


Fig. 6.7 Training and testing driver inattention actions datasets performed by group3 for similar subject recognition using TH

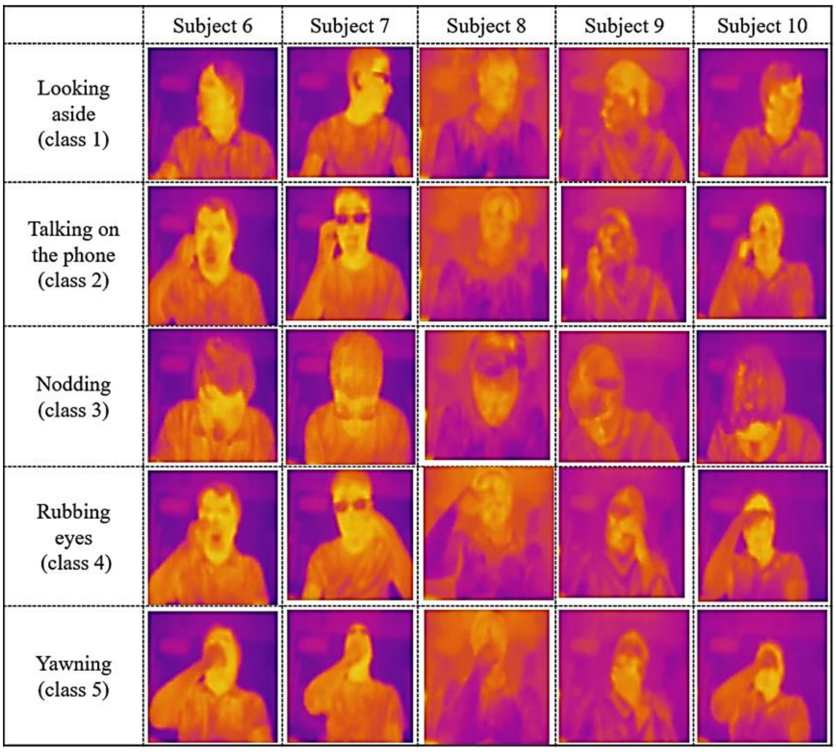


Fig. 6.8 Testing driver inattention actions datasets performed by group4 for different subject recognition using TH

6.4.2 Off-line Action Recognition

In the off-line experiment, for experiment KN data, the rest 306 instance samples conducted by the 5 subjects in group1 were utilized for similar subjects testing, and 246 instance samples conducted by the 5 subjects in group2 were utilized for different subjects testing; for experiment TH data, the rest 291 instance samples conducted by the 5 subjects in group3 were utilized for similar subjects testing, and 282 instance samples conducted by the 5 subjects in group4 were utilized for different subjects testing.

The experimental procedure was set as follow: The parameter r in the Hankel matrix was empirically set to 4. Each image was resized to 10×10 pixels and vectorised into a 100-dimensional vector. Thus, the dimension of the column vector in the Hankel matrix was 400 (100-dimensions×4 frames), which was much lower than the dimension of the other existing action features, such as STACOG (Kobayashi & Otsu, 2012). The statistical analyses for off-line datasets were performed while training and testing subjects are similar and different. The six classes including normal driving actions were considered to develop the confusion matrix for determining the false negative and false positive rate for each of the classification.

6.4.3 On-line Action Recognition

To simulate a real-time on-line operation, the testing stream data was constructed by combining action sequences from the action instance samples in the six categories. For KN action dataset, data performed by subjects in group1 was utilized for similar subjects testing and data performed by subjects in group2 was utilized for different subjects testing; for TH action dataset, data performed by subjects in group3 was utilized for similar subjects testing and data performed by subjects in group4 was utilized for different subjects testing. Additionally, the class of normal driving action sequences was inserted before each of the other actions**.** The numbers in brackets were frame numbers for normal driving actions. The total number of frames in the testing stream data (KN) was 999 and 854 for similar subjects (group1) testing and different subjects (group2) testing, respectively (**Table 6.1, Table 6.2)**;

Table 6.1 Frame numbers of each inattention categories in test stream for similar subjects testing (KN).

	Looking aside	*Talking on the phone*	*Nodding*	*Rubbing eyes*	*Yawning*
Subject1	*(16)/15*	*(12)/55*	*(11)/9*	*(15)/23*	*(17)/8*
Subject2	*(16)/20*	*(14)/54*	*(14)/10*	*(13)/22*	*(11)/25*
Subject3	*(9)/30*	*(16)/43*	*(11)/14*	*(13)/24*	*(17)/26*
Subject4	*(14)/31*	*(15)/51*	*(11)/15*	*(11)/27*	*(10)/24*
Subject5	*(11)/18*	*(11)/47*	*(11)/31*	*(11)/34*	*(10)/23*

Table 6.2 Frame numbers of each inattention categories in test stream for different subjects testing (KN).

	Looking aside	*Talking on the phone*	*Nodding*	*Rubbing eyes*	*Yawning*
Subject6	*(11)/17*	*(11)/25*	*(16)/11*	*(15)/24*	*(11)/15*
Subject7	*(16)/23*	*(19)/28*	*(16)/17*	*(16)/14*	*(12)/29*
Subject8	*(20)/23*	*(9)/24*	*(17)/22*	*(19)/14*	*(23)/23*
Subject9	*(21)/13*	*(8)/20*	*(14)/26*	*(13)/26*	*(10)/17*
Subject10	*(13)/29*	*(11)/11*	*(7)/14*	*(14)/9*	*(13)/25*

The total number of frames in the testing stream data (TH) was 864 and 798 for similar subjects (group3) testing and different subjects (group4) testing, respectively **(Table 6.3, Table 6.4**).

Table 6.3 Frame numbers of each inattention categories in test stream for similar subjects testing (TH).

	Looking aside	*Talking on the phone*	*Nodding*	*Rubbing eyes*	*Yawning*
Subject1	*(10)/27*	*(15)/21*	*(20)/15*	*(34)/13*	*(13)/17*
Subject2	*(12)/20*	*(8)/13*	*(11)/28*	*(13)/22*	*(23)/31*
Subject3	*(14)/10*	*(14)/22*	*(22)/17*	*(21)/13*	*(20)/9*
Subject4	*(13)/15*	*(13)/29*	*(14)/13*	*(13)/23*	*(15)/15*
Subject5	*(12)/35*	*(18)/19*	*(14)/9*	*(16)/17*	*(14)/19*

Table 6.4 Frame numbers of each inattention categories in test stream for different subjects testing (TH).

	Looking aside	*Talking on the phone*	*Nodding*	*Rubbing eyes*	*Yawning*
Subject6	*(12)/23*	*(19)/20*	*(14)/24*	*(22)/14*	*(19)/18*
Subject7	*(17)/12*	*(11)/22*	*(15)/13*	*(8)/18*	*(7)/16*
Subject8	*(12)/13*	*(23)/24*	*(20)/13*	*(10)/20*	*(22)/9*
Subject9	*(9)/21*	*(10)/21*	*(16)/10*	*(16)/10*	*(16)/22*
Subject10	*(17)/25*	*(17)/21*	*(14)/8*	*(15)/7*	*(15)/18*

In the on-line experiment, the width of sliding window NSW was set as 6. A sample of the testing stream for one subject and working with a sliding window was indicated (Fig. 6.9). To reduce the computational time, each image was first resized to 10×10 pixels and vectorised. Then, a 400×3 Hankel matrix was generated using Hankel with parameter

r=4. Finally, a nonlinear subspace was constructed by applying KPCA to the Hankel matrix.

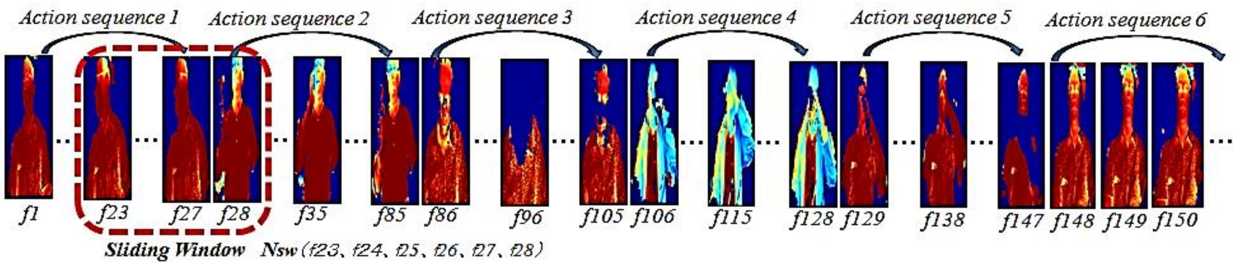


Fig. 6.9 Examples of data stream and sliding window.

6.5 Experiment Results

6.5.1 Off-line Action Recognition

The classification accuracy rate and error classification rate for each driver actions in similar subjects (group1) testing and different subjects (group2) testing are observed in Table 5 and Table 6, respectively. The values in black on the tables represent the recognition classification rate of actions on the vertical direction of the matrix, and the values in dark grey in the cell represents the rate of error classification into actions on the horizontal direction of the matrix.

6.5.1.1 Off-line Action Recognition using Kinect Camera (KN)

From the KN dataset results (Table 5 and 6), it showed that both in the similar subjects (group1) testing and in the different subjects (group2) testing, system could obtain relatively satisfactory classification accuracy rate. In the similar subjects (group1) testing arrived at a 100% and 100%, for nodding and normal driving, respectively. In the different subjects (group2) testing arrived at a 97% and 95%, for nodding and normal driving, respectively. As there exist a vast difference of the nodding and normal driving action compared with the other actions.

In the similar subjects (group1) testing, the action of looking aside possessed the lowest classification accuracy rate of 78% caused by the miss classification to the class of normal driving action; while in the different subjects (group2) testing, the rubbing eyes possessed the lowest classification accuracy rate of 61% caused by the miss classification to the class of yawning and normal driving action (Table 5 and 6). From the statistical analysis it was observed that the action recognition of inattention action was higher except for rubbing eyes in the different subjects testing. Overall, the average classification accuracy rates were 91.18% and 86.18% for similar subjects (group1) testing and different subjects (group2) testing, respectively. The EER for similar subjects testing different subjects testing were 0.192 and 0.358, respectively.

Table 6.5 Confusion matrix for recognition result of driver KN inattention dataset for similar testing and training datasets on off-line recognition method

	Looking Aside	*Taking on the Phone*	*Nodding*	*Rubbing Eyes*	*Yawning*	*Normal Driving*	*False Negative (FN)*
Looking aside	*0.78*	*0.06*	*0*	*0*	*0.02*	*0.14*	*0.22*
Talking on the Phone	*0*	*0.96*	*0*	*0*	*0.04*	*0*	*0.04*
Nodding	*0*	*0*	*1*	*0*	*0*	*0*	*0*
Rubbing Eyes	*0.02*	*0.02*	*0*	*0.86*	*0.1*	*0*	*0.14*
Yawning	*0*	*0.06*	*0*	*0*	*0.86*	*0.08*	*0.14*
Normal Driving	*0*	*0*	*0*	*0*	*0*	*1*	*0*
False Positive (FP)	*0.02*	*0.14*	*0*	*0*	*0.16*	*0.22*	

Table 6.6 Confusion matrix for recognition result of driver KN inattention dataset for different testing datasets on off-line recognition method

	Looking Aside	*Taking on the Phone*	*Nodding*	*Rubbing Eyes*	*Yawning*	*Normal Driving*	*False Negative (FN)*
Looking aside	*0.85*	*0.07*	*0*	*0*	*0.03*	*0.05*	*0.15*
Talking on the Phone	*0*	*0.85*	*0*	*0.07*	*0.03*	*0.05*	*0.15*
Nodding	*0.03*	*0*	*0.97*	*0*	*0*	*0*	*0.03*
Rubbing Eyes	*0*	*0.07*	*0*	*0.61*	*0.17*	*0.15*	*0.39*
Yawning	*0*	*0*	*0*	*0.07*	*0.9*	*0.03*	*0.1*
Normal Driving	*0*	*0*	*0*	*0.05*	*0*	*0.95*	*0.05*
False Positive	*0.03*	*0.14*	*0*	*0.19*	*0.23*	*0.28*	

6.5.1.2 Off-line Action Recognition using Thermal Camera (TH)

From the TH dataset results (Table 7 and 8), it showed that both in the similar subjects (group3) testing and in the different subjects (group4) testing, system could obtain relatively satisfactory classification accuracy rate. In the similar subjects (group3) testing also arrived at a 100% and 100%, for nodding and normal driving, respectively. As there exist a vast difference of the nodding and normal driving action compared with the other actions.

In the different subjects (group4) testing, the actions of taking on the phone and rubbing eyes possessed the lowest classification accuracy rate of 4% and 13% caused by the mainly miss classification to the class of yawning and nodding driving action, respectively. Overall, the average classification accuracy rates were 92.2% and 47.52% for similar subjects (group3) testing and different subjects (group4) testing, respectively.

Table 6.7 Confusion matrix for recognition result of driver TH inattention dataset for similar testing and training datasets on off-line recognition method

	Looking Aside	*Taking on the Phone*	*Nodding*	*Rubbing Eyes*	*Yawning*	*Normal Driving*	*False Negative (FN)*
Looking aside	*0.91*	*0*	*0*	*0*	*0.02*	*0.14*	*0.16*
Talking on the Phone	*0*	*0.89*	*0*	*0.09*	*0*	*0.02*	*0.11*
Nodding	*0*	*0*	*1*	*0*	*0*	*0*	*0*
Rubbing Eyes	*0.02*	*0.02*	*0*	*0.87*	*0.09*	*0.03*	*0.16*
Yawning	*0*	*0*	*0*	*0.11*	*0.85*	*0.04*	*0.15*
Normal Driving	*0*	*0*	*0*	*0*	*0*	*1*	*0*
False Positive (FP)	*0.02*	*0.02*	*0*	*0.2*	*0.11*	*0.23*	

Table 6.8 Confusion matrix for recognition result of driver TH inattention dataset for different testing datasets on off-line recognition method

	Looking Aside	*Taking on the Phone*	*Nodding*	*Rubbing Eyes*	*Yawning*	*Normal Driving*	*False Negative (FN)*
Looking aside	*0.70*	*0.02*	*0.065*	*0.065*	*0*	*0.15*	*0.3*
Talking on the Phone	*0.17*	*0.04*	*0*	*0*	*0.51*	*0.28*	*0.96*
Nodding	*0.02*	*0*	*0.79*	*0.04*	*0*	*0.15*	*0.21*
Rubbing Eyes	*0.04*	*0*	*0.36*	*0.13*	*0.235*	*0.235*	*0.87*
Yawning	*0*	*0*	*0.02*	*0*	*0.64*	*0.34*	*0.36*
Normal Driving	*0.09*	*0.02*	*0.28*	*0.02*	*0.04*	*0.55*	*0.45*
False Positive (FP)	*0.32*	*0.04*	*0.725*	*0.125*	*0.785*	*1.155*	

6.5.2 On-line Action Recognition

For KN experiment data, the recognition rate for the on-line experimental results in similar subjects (group1) testing and different subjects (group2) testing were 87.02% and 79.97%, respectively. And the average computation time was 0.07s for each recognizing cycle (**Fig. 6.10, 6.11**).

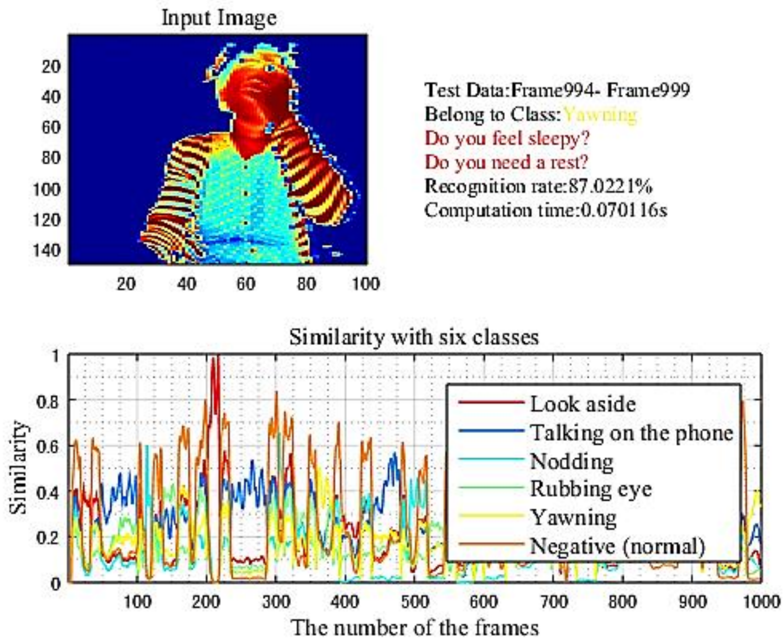


Fig. 6.10 On-line experiment (KN) result for similar subjects (group1) for training and testing use same datasets (group1).

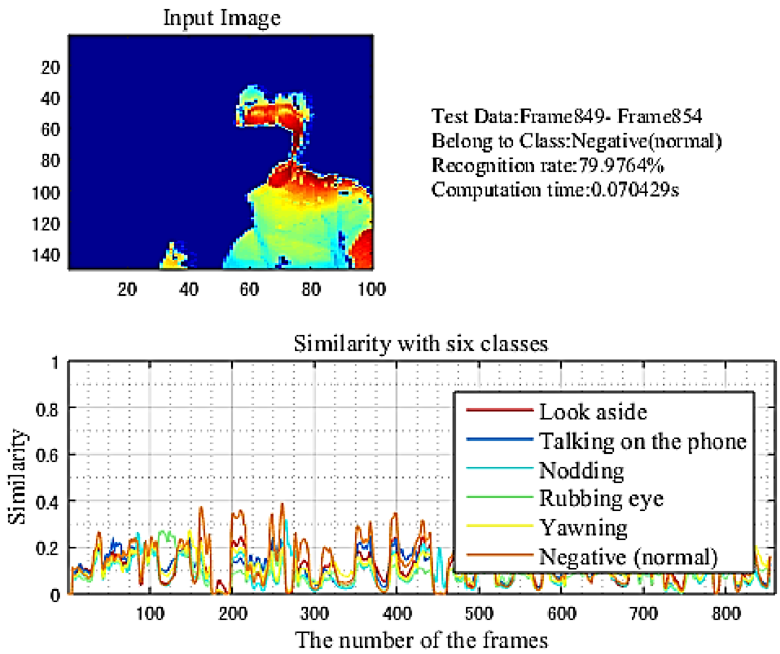


Fig. 6.11 On-line experiment result (KN) for similar subjects (group1) for training and testing use different datasets (group2).

For TH experiment data, the recognition rate for the on-line experimental results in similar subjects (group3) testing and different subjects (group4) testing were 83.24% and 42.75%, respectively. And the average computation time was 0.08s for each recognizing cycle (Fig. 6.12, 6.13)

The similarity values between each class and testing sub-sequence were compared (Fig. 6.14, 6.15, 6.16, 6.17). The experimental results showed that there exists misclassification when a testing sequence is entered from the previous testing sequence, for example when a testing sequence of looking aside started at frame No.17 (Fig 6.14 a), the first several frames were still recognized as normal driving action. As described above, the 6-width sliding window was utilized for classification, and the frames from both classes contained in the sliding window at the boundary of the two classes. It showed that with the sliding of window, the similarity rate of the previous class decreased, and the similarity rate of the next arriving class increased and the next arriving class could be correctly classified in 5 frames.

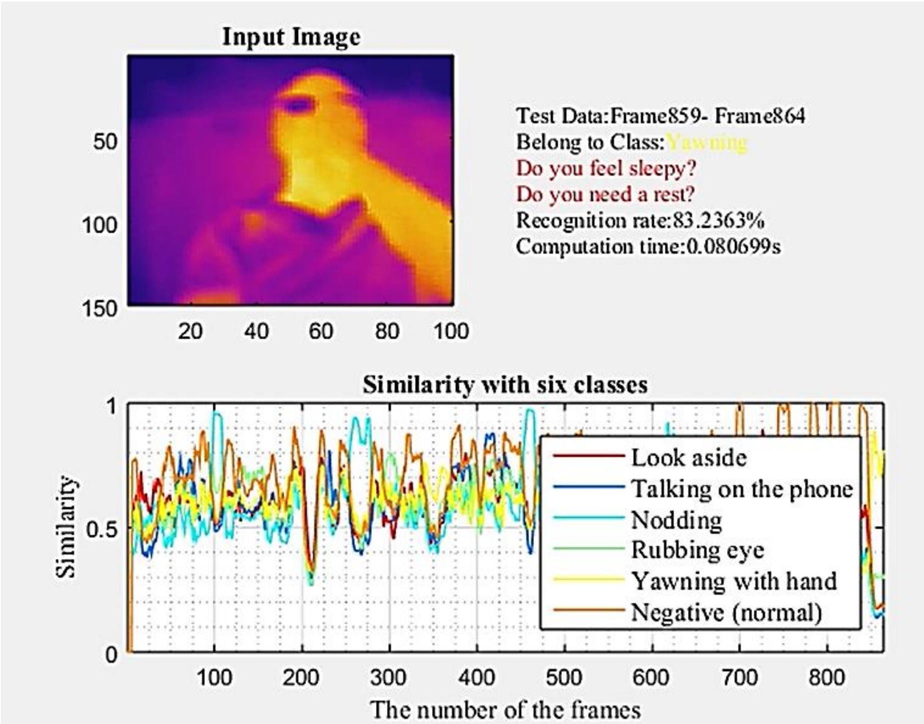


Fig. 6.12 On-line experiment (TH) result for similar subjects (group3) for training and testing use same datasets (group3).

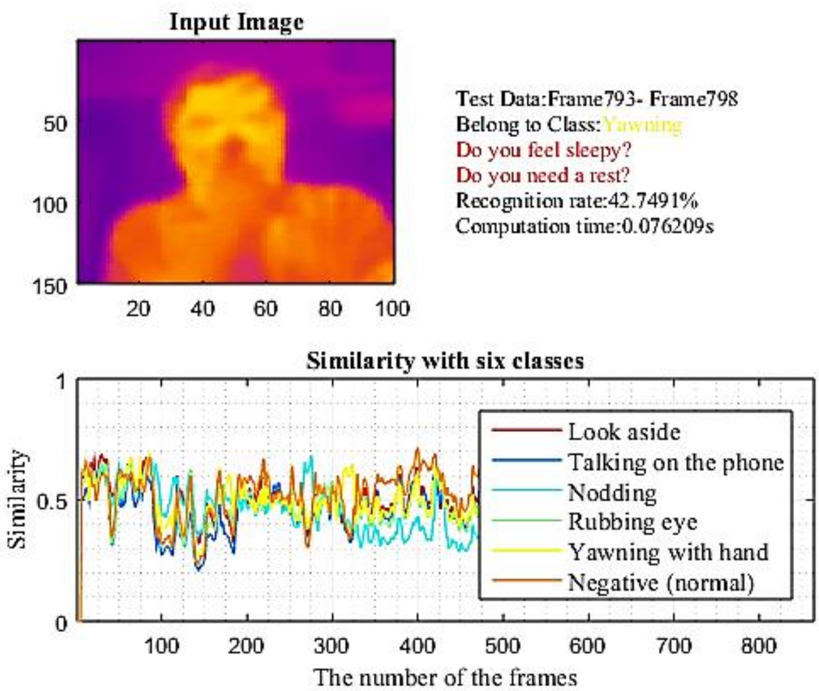
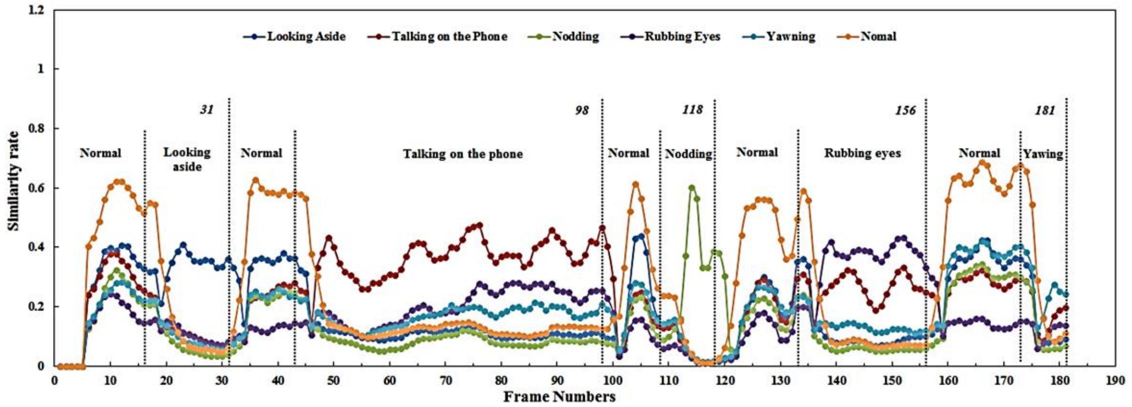
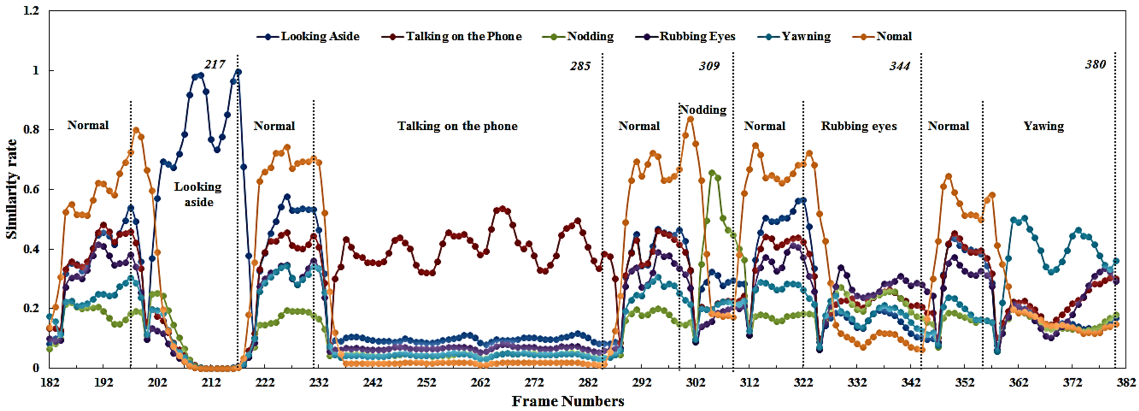


Fig. 6.13 On-line experiment result (KN) for similar subjects (group3) for training and testing use different datasets (group4).



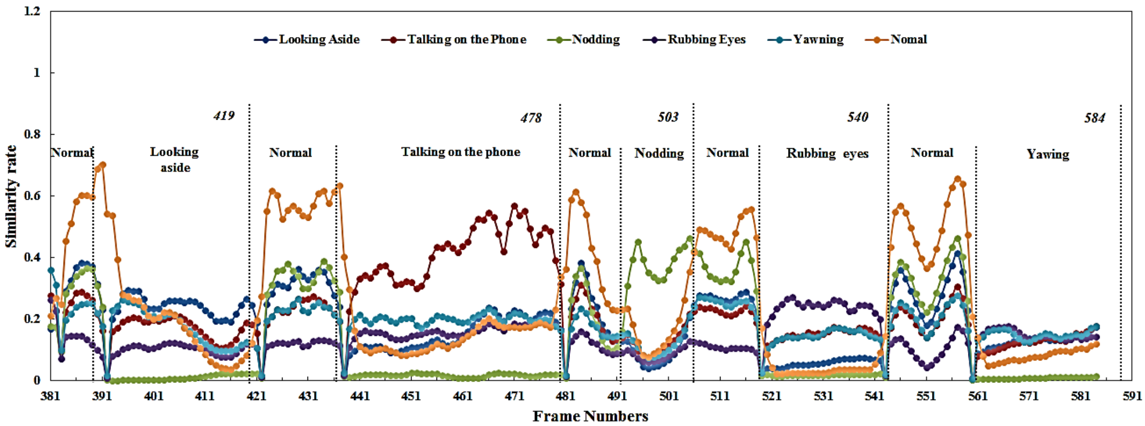
Looking Aside
Talking on the Phone
Nodding
Rubbing Eyes
Yawning
Nomal
Similarity rate
Frame Numbers
Normal
Looking aside
Normal
Talking on the phone
Normal
Nodding
Normal
Rubbing eyes
Normal
Yawing
31
98
118
156
181

(a)



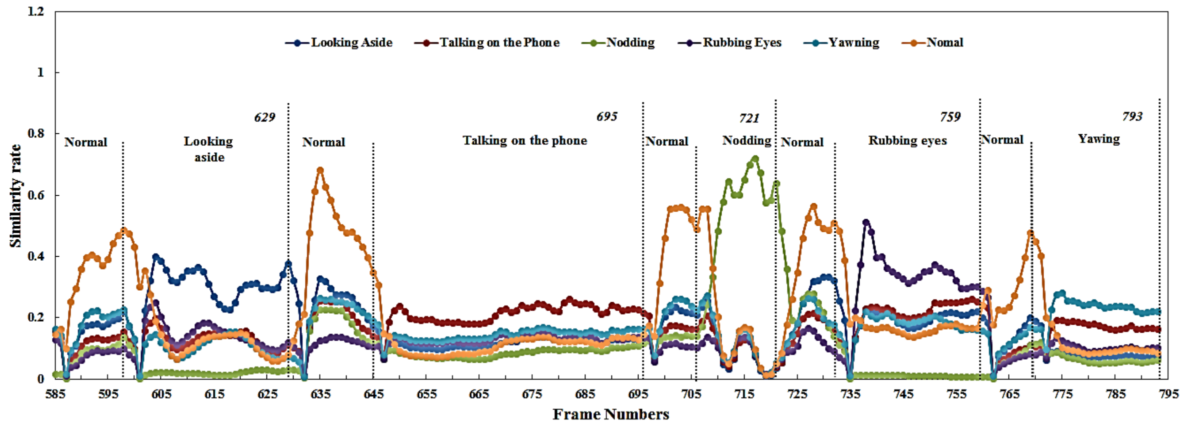
Looking Aside
Talking on the Phone
Nodding
Rubbing Eyes
Yawning
Nomal
Similarity rate
Frame Numbers
Normal
Looking aside
Normal
Talking on the phone
Normal
Nodding
Normal
Rubbing eyes
Normal
Yawing
217
285
309
344
380

(b)

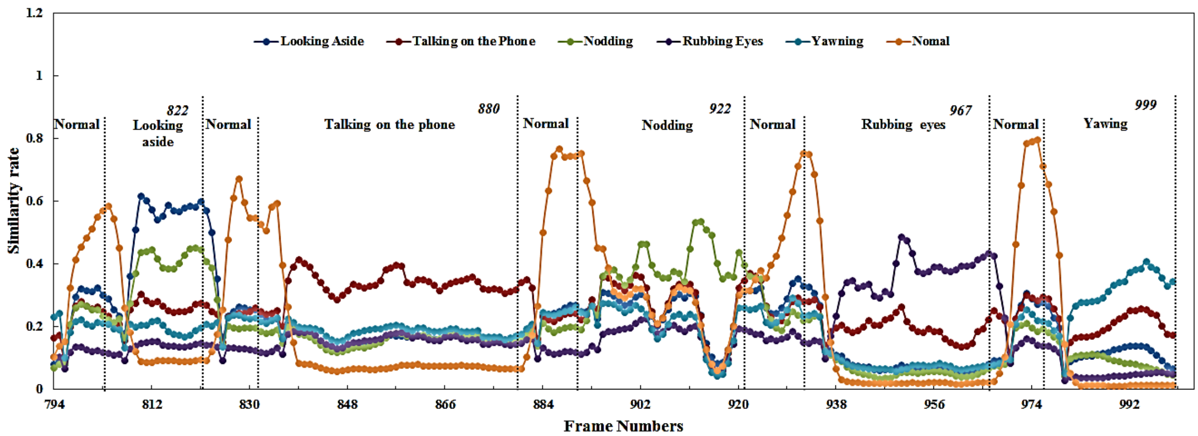


Looking Aside
Talking on the Phone
Nodding
Rubbing Eyes
Yawning
Nomal
Similarity rate
Frame Numbers
Normal
Looking aside
Normal
Talking on the phone
Normal
Nodding
Normal
Rubbing eyes
Normal
Yawing
419
478
503
540
584

(c)

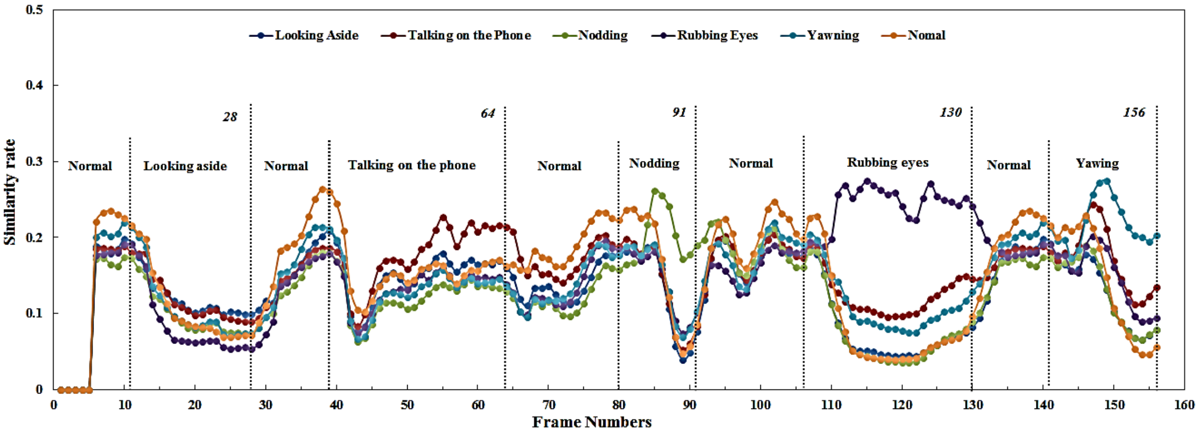


(d)



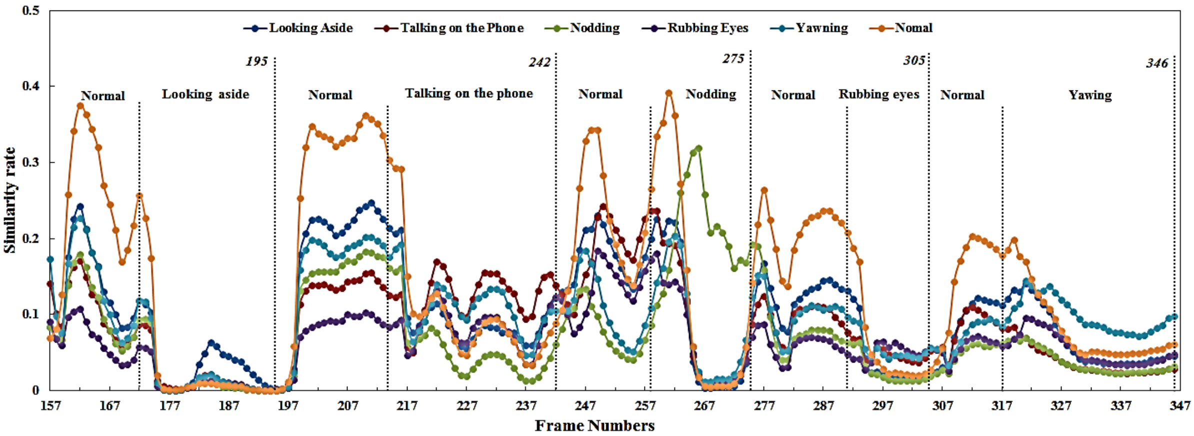
(e)

Fig. 6.14 Similarity of action sequence for KN datasets when training and Testing are same datasets (a) Subject 1, (b). Subject 2, (c) Subject 3, (d) Subject 4, (e) Subject 5.



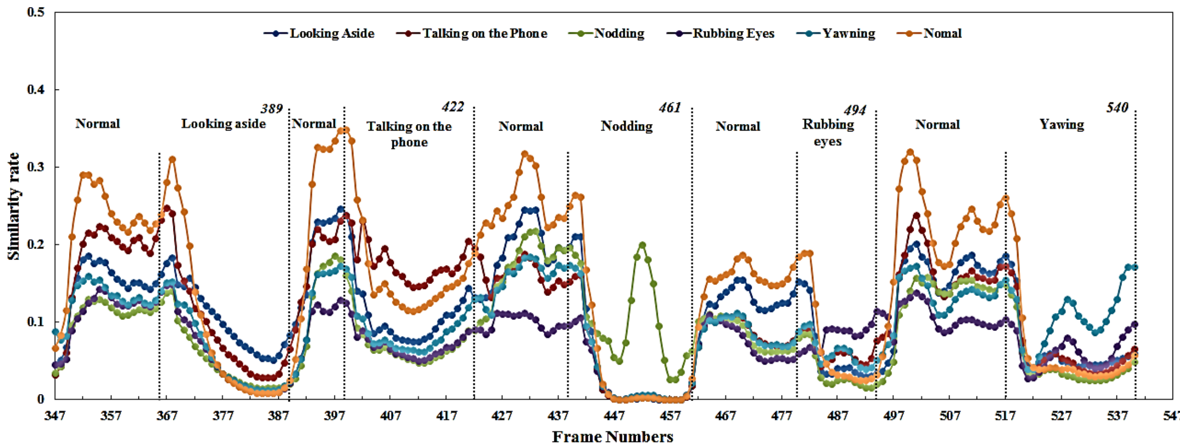
Looking Aside
Talking on the Phone
Nodding
Rubbing Eyes
Yawning
Nomal
Normal
Looking aside
Normal
Talking on the phone
Normal
Nodding
Normal
Rubbing eyes
Normal
Yawing
28
64
91
130
156
Similarity rate
Frame Numbers

(a)



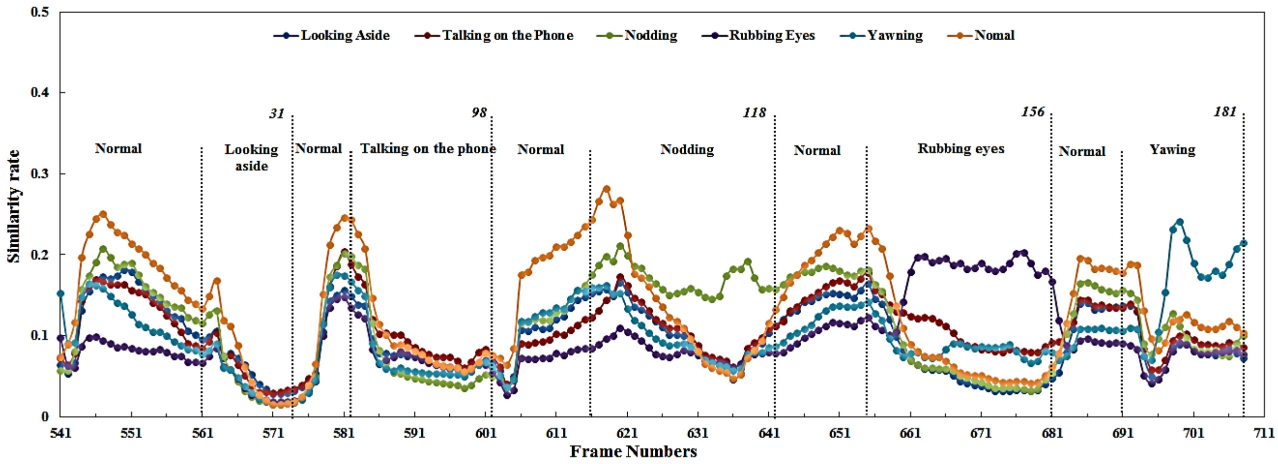
Looking Aside
Talking on the Phone
Nodding
Rubbing Eyes
Yawning
Nomal
Normal
Looking aside
Normal
Talking on the phone
Normal
Nodding
Normal
Rubbing eyes
Normal
Yawing
195
242
275
305
346
Similarity rate
Frame Numbers

(b)

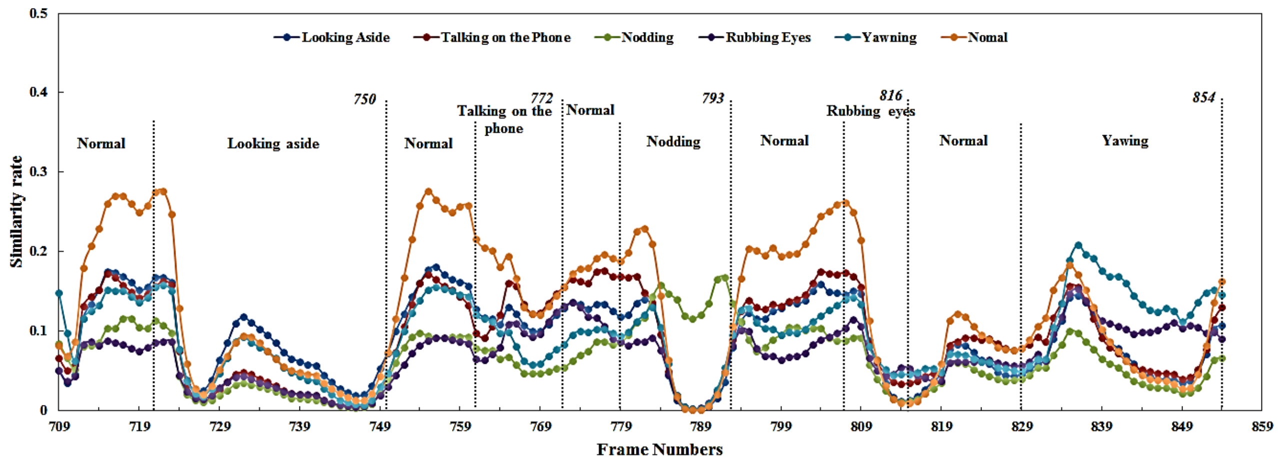


Looking Aside
Talking on the Phone
Nodding
Rubbing Eyes
Yawning
Nomal
Normal
Looking aside
Normal
Talking on the phone
Normal
Nodding
Normal
Rubbing eyes
Normal
Yawing
389
422
461
494
540
Similarity rate
Frame Numbers

(c)

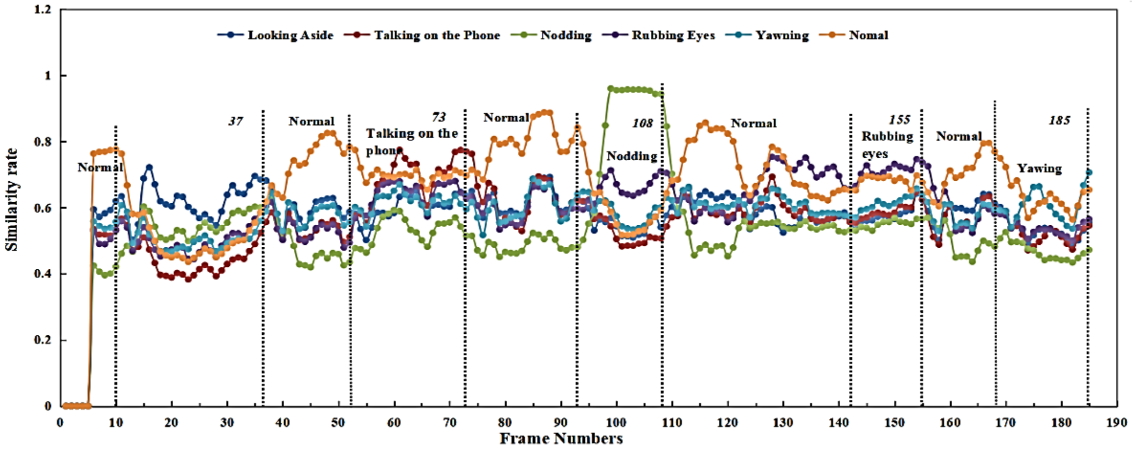


(d)

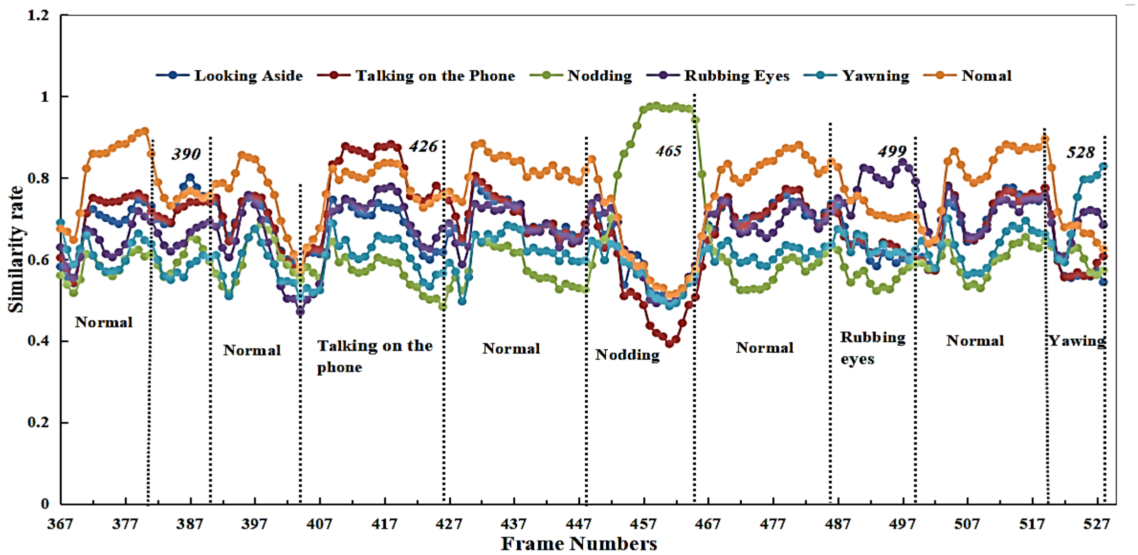


(e)

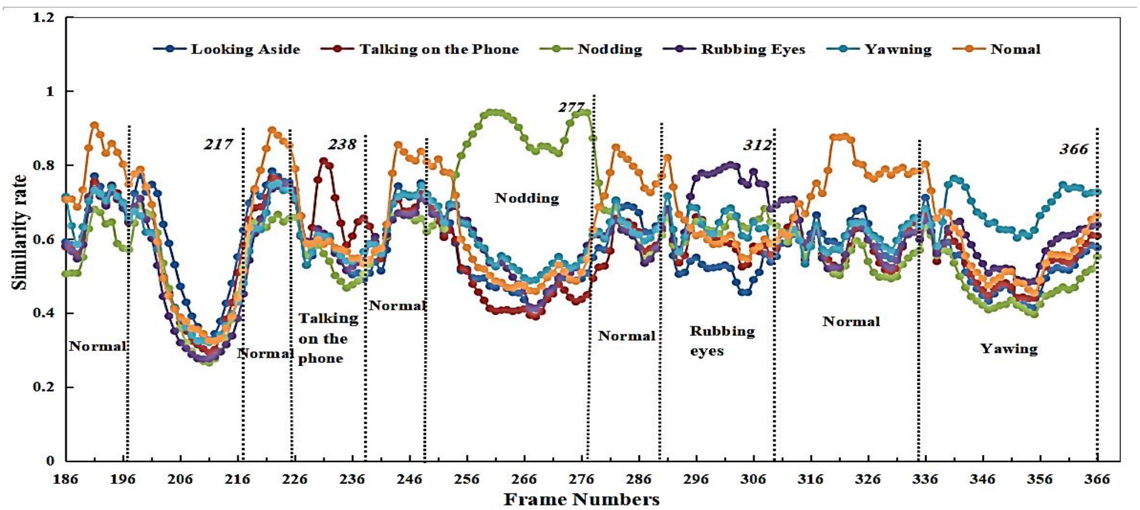
Fig. 6.15 Similarity of action sequences for KN datasets when training and testing are different (a) Subject 6, (b). Subject 7, (c) Subject 8, (d) Subject 9, (e) Subject 10.


Looking Aside
Talking on the Phone
Nodding
Rubbing Eyes
Yawning
Nomal
Similarity rate
Frame Numbers
Normal
37
Normal
Talking on the phone
73
Normal
108
Nodding
Normal
155
Rubbing eyes
Normal
185
Yawing

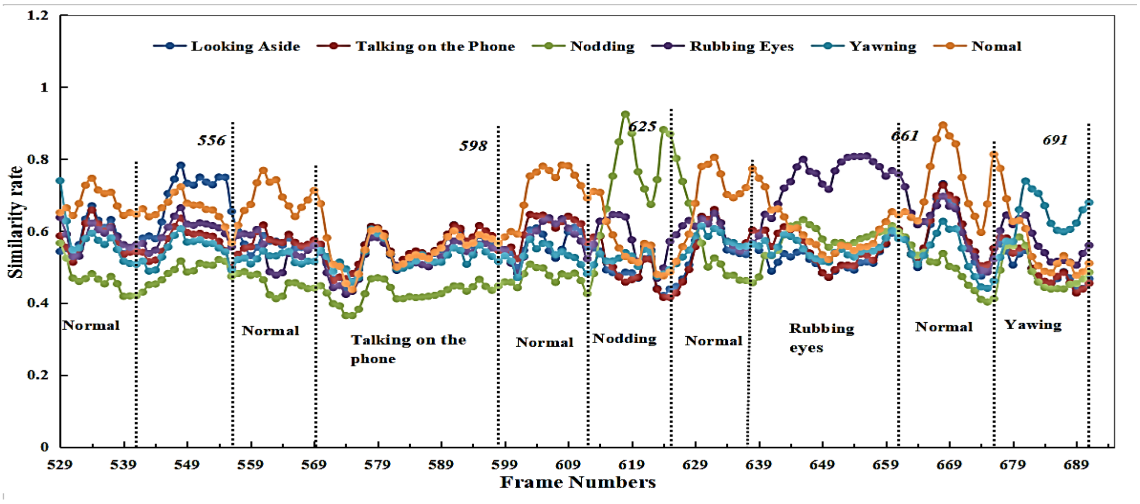
(a)


Looking Aside
Talking on the Phone
Nodding
Rubbing Eyes
Yawning
Nomal
Similarity rate
Frame Numbers
Normal
390
Normal
Talking on the phone
426
Normal
Nodding
465
Normal
Rubbing eyes
499
Normal
528
Yawing

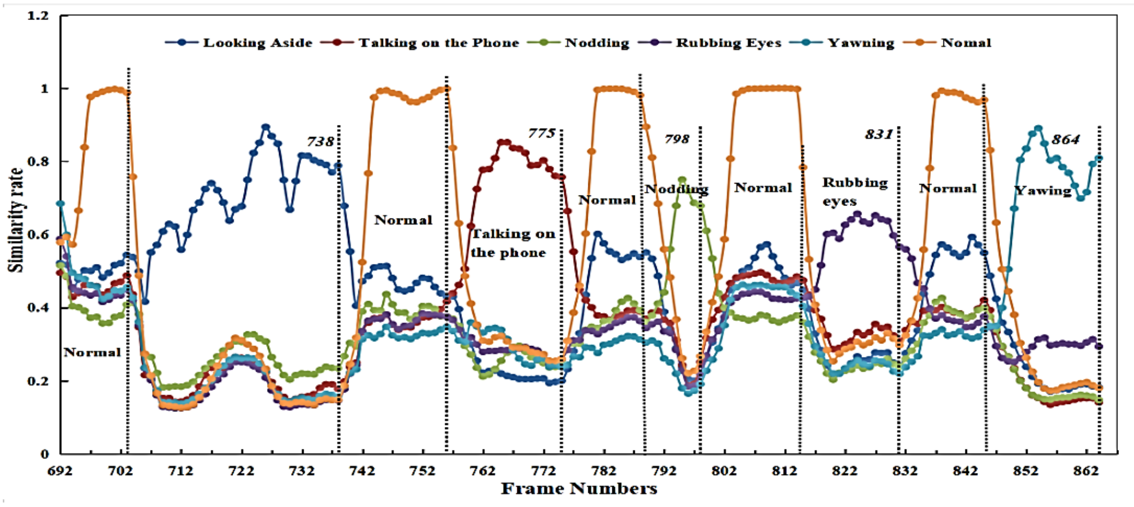
(b)


Looking Aside
Talking on the Phone
Nodding
Rubbing Eyes
Yawning
Nomal
Similarity rate
Frame Numbers
Normal
217
Normal
238
Talking on the phone
Normal
Nodding
277
Normal
Rubbing eyes
312
Normal
366
Yawing

(c)

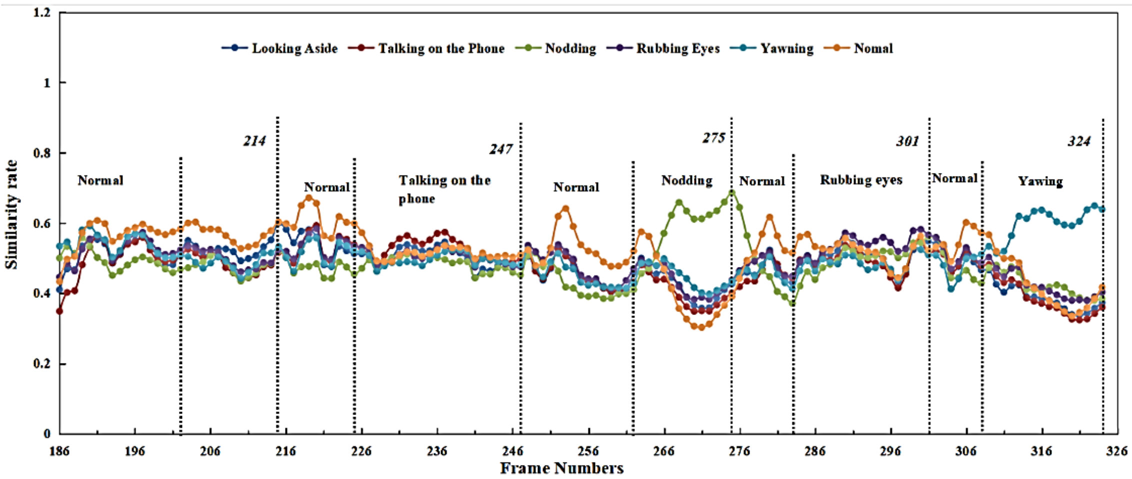


(d)



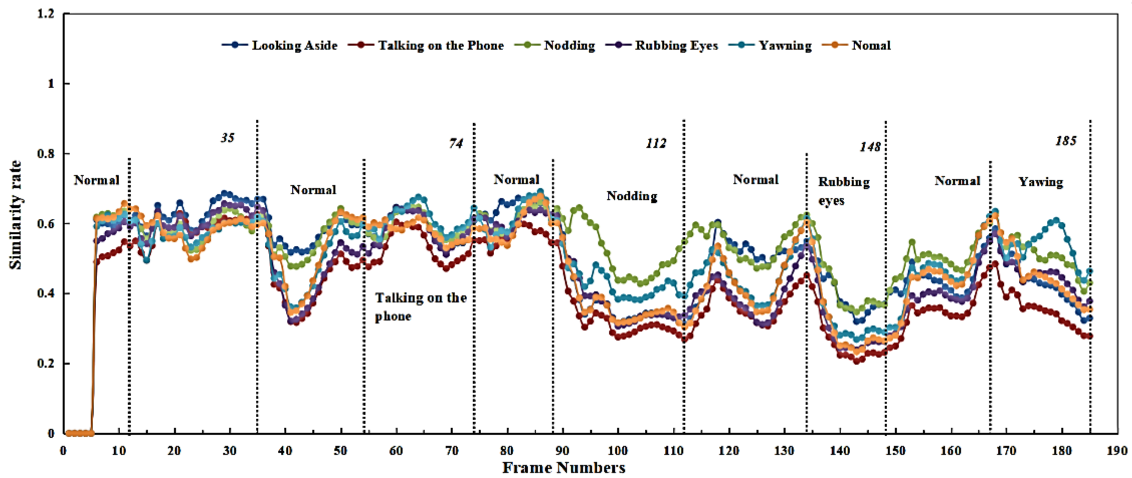
(e)

Fig. 6.16 Similarity of action sequence for TH datasets when training and Testing are same datasets (a) Subject 1, (b). Subject 2, (c) Subject 3, (d) Subject 4, (e) Subject 5.



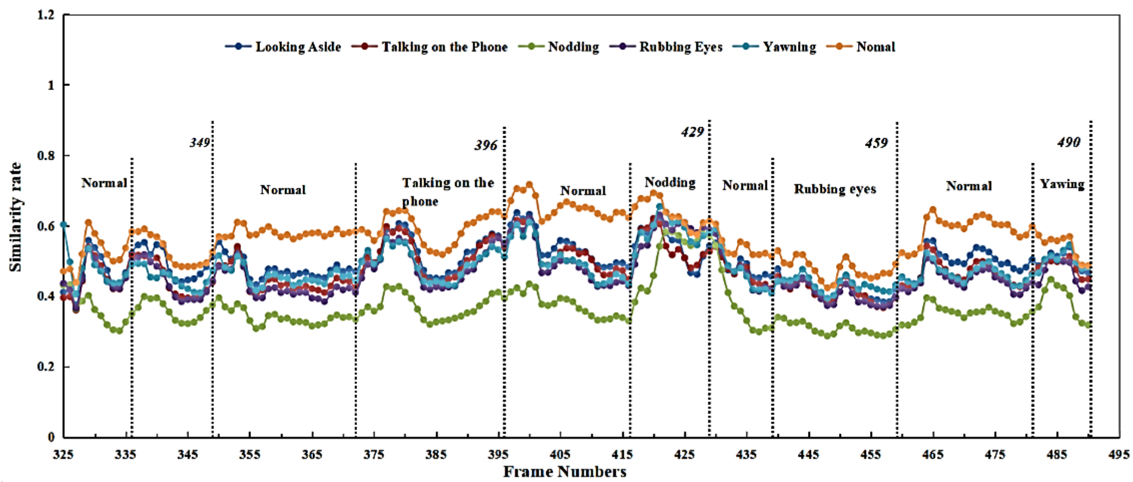
Looking Aside
Talking on the Phone
Nodding
Rubbing Eyes
Yawning
Nomal
Normal
214
Normal
Talking on the phone
247
Normal
Nodding
275
Normal
Rubbing eyes
301
Normal
Yawing
324
Similarity rate
Frame Numbers

(a)



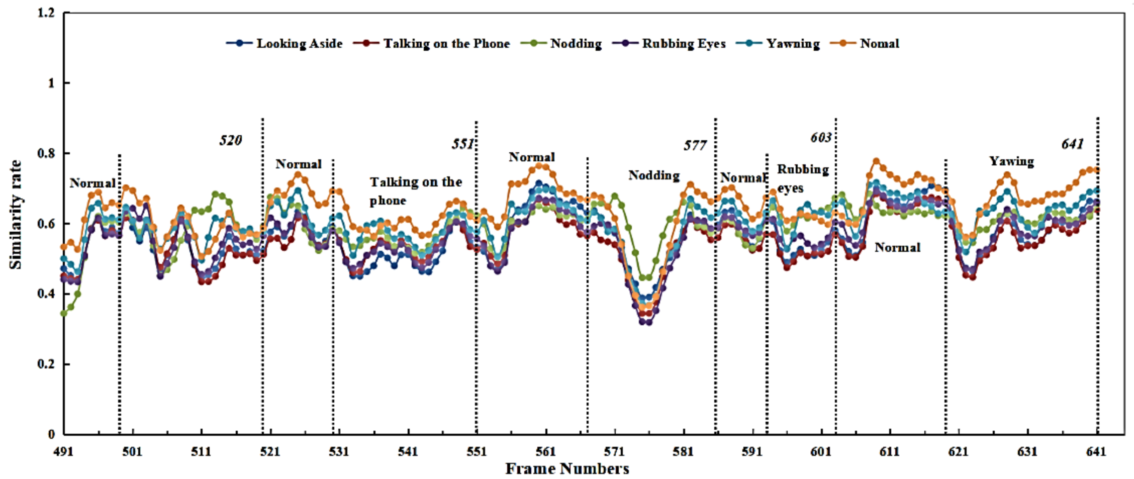
Looking Aside
Talking on the Phone
Nodding
Rubbing Eyes
Yawning
Nomal
Normal
35
Normal
Talking on the phone
74
Normal
112
Nodding
Normal
148
Rubbing eyes
Normal
185
Yawing
Similarity rate
Frame Numbers

(b)

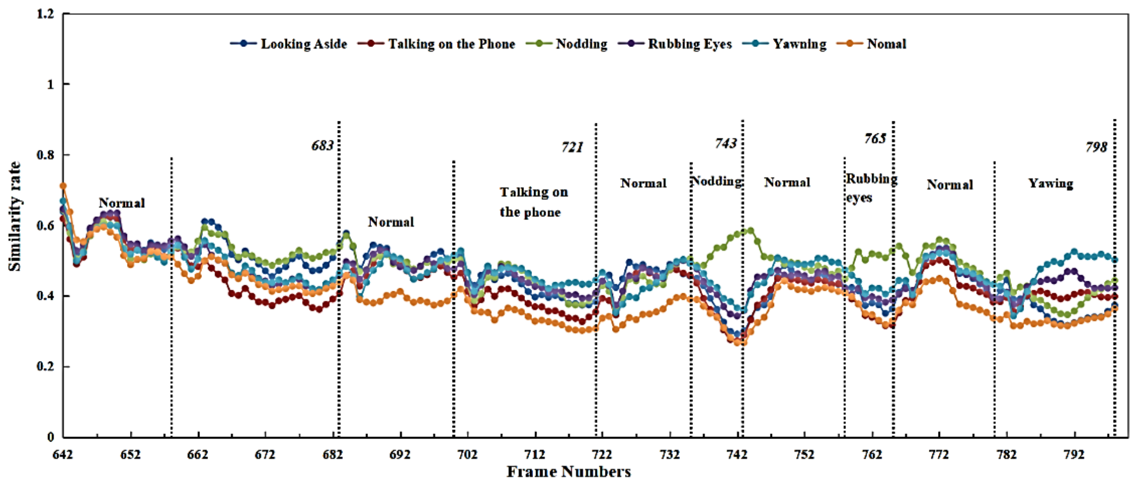


Looking Aside
Talking on the Phone
Nodding
Rubbing Eyes
Yawning
Nomal
Normal
349
Normal
Talking on the phone
396
Normal
429
Nodding
Normal
Rubbing eyes
459
Normal
490
Yawing
Similarity rate
Frame Numbers

(c)



(d)



(e)

Fig. 6.17 Similarity of action sequences for TH datasets when training and testing are different (a) Subject 6, (b). Subject 7, (c) Subject 8, (d) Subject 9, (e) Subject 10.

From the results, it was observed that for the testing of similar subjects (group1) the similarity rate (Fig. 14) was higher than the testing of different subjects (group2) (Fig. 15); the testing of similar subjects (group3) the similarity rate (Fig. 16) was higher than the testing of different subjects (group4) (Fig. 17). According to the intrinsic property of the subspace method, the smaller variance of principal component between testing date and training date (reference subspace) can generate higher similarity rate. As each object own different individual characteristics for their shape and actions, in the experiment, the variance of principal component between testing date and training date in testing of using different subjects is larger than testing of using similar subjects.

6.6 Discussion and Conclusions

6.6.1 Discussion

The automatic rescue system was developed with high accuracy and real-time performance in the recognition of danger or inattention of driving actions. Experiments were conducted using KINECT sensor-captured images to confirm the accuracy and reliability of the rescue system based on the self-established driver action dataset. The inattention action can be recognized according to the behaviours and attitudes of the drivers, such as looking aside, talking on the phone, nodding, rubbing eyes and yawning, and the performance of action recognition was satisfactory under regular conditions. In this research, inattention action recognition was determined using RGB-images, and the function of player detection provided by Kinect reduced the complexity of the driver area separation from the background. As the most widely used sensor information, the research was conducted on action recognition using RGB-image, which is obviously significant. An RGB-image is easily affected by the illumination condition, and the IR spectra obtained in natural light can also disturb Kinect for the driver area separation. Thus, weak lighting conditions were ignored, and a tractor with a cab was used to reduce the disturbances in natural light in the experiments.

The challenge in using RGB-images for action recognition was the effect of the different viewpoints and individual characteristics because a conventional subspace-based method cannot represent temporal information well. Such a problem was overcome by modifying the Kernel Mutual Subspace Method to make it possible to capture the temporal relations among the sequential images using the Hankel matrix. In the offline and online experiments, the developed Hankel-based KMSM made it possible to classify and recognize most of the inattention actions and realized high recognition accuracy. However, some of the actions in the subtle motions obtained a low classification rate, owing to similarities with others actions, for example, talking on the phone, yawning or checking the rear-view mirror under normal driving conditions.

The driver action recognition in the subtle motions is difficult to distinguish and action recognition rate was lower. In addition, the training and testing datasets had variations for subjects to subjects. To improve further accuracy of recognition of the inattention actions, it is meaningful to build a larger dataset by collecting more training data of aged farmers under various conditions. In the same time, tractor's cabin-based training and testing on similar subject could have a wider possibility, as the drivers are limited in the farms. In further research, the possibility of using depth-images will also be discussed for driver inattention action recognition under weak lighting conditions or at night.

6.6.2 Conclusions

In conclusion, a vision-based real-time recognition system of driver inattention actions was proposed using the Hankel-based KMSM, where the sequences of images were represented by a nonlinear subspace. By introducing the Hankel, the limitation of the subspaces in the conventional KMSM method only encoding the appearance information of the sequential images was overcome. Furthermore, the temporal information of the sub-actions such as Hankel-based KMSM could describe the dynamic properties of short sub-actions. In addition, compared with other human action recognition approaches, the image-based low-dimensional block Hankel matrix in this research greatly reduced the computational complexity, which meets the requirements for real-time recognition. To evaluate driver inattention, a driver action dataset was established. In addition, the validity of the Hankel-based KMSM method was demonstrated through experiments using the driver inattention action dataset for similar subjects and different subjects both for off-line and on-line experiments. As a result, the off-line classification accuracy rates were 91.18% and 86.18% for similar subjects testing and different subjects testing, respectively for RGB image; and with accuracy of 92.2% and 47.52% for similar subjects testing and different subjects testing, respectively for thermal image. And the on-line classification accuracy rates were 87.02 and 79.97% for similar subjects testing and different subjects testing, respectively RGB image; and with accuracy of 83.24% and 42.75% for similar subjects testing and different subjects testing, respectively thermal image. The average computational time was 0.07 s and 0.08 for RGB image and thermal image, respectively. The results showed that the proposed method could satisfy the real-time and accuracy requirements of the rescue system.

Chapter 7
Conclusions

This thesis aimed to develop a driving rescue system for agriculture vehicles to ensure driving safety. Series of works including system architecture design, road type classify algorithm development, driver inattention actions and fatigue detection recognize algorithm development, road and driver action dataset establishment and field experiments have been performed. The major contributions of this study and future work are drawn as follows;

7.1 Summary of Research Findings

7.1.1 Designed Driving Rescue System Architecture

Systematically analyzed the relationship between driver condition, road condition and driving safety. Three categories of road condition and five types of inattention actions were defined, and MRM was designed regarding the driver status and road condition.

7.1.2 Designed Driving Environment and Driver Status Sensing System

Machine vision system was incorporated for monitoring road condition and driver status to establish a no-contact rescue sensing system. A monocular camera was used for road environmental monitoring and types of road recognition; Kinect device and thermal camera were used for driver status monitoring. Utilizing the Kinect player extract function, a driver can be easily segmented from the image, and the thermal camera ensured the effective driver action monitoring even under low illumination condition.

7.1.3 Established Road Type and Driver Action Datasets

For road type datasets, a dataset includes structure roads and unstructured roads, curve roads, straight road and cross roads under structured roads classifier. On the other hand, curve roads, straight road and cross roads are referred under the classifier of unstructured roads. For driver action datasets, dataset included actions of looking aside; talking on the phone, nodding, rubbing eyes, yawing and normal driving was established.

7.1.4 Recognition of Road Types Classification Algorithm

A KMSM based road type classify algorithm was designed for road classification. Using the established road type dataset, offline and online experiment was conducted to evaluate the effectiveness of the developed algorithm. The off-line classification accuracy rates were 97.7%, 98.1% and 95.4% for structured road and unstructured road, S road, straight road and T road under structured road and unstructured road, respectively. The on-line classification accuracy rates were 100%, 85.5% and 91.55%, for structured road and unstructured road, S road, straight road and T road under structured road and unstructured road, respectively. The average computational time was noted 0.03s for recognizing the each of the classifier.

7.1.5 Development of Driver Action Recognition Algorithm

To overcome the limitations of the conventional subspace method in addressing temporal information, which is important for a driver's inattention action recognition, the Hankel-based KMSM was designed. Offline and online experiments were conducted to evaluate the effectiveness of the developed algorithm. As a result, the off-line classification accuracy rates were 91.18% and 86.18% for similar subjects testing and different subjects testing, respectively for RGB image; and with accuracy of 92.2% and 47.52% for similar subjects testing and different subjects testing, respectively for thermal image. The on-line classification accuracy rates were 87.02 and 79.97% for similar subjects testing and different subjects testing, respectively RGB image; and with accuracy of 83.24% and 42.75% for similar subjects testing and different subjects testing, respectively using thermal image. The average computational time was 0.07 s and 0.08 for RGB image and thermal image, respectively.

7.2 Future Work

To complete the research in this thesis, further attempt will be taken to complete the rescue system for the driver and roads as an integrated unit.

7.2.1 Extend the Road Type Datasets and Driver Action Datasets

Unstructured road under agricultural environment covers large area in Japan. In our further research, we will collect most of the types of road structured in rural and semi urban areas. As the different area, due to the landform and different crop production and operation style, the road image shows different features. As the same reason, driving action form additional operators is also required. Because the different persons have different features and action characteristics, in our future research, we will collect more data in different volunteers to for increasing training datasets of driver action, which will help in robust recognition of the accuracy system.

7.2.2 Robust Rescue System Development

The rescue system developed in this research had not considered disturbances such as week illumination, backlighting, special temperature, and drive sub motions. From the robust point of view the secure system should own performance to cover those conditions. For this object, along with extended datasets, other Artificial Intelligence (AI) and deep learning solutions are also considered to introduce with the subspace method.

Acknowledgements

First, I would like to express my profound gratitude and sincere appreciation to my supervisor, Associate Professor Dr. Tofael Ahamed, for his availability, expertise and advice, motivation, and encouragement to move forward. It is my great pleasure to express deepest sense of gratitude, sincere appreciation to Professor Dr. Tomohiro Takigawa and Associate Professor Dr. Noguchi Ryozo, their kind advice, scholastic guidance, and encouragement guided me to do the research. I would like to express my gratitude to Assistant Professor Takuma Genkawa, Professor Shusuke Matsushita and Professor Kota Motobayashi, members of the advisory and thesis evaluation committee for their kind comments, advice, and suggestions during the course of this study.

Special gratitude is expressed to all lab members who give me grate help and support during the doctor course. In addition, special thanks should be expressed staffs of Agricultural and Forestry Research Center and to all volunteers who help me to collect driver experiment data for my papers and this thesis. Sincere gratitude and thanks to my husband Dr. Linhuan Zhang, who given useful discussion and data collection for the research.

Finally, Sincere gratitude and thanks to my parents Zhang Xuegen and Han Yuhua, parents-in-law Zhang Bingfan and Weng Yueqing, and my two sons Zhang Jiahang and Zhang Yihang for their moral support and encouragements.

Reference

Ahamed T, Tian L, Takigawa T, et al. Development of auto-hitching navigation system for farm implements using laser range finder[J]. Transactions of the ASABE, 2009, 52(5): 1793-1803.

Ahamed, T., Tomohiro, T., Noguchi, R., &Lei, T. (2014). A Road Map for Sustainable Agricultural Practices, Bioproduction Engineering, Nova Science Publishers, New York, USA, Inc, ISBN: 978-1-62618-122-9, 246.

Chaudhry R, Ofli F, Kurillo G, et al. Bio-inspired dynamic 3d discriminative skeletal features for human action recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. 2013: 471-478.

Craye C, Karray F. Driver distraction detection and recognition using RGB-D sensor[J]. arXiv preprint arXiv:1502.00250, 2015.

Coronel E, Pojomovsky A, Gaona F. Reliable navigation-path extraction system for an autonomous mobile vehicle[C]//Digital Information Management (ICDIM), 2015 Tenth International Conference on. IEEE, 2015: 175-181.

Cyganek B, Gruszczyński S. Hybrid computer vision system for drivers' eye recognition and fatigue monitoring[J]. Neurocomputing, 2014, 126: 78-94.

Fukui K, Yamaguchi O. Face recognition using multi-viewpoint patterns for robot vision[C]//Robotics Research. The Eleventh International Symposium. Springer, Berlin, Heidelberg, 2005: 192-201.

Gonzalez R, Iagnemma K. DeepTerramechanics: Terrain Classification and Slip Estimation for Ground Robots via Deep Learning[J]. arXiv preprint arXiv:1806.07379, 2018.

Hadsell R, Sermanet P, Ben J, et al. Learning long - range vision for autonomous off - road driving[J]. Journal of Field Robotics, 2009, 26(2): 120-144.

Han J, Kim D, Lee M, et al. Road boundary detection and tracking for structured and unstructured roads using a 2D lidar sensor[J]. International Journal of Automotive Technology, 2014, 15(4): 611-623.

Jung S J, Shin H S, Chung W Y. Driver fatigue and drowsiness monitoring system with embedded electrocardiogram sensor on steering wheel[J]. IET Intelligent Transport Systems, 2014, 8(1): 43-50.

Kobayashi T, Otsu N. Motion recognition using local auto-correlation of space–time gradients[J]. Pattern Recognition Letters, 2012, 33(9): 1188-1195.

Laddha A, Kocamaz M K, Navarro-Serment L E, et al. Map-supervised road detection[C]//Intelligent Vehicles Symposium (IV), 2016 IEEE. IEEE, 2016: 118-123.

Lee J D, Li J D, Liu L C, et al. A novel driving pattern recognition and status monitoring system[C]//Pacific-Rim Symposium on Image and Video Technology. Springer, Berlin, Heidelberg, 2006: 504-512.

Li B, Ayazoglu M, Mao T, et al. Activity recognition using dynamic subspace angles[C]//Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on. IEEE, 2011: 3193-3200.

Li, G., & Chung, W. Y. Detection of Driver Drowsiness Using Wavelet Analysis of Heart Rate Variability and a Support Vector Machine Classifier. Sensors 2013, 13(12), 16494-16511

Li W, Zhang Z, Liu Z. Action recognition based on a bag of 3d points[C]//Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer

Society Conference on. IEEE, 2010: 9-14.

Lundahl K, Olofsson B, Berntorp K, et al. Towards lane-keeping electronic stability control for road-vehicles[C]//19th IFAC World Congress, August 24-28, Cape Town, South Africa. International Federation of Automatic Control, 2014, 19: 6319-6325.

Maeda, K. I., & Watanabe, S. A pattern matching method with local structure. Trans. IEICE, (1985), 68(3), 345-352.

Masala G L, Grosso E. Real time detection of driver attention: Emerging solutions based on robust iconic classifiers and dictionary of poses[J]. Transportation research part C: emerging technologies, 2014, 49: 32-42.

Matthew, B., Joseph, F., Stephen, E., & J. Christian Gerdes. Safe driving envelopes for path tracking in autonomous vehicles. Control Engineering Practice, (2016). Volume 61, 307-316.

Moonen M, De Moor B, Vandenberghe L, et al. On-and off-line identification of linear state-space models[J]. International Journal of Control, 1989, 49(1): 219-232.

Nilsson K, Pinzke S, Lundqvist P. Occupational injuries to senior farmers in Sweden[J]. Journal of agricultural safety and health, 2010, 16(1): 19-29.

Oreifej O, Liu Z. Hon4d: Histogram of oriented 4d normals for activity recognition from depth sequences[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2013: 716-723.

Reina G, Milella A, Worst R. LIDAR and stereo combination for traversability assessment of off-road robotic vehicles[J]. Robotica, 2016, 34(12): 2823-2841.

Sakano, H., Mukawa, N., & Nakamura, T. Kernel mutual subspace method and its application for object recognition. Electronics and Communications in Japan (Part

II: Electronics), 2005, 88(6), 45-53.

Sigari, M. H., Pourshahabi, M. R., Soryani, M., & Fathy, M. A Review on Driver Face Monitoring Systems for Fatigue and Distraction Detection. International Journal of Advanced Science and Technology. (2014) Vol. 64, 73-100.

Sun X, Cai Y, Chen L, et al. Vehicle height and posture control of the electronic air suspension system using the hybrid system approach[J]. Vehicle System Dynamics, 2016, 54(3): 328-352.

Tawari A, Trivedi M M. Robust and continuous estimation of driver gaze zone by dynamic analysis of multiple face videos[C]//Intelligent Vehicles Symposium Proceedings, 2014 IEEE. IEEE, 2014: 344-349.

Van Overschee P, De Moor B. N4SID: Subspace algorithms for the identification of combined deterministic-stochastic systems[J]. Automatica, 1994, 30(1): 75-93.

Wof L, Shashua A. Kernel principal angles for classification machines with applications to image sequence interpretation[C]//Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on. IEEE, 2003, 1: I-I.

Wang H, Yuan C, Luo G, et al. Action recognition using linear dynamic systems[J]. Pattern Recognition, 2013, 46(6): 1710-1718.

Yamato J, Ohya J, Ishii K. Recognizing human action in time-sequential images using hidden markov model[C]//Computer Vision and Pattern Recognition, 1992. Proceedings CVPR'92., 1992 IEEE Computer Society Conference on. IEEE, 1992: 379-385.

Zhang F, Su J, Geng L, et al. Driver fatigue detection based on eye state recognition[C]//Machine Vision and Information Technology (CMVIT), International Conference on. IEEE, 2017: 105-110.

Zhang L, Ahamed T, Zhang Y, et al. Vision-based leader vehicle trajectory tracking for multiple agricultural vehicles[J]. Sensors, 2016, 16(4): 578.

Zhang, Y., Suryanto, C. H., &Fukui, K. Hand Shape Recognition using Position Invariant Feature from the Sequence of Depth Images. Forum on Information Technology 13(3), Japan (2014), 289-291.

Zhu Y, Luo K, Ma C, et al. Superpixel Segmentation Based Synthetic Classifications with Clear Boundary Information for a Legged Robot[J]. Sensors, 2018, 18(9): 2808.