Research Article

Development of a Real-Time Detection System for Augmented Reality Driving

Kuei-Shu Hsu,[1] Chia-Sui Wang,[2] Jinn-Feng Jiang,[3] and Hung-Yuan Wei[3]

[1]*Department of Applied Geoinformatics, Chia Nan University of Pharmacy & Science, No. 60, Section 1, Erren Road,
Rende District, Tainan 71710, Taiwan*

[2]*Department of Information Management, Chia Nan University of Pharmacy & Science, Tainan 717, Taiwan*

[3]*Project Coordination and Liaison Section, Planning and Promotion Department, Metal Industries Research and Development Center,
Taipei, Taiwan*

Correspondence should be addressed to Kuei-Shu Hsu; kshsu888@gmail.com

Augmented reality technology is applied so that driving tests may be performed in various environments using a virtual reality
scenario with the ultimate goal of improving visual and interactive effects of simulated drivers. Environmental conditions simulating
a real scenario are created using an augmented reality structure, which guarantees the test taker's security since they are not subject
to real-life elements and dangers. Furthermore, the accuracy of tests conducted through virtual reality is not influenced by either
environmental or human factors. Driver posture is captured in real time using Kinect's depth perception function and then applied
to driving simulation effects that are emulated by Unity3D's gaming technology. Subsequently, different driving models may be
collected through different drivers. In this research, nearly true and realistic street environments are simulated to evaluate driver
behavior. A variety of different visual effects are easily available to effectively reduce error rates, thereby significantly improving
test security as well as the reliability and reality of this project. Different situation designs are simulated and evaluated to increase
development efficiency and build more security verification test platforms using such technology in conjunction with driving tests,
vehicle fittings, environmental factors, and so forth.

1. Introduction

The application of virtual reality to games that are available
on computers or mobile phones is currently a very popular
topic. Multiple cameras may be utilized to observe a user as
that user simulates driving a vehicle. Moreover, to analyze
the conditions of operating a vehicle, situation simulations
(in addition to environmental factors and external variable
conditions) may be employed to test the user's response,
action changes, and so forth. Such simulations may even be
a criterion for developing vehicle equipment and parts. User
habits are obtainable through the direct operation of virtual
reality.

The subject of this study is to apply a stereo vision
system to a virtual reality system. A computer's vision system
may be divided roughly into plane vision and stereo vision,
which are different in their capabilities to estimate the depth

information of target objects in an image (also referred to as
"depth perception"); this division makes a computer vision
system more practical. Some of its many practical applica-
tions include remote sensing and monitoring, medical image
processing, robot vision systems, military reconnaissance,
mineral exploration, and cartography.

In recent years, computer vision systems have made
big advances in both the theory and algorithms of image
processing as well as in practical applications. For example, a
3D stereo object in space may be captured as a 2D plane image
by a charge-coupled device camera through image processing
methods—such as image transforms, image enhancement,
thresholding, filtering, edge detection, and thinning [1, 2].
That object may then be used as image information for appli-
cations of feature extraction, recognition, visual servicing,
and so forth [2–4]. Furthermore, advanced techniques have
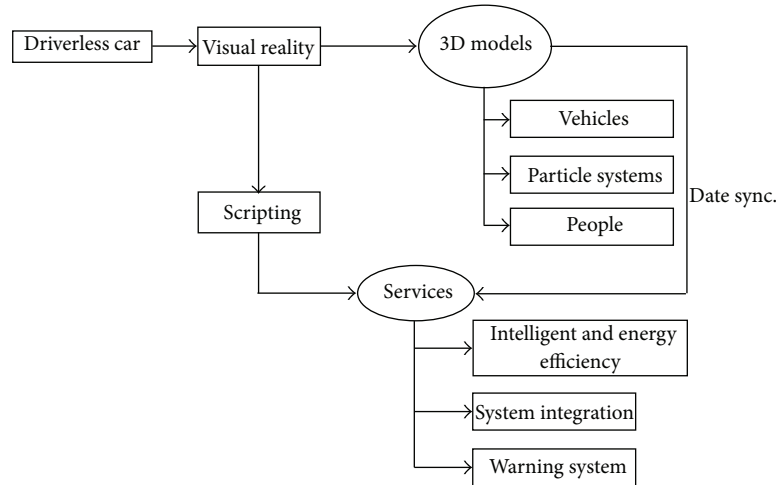led to the development of 3D stereo vision systems. However,

FIGURE 1: Research flowchart.

there are still many core technologies to be achieved before computer vision can be compared to human vision [5, 6].

To perform tests in different environments using virtual reality scenarios and create a realistic feel, environmental conditions that simulate real scenarios must be created. Therefore, we employ Unity software to create virtual reality systems and then use those systems in conjunction with various scenarios. Our systems may subsequently be promoted for use to smart phone platforms. Using virtual reality for simulations is not only safe but also convenient. Furthermore, testing accuracy is not affected by environmental factors that may be induced by tests in the real world.

2. Research Method

This research uses streetscape, which is commonly used in driving, as the scenario. SketchUp software is used to model and map an entire city. After the model is drawn, it is imported to Unity to create a virtual reality interface. Other representative scenarios for testing will be created after discussion of this research to increase the accuracy and sufficiency of multidimensional data collection.

Figure 1 shows a picture an "augmented reality driver test system" is constructed. The functions of this platform include both virtual reality and augmented reality, making it one of the most complex human-computer cooperation machine systems. In the system, the interactive motion simulator allows the behavioral model of the driver to enter the virtual scenario and operate the vehicle with a steering wheel. Active and augmented reality technologies are integrated, and two different interactive models of augmented reality interactions are proposed to invite the driver to perform a practicality test for different emergency conditions. Several streetscape 3D models are established for the project.

Therefore, the project aims to provide unique, prior-to-development test operations for a driver behavior test. First, the driver is allowed to perform driving behavior analysis in an emulated environment for driver behavior testing that may also be used as a test platform for vehicle products. Second,

TABLE 1: Abstract table of illustrative statistics for different background variable items.

Background variable item	Group	Number	Percentage
Gender	Male	100	50.0%
	Female	100	50.0%
Major use purpose	Goal	30	30.0%
	Entertainment	37	37.0%
	Development	33	33.0%

virtual reality is used to derive new teaching and learning strategies; such a learning model may be applied although the operator has no prior computer experience. By allowing learners to be immersed in learning content through the use of virtual reality, drivers are allowed to apply information from the test to product development (see Table 1).

3. Real-Time Detection of Drivers Using Augmented Reality

3.1. Obtaining Gesture Trajectory Information. Kinect separates human hands from the background environment as a division strategy. OpenNI provides the API interface, that is, a multilinguistic and cross platform framework that is capable of communication with external hardware. Additionally, OpenNI includes an intermediate space for processing gesture recognition, tracking, and so forth. Currently, there are four types of hardware supported by OpenNI: 3DSensor, RGBCamera, IRCamera, and Audio Device.

Figure 2 shows a picture that was acquired with a Kinect sensor. First, the skin color is detected; then, the hands are deeply subdivided for separation into 2D and 3D. The hand model is created in 3D, including an appropriately arranged and assembled geometric primitive. Every hand gesture is represented by 27 vector points. With respect to estimation of the hand joints, 27 vector points are specified and tracked to
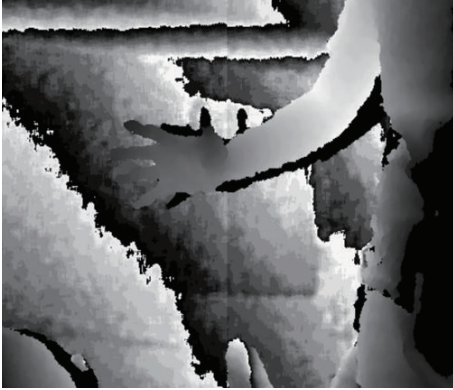
FIGURE 2: Schematic view for Kinect gesture tracking.

FIGURE 3: Schematic view of Kinect facial recognition.

minimize different hypotheses between hands and real observation of model parameters. To quantify such differences, the graphic rendering technology used generates similar skin and constructs depth mapping for given hands. Therefore, an appropriate objective function and a variance of PSO are both defined to obtain the optimal hand configuration. To output the result of such optimization, the hand joints are tracked with time continuity.

3.2. Facial Feature Analysis. Real-time detection is also applied to facial recognition using Kinect's built-in SDK to create a facial recognition structure. The subsequent application incorporates this facial recognition structure into Unity to create an interactive function between the user and the virtual system (as shown in Figure 3).

3.3. Eye Recognition Analysis. Eye recognition generally includes preprocessing, feature extraction, and sample learning and recognition. Many technologies and methods can be used for eye recognition, including the following.

  (i) Projection: here, eye location is detected according to the projection distribution feature of a picture in a certain direction. Projection is a statistic method that utilizes the eye's gray-level information to detect the ordinate and abscissa of the pupil through horizontal and vertical projections to obtain an accurate position of human eyes. The most common projection functions are the integral, mean, and variance projection functions.

  (ii) Hough transformation: Hough transformation transforms pictures from a spatial domain to a parameter space. Certain parameter forms satisfied by most boundary points are used to illustrate a curve in a picture. The pupil is regarded as a standard circle. We may locate the eye's pupil accurately using Hough transformation through the standard equation of a circle. The robustness of Hough transformation is enhanced considerably because of significant geometric analyticity: $(x - a)^2 + (y - b)^2 = r^2$.

  (iii) AdaBoost classification: in the machine-learning field, the AdaBoost algorithm is a high-efficiency, iterative computing algorithm that trains different weak classifiers for the same training set, followed by gathering these weak classifiers to construct a strong classifier. This algorithm is advantageous of its higher classification accuracy and rapid human eye recognition. However, the effectiveness of the algorithm depends on the selection of weak classifiers. It is very important for the application of rapid human eye detection.

  (iv) Sample matching: the location of the pupil may be searched for dynamically from left to right and from up to down in a different picture window using a circle sample according to pupil shape. Sample matching searches for a small picture of interest within a larger picture. It identifies the location of the target with the most similar location as the matching point by calculating the similarity of the sample and matching zone. A sample matching algorithm exists within the scope of the machine-learning field and is an effective eye recognition algorithm.

3.4. Gesture Trajectory Creation. The gesture trajectory recognition used in this study is based on the hidden Markov mode (HMM).

HMM Model Definition. HMM is a very mature technology that matches time-variant data. It is a random state machine. In a gesture recognition system, the observations of the vector sequence and coded symbolic sequence are jointly called the "observation sequence" (designated "O"), which is a random sequence. An HMM with N states (designated as $S_1, S_2, \ldots, S_N$) is represented by a triple parameter, $\lambda = \{\pi, A, B\}$, wherein

  (i) $\pi = [\pi_1, \pi_2, \ldots, \pi_N]$ is the initial distribution for describing the probability distribution of observation
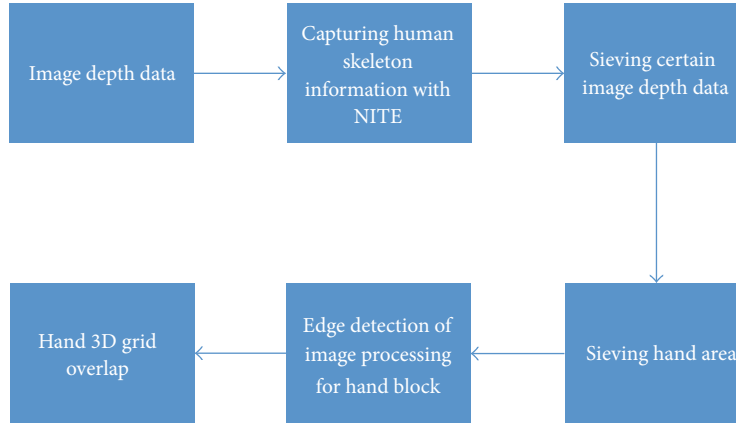
FIGURE 4: Kinect hand recognition flow.

TABLE 2: Single factor variance analysis of individual aspect and entire aspect for different major use purposes with respect to the evaluation test system of improvement for vehicle virtual reality.

Aspect	Variation source	SS	df	MS	F
Goal motivation	Intergroup	8.97	2	4.48	0.57
	Intragroup	769.21	97	7.93	
	Sum	**778.18**	**99**		
Entertainment motivation	Intergroup	216.84	2	108.42	10.65***
	Intragroup	987.46	97	10.18	
	Sum	**1204.33**	**99**		
Entire aspect	Intergroup	317.36	2	158.68	2.67
	Intragroup	1193.1	97	59.37	
	Sum	**1510.46**	**99**		

*** $p < 0.001$.

sequence O under the state at T_i and $\pi_i = P(q_i = S_i)$, $i = 1, 2, \ldots, N$, which satisfies $\Sigma N_i = 1 \pi_i = 1$;

(ii) $A = \{a_{ij} \mid i, j = 1, 2, \ldots, N\}$ is a state transition probability matrix; that is, $a_{ij} = P(q_i = S_j \mid q_{i-1} = S_i, q_{i-2} = S_k, \ldots) = P(q_i = S_j \mid q_{i-1} = S_i)$, which satisfies $\Sigma N_i = 1 a_{ij} = 1$;

(iii) B is a state output probability matrix, which indicates the distribution of random variables or random vectors in the observation probability space of various states and $B = \{b_j(k), j = 1, 2, \ldots, N\}, k = 1, 2, \ldots, M$, which satisfies $\Sigma N_i = 1 b_j, k = 1$.

The dynamic gesture trajectory feature selects and extracts dynamic gestures that are different from static gestures. For static gestures, the change is only on a spatial location and morphology, whereas, for dynamic gestures, the change relates to both time and space. The experiments show that human motion may be recognized only by motion information; therefore, gestures may be transformed into spatial motion trajectories for recognition.

The basic features of a gesture trajectory are location, speed, and angle. Speed is not practical for our use because there can be large differences in speed for the same gestures. The coordinate locations of points on a gesture trajectory curve may be significant features for gesture trajectory recognition; however, differences are still present even for the same gesture trajectories. Thus, there is a change in coordinate points. The change in the tangential angles of a gesture trajectory at different times is used as a trajectory feature. With Kinect, the sensor may acquire palm locations

$$\varphi^t = \tan^{-1}\left(\frac{y_c^t - y_c^{t-1}}{x_c^t - x_c^{t-1}}\right). \tag{1}$$

According to the obtained angle, a 12-direction chain code is used for discretization (see Figure 4).

We then use the Unity software to create a virtual reality system in conjunction with using various scenarios to build virtual reality streets and scenarios. Real vehicle operation is simulated through virtual reality, and environmental (capability of simulating special weather, such as rain and snow) and human (e.g., a pedestrian crossing the road suddenly, appearance of an automobile or a motorcycle from a dead sector, or appearance of roadblock) variables are used as criteria for building the system. Furthermore, it is built according to potential sudden conditions in vehicle traveling in conjunction with human factor engineering design (see Table 2).

After the system is built, Unity3D applies the virtual reality scenario to augmented reality in combination with driver detection. The driver operation model is available from

Figure 5: Detection of the driver's real-time posture.

Figure 6: Virtual reality scene.

the driving process of a scenario test, and the driver's posture and gaze angle may also be used as test criteria.

Presently, people's requirements of visual enjoyment are increasing with the development of 3D stereo vision (as shown in Figures 5 and 6). For people to perform the test in different environments using virtual reality scenarios and have a realistic feeling, environmental conditions simulating real scenarios may be created. Virtual reality simulation is not only safe but also convenient. Furthermore, the accuracy of tests conducted with virtual reality is not affected by environmental factors that may be induced by tests in the real world.

4. Conclusion

Streetscape is the main scenario of this study. 3D drawing software is employed for modeling and mapping to draw the entire street and for building models. After all building models are drawn, they are imported to Unity to create a virtual reality interface. For greater accuracy and multidimensional data collection, a realistic streetscape is developed, complete with a variety of sudden conditions involving vehicles, pedestrians, roadblocks, and animals. Kinect is used to perform augmented reality effects and improve the system design for feedback regarding the operating behavior of the driver. New 3D technology and program design are utilized to construct lifelike road environments and traffic characteristics, including dynamic traffic flow and cross signal control. Planned future work for additional application is to research the interactions between people, vehicles, and roads; for example, suitable vehicles, road systems, and situations will be studied. Furthermore, both software and hardware will be updated to benefit from the increasing reality of simulated environments and for future extensibility of the driving simulation system.

In the project, a real streetscape is simulated, followed by an evaluation of driver behavior and vehicular products. A variety of different visual effects are easily available to effectively reduce error rates. Moreover, the reliability and reality of this project may be improved and the test security may be increased considerably. Different situation designs are simulated and evaluated to increase development efficiency and add more security verification test platforms using such technology in conjunction with driving tests, vehicle fittings, environment factors, and so forth.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

[1] G. Tsechpenakis, K. Rapantzikos, N. Tsapatsoulis, and S. Kollias, "A snake model for object tracking in natural sequences," *Signal Processing: Image Communication*, vol. 19, no. 3, pp. 219–238, 2004.

[2] W. Admiraal, J. Huizenga, S. Akkerman, and G. T. Dam, "The concept of flow in collaborative game-based learning," *Computers in Human Behavior*, vol. 27, no. 3, pp. 1185–1194, 2011.

[3] L. M. Bergasa, P. F. Alcantarilla, and D. Schleicher, "Non-linearity analysis of depth and angular indexes for optimal stereo SLAM," *Sensors*, vol. 10, no. 4, pp. 4159–4179, 2010.

[4] A. Carleton-Hug and J. W. Hug, "Challenges and opportunities for evaluating environmental education programs," *Evaluation and Program Planning*, vol. 33, no. 2, pp. 159–164, 2010.

[5] G. Klančar, M. Kristan, and R. Karba, "Wide-angle camera distortions and non-uniform illumination in mobile robot tracking," *Robotics and Autonomous Systems*, vol. 46, no. 2, pp. 125–133, 2004.

[6] P. KaewTrakulPong and R. Bowden, "A real time adaptive visual surveillance system for tracking low-resolution colour targets in dynamically changing scenes," *Image and Vision Computing*, vol. 21, no. 10, pp. 913–929, 2003.