

Research Article

Tourist Street View Navigation and Tourist Positioning Based on Multimodal Wireless Virtual Reality

Shousheng Chai,¹ Jun Shan ,¹ and Yi Hui Chan²

¹School of Management, Ocean University of China, Qingdao, China

²Sunway University, Jalan Austin Heights Utama, Taman Mount Austin, Johor 81800, Malaysia

Correspondence should be addressed to Jun Shan; shanjun564@163.com

Received 28 September 2021; Accepted 23 October 2021; Published 16 November 2021

Academic Editor: Guolong Shi

Copyright © 2021 Shousheng Chai et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Based on the multimodal wireless virtual reality technology, this article researches and implements two different spatial positioning estimation systems based on intelligent networks and mixed reality devices, which can estimate their position and attitude in space with higher accuracy. Based on the multimodal protocol, an extensible information exchange protocol is designed to meet the requirements of the terminal and the background service in the scenic spot, which realizes the information interaction between the terminal and the background service and supports the realization of system functions. Second, a real-time data management module is designed for the two types of terminal historical location query and real-time location query. Through the research and summary of GPS positioning technology and existing matching algorithms, as well as the analysis of positioning errors, a scenic map matching algorithm based on an improved virtual reality model is proposed. Later, in the electronic map design, the traditional linear street view model was improved, the width element of the street view was increased, and the rectangular street view model was introduced. When the candidate road section was selected, the candidate road section was determined by the area intersection method of the error rectangle and the rectangular street view, and the search path was improved, while reducing omissions in the selection of candidate road sections. In the process of adaptive moving target feature learning, it can effectively remove the inaccurate foreground extraction caused by the misjudgement and misdetection of the multimodal background modeling method. This plays an important role in the later trajectory tracking and reduces the uninterruptedness. The simulation results show that the real-time data is cached in the memory, and the real-time data is divided and organized on the basis of region, which realizes efficient designated terminal location query, range query, neighbour query, and many other typical real-time location-based queries. Finally, using the characteristics of the database management system adopted, through a reasonable design of the database table structure, efficient access to historical data is realized. System test results show that the system has realized multimedia applications, electronic maps, travel notes, and other functions and has a friendly interface, easy operation, and good maintainability and scalability.

1. Introduction

Tourism, as an important growth point of the country's economy and an important way of people's leisure life, has developed rapidly in recent years and has a broad market prospect. However, due to the rapid increase in the number of tourists, the improvement of tourist requirements for service levels, and the sustainable development of scenic spots, the information level puts forward higher requirements [1]. The location service formed by the integration of navigation

and positioning, mobile interconnection, cloud computing, and other technologies is applied to the tourism industry (smart tourism), which will significantly improve the level of tourist services, refined management, and emergency handling in scenic spots [2]. Since mobile smart devices need to locate themselves before use to further perform path planning based on the data collected by sensors and other tasks that require high positioning accuracy, the research needs for spatial positioning technology in smart device applications are also increasing [3].

At present, the positioning characteristics of the navigation system in scenic spots are mainly manifested in the slow and irregular movement of users, the complex environment of scenic spots, and the susceptibility to GPS signals of smartphones [4]. Commonly used matching algorithms are mainly aimed at vehicle data models, and this article applies matching algorithms to pedestrian positioning data in scenic spots. Therefore, the research of map matching technology based on scenic spot navigation has certain significance and value [5]. Among these advanced technologies, location service (LBS) has undoubtedly become the preferred choice of scenic area managers, because LBS is an effective combination of modern technologies such as mobile communication technology, Internet technology, and geographic information system for effectively link people, vehicles, and infrastructure in the scenic area to achieve the effect of joint adjustment and linkage [6]. At present, the commonly used visual positioning system can use the visual odometer to estimate the motion trajectory of the carrier and restore the spatial structure of the scene according to the adjacent image frames obtained by the camera, but the use of the visual odometer is only to estimate the camera by calculating the image changes at adjacent moments. The vision sensor itself is susceptible to equipment performance, environmental lighting changes, and equipment movement speed, so when using the visual positioning system to locate the equipment carrier, there will inevitably be a cumulative drift of positioning [7–9]. At the same time, with the help of LBS, the staff of the scenic spot can obtain the location information of tourists in time, dig out tourists' hobbies, provide personalized recommendation services for tourists, and provide timely rescue services for tourists in distress. For the same reason, tourists can also obtain themselves through LBS or companion positioning information, finding route information of scenic spots, and obtaining tourism resources and other services, greatly facilitating the travel needs of tourists [10].

This paper proposes a method of coupling visual multimodality and inertial measurement unit based on virtual reality algorithm to realize a spatial positioning system for tourists in scenic spots. The system can improve the accuracy of the device's attitude estimation during operation by coupling the device positioning and motion data obtained by two different sensors. The fusion application of indoor positioning technologies of different precisions realizes the coordinated utilization of multiprecision cross-device and cross-platform positioning data resources and improves the adaptability and stability of the indoor positioning system. Combining navigation and positioning, short message communication, pseudolite enhancement, CORS high-precision positioning, 3G mobile communications, intelligent location services, and other technologies, designed and developed three applications for scenic location services, scenic area management and search and rescue scenic area management system are given. Based on the C/S architecture, the system has developed server-side software, which solves key technologies such as data exchange protocol design, real-time data management, and efficient data storage and satisfies the real-time location query, group location query, and historical location of 50,000 terminals, personalized information push, and communication access and other functions.

2. Related Work

For the currently evolving field of equipment automation, in order to enable accurate operation of equipment, efficient and accurate positioning technology is required to provide machine positioning information to the control system during the working process of the machine. With the continuous development of equipment in the direction of intelligence and automation, many fields have begun to apply autonomous robots to complex, high-risk, and highly repetitive tasks. Environmental data collection and autonomous positioning of smart devices are currently indispensable practical technology and research focus [11–13].

Blum et al. [14] introduced the AGV storage robot characterized by autonomous navigation and multisensor cooperation. By using lidar, inertial navigation sensors, and odometers for environmental data collection and motion data measurement, single-sensor operation is eliminated by multisensor cooperation. The resulting information is incomplete, the accuracy of the sensor hardware is limited, and the probability density function is used to represent the movement state of the robot to realize the global positioning of the warehouse robot. At present, there is also a method of assisting equipment to complete positioning by setting recognizable marks at specific locations. Namoun et al. [15] explained the principles and applications of QR code positioning technology. With the characteristics of high storage capacity and high reliability of the code, only the two-dimensional code identification is acquired through the visual sensor, and the positioning system can calculate the position and direction of the device according to the position coordinate information stored in the two-dimensional code, realizing efficient and fast positioning. Robben et al. [16] integrated direction, distance, and spatial correlation and proposed an improved topological relationship map matching method. The projection distance between the positioning points and the road section, the positioning track direction and the street view direction, and the street view correlation are calculated by the factor weighting method. Afterward, many research scholars added street view turning restrictions and heading changes as new weights to the matching process. In order to improve the real-time problem of matching, Willing et al. [17] proposed a dynamic weight real-time map matching algorithm, which uses the road network topology and GPS trajectory points to establish a multielement dynamic weight model and calculates the matching road section according to the confidence level.

Compared with geometric map matching algorithms, topological structure-based map matching algorithms are more accurate and faster, but there are still problems in complex road network conditions. The geometric matching algorithm proposed by Del et al. [18] is the first prototype, including point-to-point matching, point-to-arc matching, and arc-to-arc matching. Since the geometric map-based matching method only considers the geometric shape of the street view, the implementation is simple, but the matching accuracy is not high. Some scholars have proposed an improved algorithm for map matching based on location

points. The positioning points are projected to candidate road sections, taking into account the continuity of user movement, the projection distance, and the moving direction, and the x -axis sum is performed on the projection points on each candidate road section. The y -axis analysis can quickly match the positioning point to the correct road section. Some scholars have proposed a matching method that considers the tangent of the curve. On the basis of point-to-arc matching, the tangent distance between the point and the arc and the deflection angle of the arc street view are considered to improve the positioning error of GPS [19]. Some scholars have proposed a map matching algorithm based on trajectory similarity, which quantifies the position and direction of the trajectory, analyzes the shape of the trajectory and street scene, uses fuzzy logic inference method to find the matching road section for the first positioning point, and then tracks it. In the matching mode, another fuzzy logic reasoning method is used to judge whether the matching road section of the current location point is the last matching road section and considering the constraint conditions of vehicle speed and direction, the wireless virtual reality model is used to achieve curve-to-curve map matching [20, 21].

3. Construction of Tourist Street View Navigation and Tourist Positioning Model Based on Multimodal Wireless Virtual Reality

3.1. Multimodal Network Hierarchy. As an independent conceptual design, multimodal network equipment combines the advantages of virtual reality technology and augmented reality technology, providing a new development direction for the interaction between users, computers, and the environment. Mixed reality technology is gradually realized with the advancement of computer vision, graphics and image processing technology, smart device display technology, and multidimensional data input systems. Figure 1 shows the multimodal network hierarchy topology.

Location service module: it mainly conducts two-way information exchange with the user's intelligent terminal and connects to the network with each subsystem to divide, store, process, and distribute information. Various location information provided include map service, route planning, location query, and location monitoring.

$$I = \{C \in (x, y), x \in (0, 1, 2, \dots, n), y \in (0, 1, 2, \dots, n-1), n \in N\} \\ I = \{C \in (x, y), x \in (0, 1, 2, \dots, n), y \in (0, 1, 2, \dots, n-1), n \in N\}, \quad (1)$$

$$L(x, y) = \sqrt{\sum (C_a(x, y) - C_b(x, y))^2}. \quad (2)$$

The positioning system includes satellite navigation, differential enhancement, pseudolite technology, IRID technology, mobile network technology positioning, and other methods. The spatial positioning technology is used to obtain the location information (latitude, longitude, speed, and time) of the mobile terminal, which is the key to the

realization of the entire system. In realizing applications, multiple technologies should be combined to ensure the continuity of positioning, as well as the accuracy, availability, and integrity of positioning.

$$G(x, y) - \iint I(x, y) \times g(x-1, y-1) dx dy = 0. \quad (3)$$

Mobile Internet: the link between users and the location service module, it must be able to effectively transmit user requests and respond to the processing results of the location service module in real time. Based on the current 3G or 4G communication, mobile value-added services with location-based services can be developed.

$$d^2(x, y) - (x - y)^T(x - y) = 0, \quad (4)$$

$$H(x) = \begin{cases} 0, & x > t, \\ \text{sign} \sum_{t=1}^T a(t) \times g(x, t), & x < t. \end{cases} \quad (5)$$

User intelligent terminal: terminal equipment used by users, such as smart phones and walkie-talkies, can be used as terminal equipment. These terminals must have excellent graphics effects, a smooth communication interface, a reasonable user display interface, and convenient input methods. Therefore, PDAs and certain types of smart phones will become the first choice for personal terminals.

$$g(x) = \sum_{i,j=1}^n \frac{[s(1, i) + s(2, i) + \dots + s(j, i)]}{s(i, j)} = 0. \quad (6)$$

Usually, the positioning error and is covariance generated during the operation of the equipment are calculated by the arithmetic difference between the machine motion state vector and the calculated positioning estimate. However, there are constraints in the algorithm when using the Kalman filter algorithm, so a quaternion error needs to be introduced in the calculation to represent the error between the estimated value of the algorithm and the positioning in the actual coordinate system.

3.2. Virtual Reality Algorithm Framework. Virtual reality technology is an indispensable part of current simulation technology. It is the integration of simulation technology, computer graphics, and sensing technology. It enables users to perceive the virtual reality generated by virtual reality technology through vision, hearing, and even touch environment, and users can interact with virtual reality applications through sensing devices. Mixed reality can combine reality with virtual environments, allowing users to directly interact with virtual applications using gestures, voice, or perspective in an environment with an immersive experience. It can be regarded as a combination of virtual reality technology and augmented reality technology with independent technical concept of advantages. The single use of satellite navigation and positioning can achieve a positioning accuracy of about 10 meters. If higher accuracy (decimeter level and

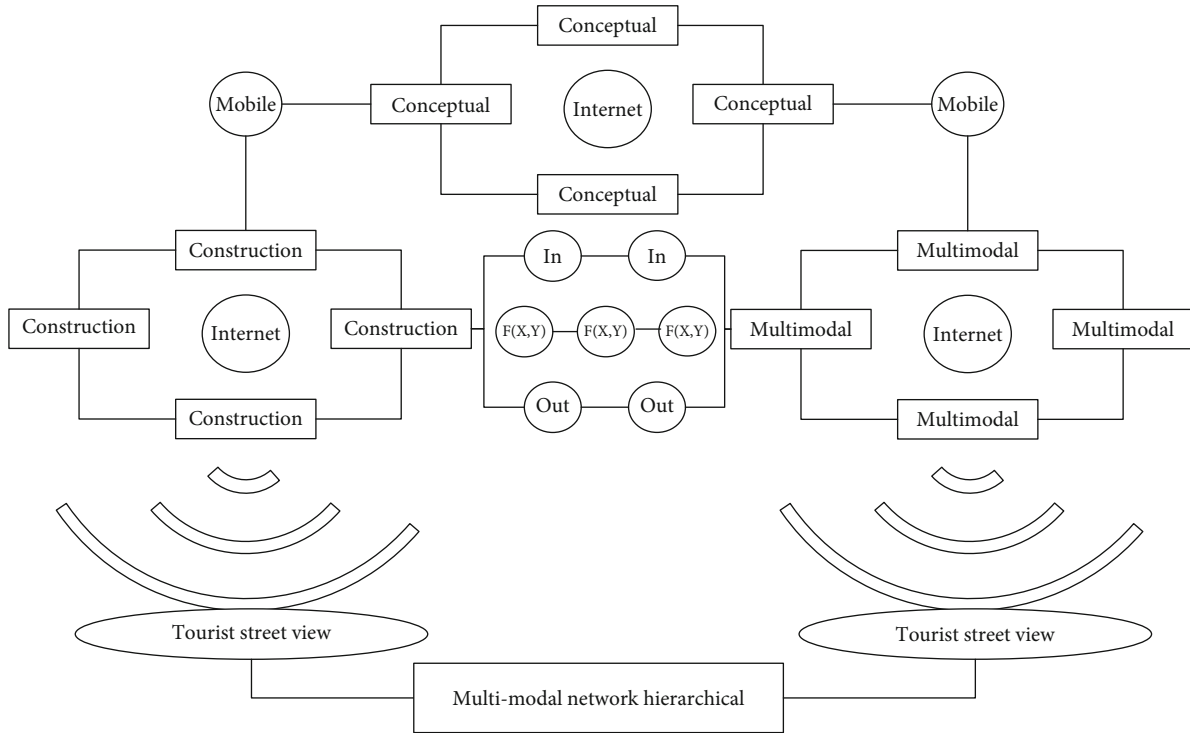


FIGURE 1: Multimodal network hierarchical topology.

centimeter level) is required, differential enhancement technology is required. The working principle of this technology is the reference station that obtains the distance between it and the satellite through the receiver and then compares the obtained value with error with the real value obtained by calculation, calculates the difference of the deviation, and sends it to the user. The user passes the difference to get the pseudorange and then solve the location coordinates of the user. Through this series of calculations, the purpose of accurate positioning is achieved. In general, the pseudorange differential information transmitted by the base station to the user station is the chat and its rate of change. Figure 2 shows the histogram of the rate of change of the multimodal data communication chain. After the user station receives the abovementioned difference information through the data communication link, it starts positioning calculation. The calculation mode of the user station is similar to the single-point positioning mode, except that the pseudorange observed by the user station must be corrected accordingly using the abovementioned difference information.

The current more common method is the linear method of two sets of three orthogonal motions, that is, according to the two sets of plane images and the three sets of plane orthogonal motions, the camera can be linearly calibrated by the pole information in the image, but this method is for the operation of the motion system. The high load and hardware cost requirements are not suitable for situations where the camera motion parameters are unknown, or the camera cannot be precisely controlled to perform a specified motion. It can be seen from the steps that Kalman filtering has two calculation loops, specifically a filtering calculation

loop and a gain calculation loop. Because it has a good denoising ability for many random noises, it is especially effective for filtering out impulse interference and image scanning noise. The gain calculation loop can calculate the gain value independently, but the filter calculation loop needs to use the output value obtained by the gain calculation loop as the input value to calculate the final state estimation value. The map matching process is divided into four stages: (1) data preprocessing: bit data is repaired and filtered, as well as coordinate conversion. It is a map, which can ensure the effectiveness of the map matching algorithm. (2) Screening candidate road sections: according to the location address, obtain the location that will be projected to nearby road sections. The amount of street view data on the electronic map is very large, so the selection of candidate areas can greatly reduce the workload of road segment search. (3) Determine the matching road section: according to the matching algorithm, we find out the street scene that matches the anchor point from many street scenes. (4) Projection matching point: we map the anchor point to the matching street scene in a certain projection method. After the matching section is selected, the vertical projection method is generally used to obtain the projected matching points. The self-calibration method is to use the camera to obtain the information in the scene, obtain two parallel lines in the scene, and use the intersection of the corresponding two lines in the camera image as a vanishing point, which can be calculated by connecting the vanishing point to the optical center. The conversion relationship between the camera coordinate system and the imaging image coordinate system is used to calculate the camera parameters.

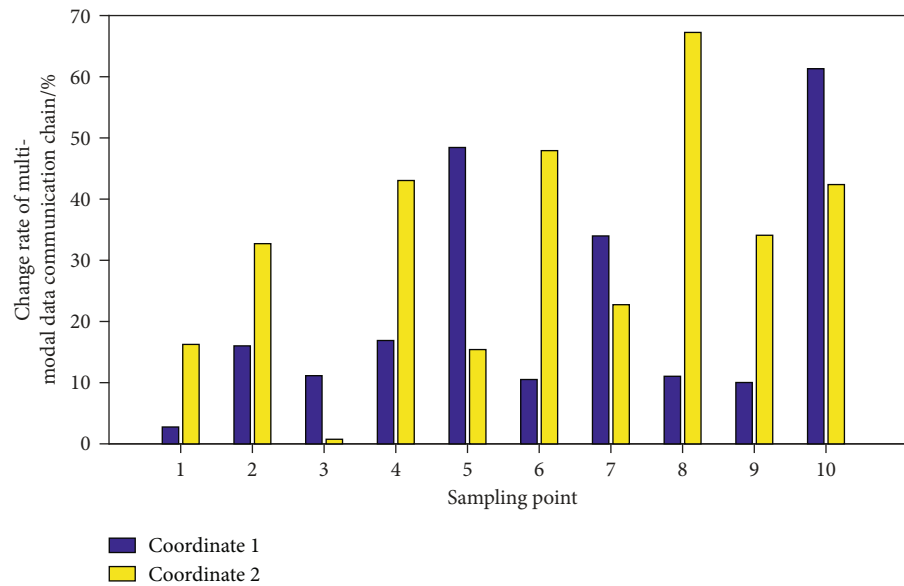


FIGURE 2: Histogram of rate of change of multimodal data communication chain.

3.3. Tourist Street View Navigation Design. The positioning method that this article needs to achieve is to obtain the positioning transformation of the object movement in reality through an external binocular camera and transfer the positioning data of the target object to the Hololens device application after the device coordinate system transformation, so that the Hololens application program matches the appearance of the real object. The same virtual model is projected to the same position as the real object, and the virtual model can realize synchronous movement with the real object according to the model positioning and transformation received by the application program and can provide users with services related to location information, which should be called location services. In this sense, the previous vehicle and surveillance and navigation industries can also be called location services. In a narrow sense, LBS specifically refers to personal neighbors location services, that is, wireless mobile location services for individuals. In general, location service (LBS) is a master of communication technology and mobile positioning technology and an effective combination of positioning services, geographic information, the Internet, the Internet of Things, communication technology, and mobile smart devices. Based on smart devices with different capabilities, location services can be divided into two categories. The first type is mobile phones and other mobile devices with few functions, mainly for mobile viewing and obtaining geographic information, and mainly with pictures and text; the second type is intelligent terminals with navigation capabilities, such as car navigation equipment, and its services are not limited to vehicles and people. It also provides the ability to interact with the server, such as being able to obtain real-time location information. However, due to the large base of mobile terminals such as mobile phones and a wide range of services, the first category has a lot of room for development, and the second category's navigation functions have been greatly improved on the previous basis, so the market prospects are also broad.

Figure 3 shows the navigation frame of tourist street view. By obtaining the spatial coordinates $P : (X, Y, Z)$ of the pixels in the scene in the world coordinate system, it can be converted according to the coordinate system conversion relationship between the world coordinate system and the imaging plane coordinate system in the pinhole camera model. The pixel points $p : (u, v)$ on the two-dimensional image are imaged by the camera, and the method of using the checkerboard image for calibration is essentially to regard the checkerboard used for calibration as a plane object in the scene space. After the checkerboard image is obtained, the checkerboard in the camera image is regarded as a two-dimensional plane calibration object in the camera image. After obtaining the corner coordinates of the checkerboard in the camera image, the coordinates of the corner points of the checkerboard in reality can be compared with it. The coordinates of the corresponding checkerboard corner points in the image are calculated for the coordinate correspondence of the corner points. First, the RFID identifier generates a magnetic field with a radio frequency signal. When the RFID tag comes to the magnetic field, it induces the signal from the magnetic field and transmits the information contained in the tag through the energy transmitted when receiving the signal. The RFID identifier obtains the information and decodes it. After success, the data is transmitted to the information system for subsequent data processing. SCDMA is developed on the basis of CDMA. In TDD mode, the nature of TDMA is used, that is, time is divided into several frames, and each frame is divided into several time slots. Users in each time slot can pass different codes. To make a difference, its main features are fusion of synchronous CDMA, smart antenna, and wireless transmission technology, while using joint detection technology to maintain the time slot width and reduce interference between uplink users, increase spectrum utilization, and reduce multiple access interference.

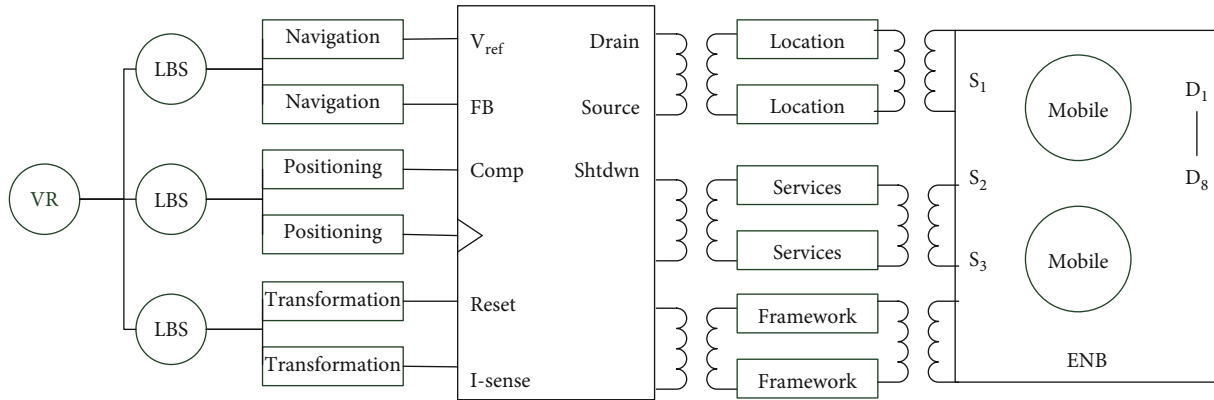


FIGURE 3: Tourist street view navigation framework.

3.4. Analysis of Tourist Positioning Needs. Tourist positioning refers to determining the location of physical entities in a temporal and spatial reference system. Among them, satellite navigation and positioning technology are a modern positioning technology produced in the second half of the last century. The satellite navigation and positioning system are composed of a navigation constellation composed of multiple satellites responsible for navigation, a ground monitoring network, and a customer navigation and positioning terminal (user machine for short). The basic principle is to use the satellite as a spatial reference system by obtaining the real-time speed and position information of the satellite. This system eliminates irrelevant information in the image, restores useful real information, enhances the detectability of related information, and simplifies data to the greatest extent, thereby improving the reliability of feature extraction, image segmentation, matching, and recognition and improving target detection and recognition. The navigation users (including ground, ocean, air, and space users) equipped with the receiver can use the received satellite positioning data to measure the satellite. We use data such as Doppler shift to calculate the user's speed and position. An activity is a screen of the device screen, it can add view UI view, such as button and list view, and set up a listener to respond to its user click event, complete the corresponding logical operation function, and realize user interaction. Activity realizes program interaction by jumping. When running a new activity, the previous activity will be placed in a suspended state and placed in the activity management stack of the activity manager. The service component runs in the background and cannot interact with users. It is mainly used for complex logic implementation or processing time-consuming background operations, such as reading network streams and playing audio. Service life cycle is less than activity, which is related to the way the service is started.

When running the visual inertial navigation coupling method, we compare the positioning data obtained by positioning with the measured positioning true value, obtain the absolute positioning error data by calculating the absolute value of the difference between the two, and then use the absolute positioning error data in operation for statistics. The root mean square error and the mean value of the error during device positioning can be obtained. Figure 4 shows a

ladder diagram of the amount of multimodal tourist positioning data. Statistics of the absolute positioning error and the root-mean-square error of the combined operation of visual SLAM and visual inertial navigation can show that after combining the motion data output by the inertial navigation, there will be large errors in a short time after the machine is initialized. However, in the long-term operation of the equipment, the positioning error of the output can be reduced by using the inertial navigation to assist the equipment. The positioning root mean square error of the device under cooperative work is greater than the positioning root mean square error output by the original visual SLAM odometer.

4. Application and Analysis of Tourist Street View Navigation and Tourist Positioning Model Based on Multimodal Wireless Virtual Reality

4.1. Virtual Reality Data Preprocessing. In order to ensure the accurate transmission of data in the process of switching between interfaces, it was decided to install the above three interfaces in an activity. After comprehensive consideration, it was decided to use frame layout. The first layer adopts a vertical linear layout, which is one-to-one, place a map interface map view and frame layout (specified as framelayout1) at a ratio of, and use a frame layout (specified as frame layout) for the second layer; camera view and other 3D virtual navigation information drawn by OpenGL ES can be added to frame layout. Since the camera does not have a hardware trigger mechanism, it is necessary to match the time stamp of the camera output data with the time stamp of the IMU output data in the calculation and realize the time synchronization between the camera output data and the IMU output data by finding the correspondence between the time stamps. The object area has some noise holes, and some small noise objects are scattered on the background area. Continuous opening and closing operations can effectively improve this situation. Sometimes it takes multiple corruptions and the same number of expansions to produce results. When running the coupling program of the vision sensor and IMU, the robot calls the inertial navigation sensor and

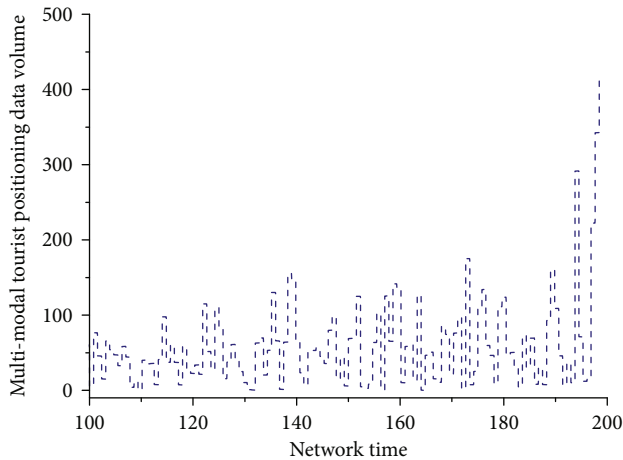


FIGURE 4: Multimodal tourist positioning data volume ladder diagram.

the vision sensor through the application program of the industrial computer to obtain environmental data in the form of multithreading. The core chip of this module adopts AIDS1086 of IMS company to realize. AMS is a low-dropout voltage regulator chip produced by IMS. It can maintain a small voltage difference between the power supply voltage (input terminal) and the load voltage (output terminal). It is widely used in high-efficiency linear regulators, CPU power supplies, and battery charging module, embedded device power supply. In this circuit, VDD-5V is the input, and VDD-10 is the regulated output, which provides the reference input voltage for the follow-up circuit.

Figure 5 shows the multimodal position sensing data curve. From the displayed results, the device's operating pose angle can obtain more accurate data in the short term when only relying on the visual sensor to run the visual SLAM. However, in the long-term operation process, the observation error is caused by the sensor and the long-term. The cumulative error caused by the operation causes the positioning error of the device positioning to continue to increase; while the coupling method is combined with the inertial navigation sensor data, although the error will occur during the long-term operation, it is compared with the visual SLAM alone. In terms of the situation, the mean value of the error between the running positioning and the real positioning is lower. By comparing the error comparison between the operation output positioning of the visual inertial navigation fusion method and the original visual SLAM method, it can be seen that the position output by the visual odometer is used as the observation value input of the Kalman filter algorithm, combined with the six freedoms obtained by the IMU. The degree of motion data can optimize the positioning results obtained by tourists in the scenic spot and make the positioning closer to the real positioning.

4.2. Realization of Tourism Street View Navigation Simulation. The SDRAM module of the travel terminal is used to process data. The module uses the K4X51163PC produced by Samsung, which is a 32 MB SDRAM used in

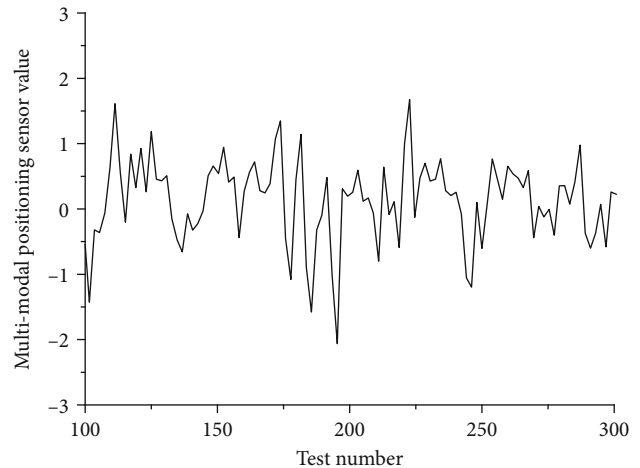


FIGURE 5: Multimodal positioning sensor data curve.

embedded devices. The maximum operating frequency of the K4X51163PC is 133 MHz, and the chip uses a differential clock input. The operating voltage is 1.7 V to 1.95 V, the CAS delay is 2 or 3 clock cycles, the maximum burst length is 16 bits, and it supports 4 channels. The SDRAM module is connected through the external memory module of the main controller. The chip contains with reference voltage source, error amplifier, drive transistor, and phase compensator, and the output voltage is at a minimum interval of 50 mV and can be selected from 0.9 V to 5.0 V. The XC6219A12CMR chip has excellent dynamic response performance and can be used in a wide range. There is a good power supply rejection ratio in the frequency range, even when the load frequency changes, the voltage output still has high stability. The value of r is approximately equal to 0.1, so when the reference input is a step signal, the manipulated variable does not contain impulse function, but contains a sharp pulse function. This phenomenon is called fixed-point shock. Considering the basic PID control system, in order to avoid fixed-point shock phenomenon, we arrange the derivative action only in the feedback path, so that the derivative only occurs on the feedback signal, and it does not occur on the reference signal. Figure 6 shows the distribution of the accuracy of the multimodal street view navigation feedback signal.

First, in the ArcGIS map, we look for the corresponding street view and mark the position of the waypoint based on the center line of the street view. The waypoint includes the starting point and intersection node of the road section and is marked as the waypoint position. We save the latitude and longitude information of each marked waypoint as a kml file, use coordinate conversion processing to extract the coordinate information, and store it in the waypoint data format. According to the improved street view model, it is also necessary to measure the width information of the street view. In order to improve the accuracy of the street view width, it is carried out from the two ends and the middle of the street view, and the average value is taken after multiple measurements as the final street view width information. In the actual calculation process, in order to enhance the

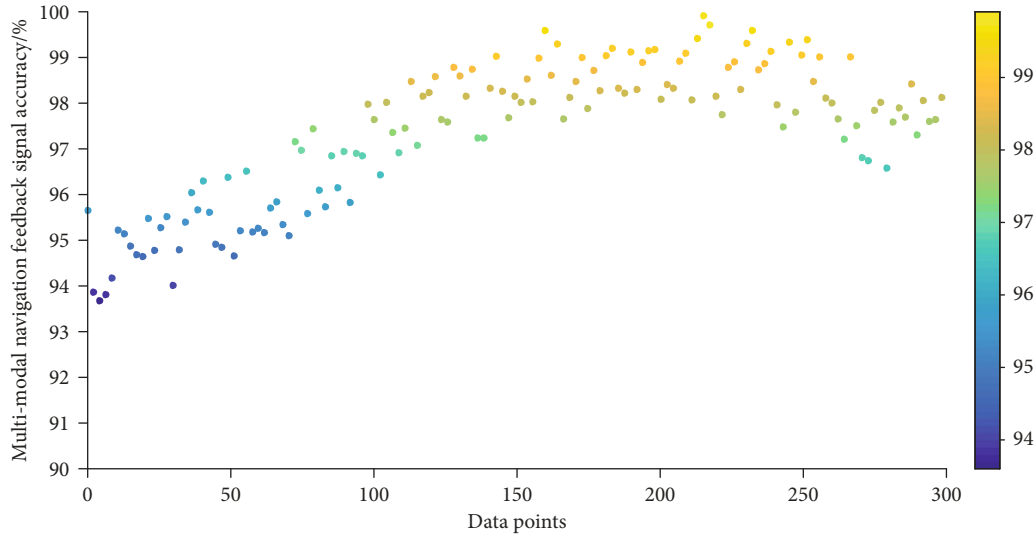


FIGURE 6: Distribution of the accuracy of multimodal street view navigation feedback signals.

robustness of the matching, a total of 16 seed points of 4×4 are used for each key point to describe, so that 128 data can be generated for one key point, that is, a 128-dimensional SIFT feature vector is finally formed. At this time, the SIFT feature vector has removed the influence of geometric deformation factors such as scale change and rotation and then continue to normalize the length of the feature vector to further remove the impact of illumination changes. After the SIFT feature vectors of the two images are generated, in the next step, we use the Euclidean distance of the key point feature vectors as the similarity judgment measure of the key points in the two images. Among these two key points, if the closest distance divided by the next closest distance is less than a certain proportion threshold, then, this pair of matching points is accepted. Every time a user requests another protected page, ASENET authentication will check whether the user is authenticated and then allow the user to view the page or direct the user to the login page accordingly. By default, the authentication cookie is always valid during the user's session. Lowering this ratio threshold will reduce the number of SIFT matching points, but it will be more stable.

4.3. Example Application and Analysis. This experiment completed the gesture recognition based on Kinect in the Unity environment. We use MS-SDK to obtain relevant variables from the system and calculate the information from the Kinect sensor. First, we obtain the position and rotation information of each joint, and then the state of each hand is sent every frame and obtained by the machine learning algorithm obtained by the Kinect's depth camera. In order to compensate for the noise in the data provided by Kinect, first, we calculate three orthogonal axes based on the positions of three adjacent joints and then estimate the direction of each joint for the scene. In addition to the position and orientation variables, two additional variables are obtained: left-handed and right-handed states. The model used in the

API is an SSD model with mobile net, which is used as an initialization checkpoint for training. On the basis of this model, further use gesture images for training, that is, use the data collected by yourself to train and test the pretraining model. When calculating HOG directional gradient histogram features, if you directly calculate the gradient, gradient amplitude, and angle for each small block, and then perform statistics, the calculation complexity of HOG is relatively large. Therefore, this article introduces the concept of integral graph into HOG Calculation in order to accelerate the calculation performance of HOG. The data collected in this paper are 1, 2, and 5 digital gestures. The process involves data set collection and labeling, VOC2012 data set production, tfrecord data generation, SSD migration learning and model export, and real-time detection.

Figure 7 shows the convergence curve of the multimodal street view navigation training data. The data shows that the convergence speed of the last 10 iterations is significantly faster than the previous one. This is because when the nine is larger, the LMA is similar to the gentle convergence speed of the gradient descent. When the nine is small enough, the LMA exhibits the properties of GNA to obtain a faster local convergence speed. Finally, for image stitching, a drawing surface needs to be selected. If an image plane is used as a reference, and other images undergo perspective transformation to this reference plane, the vertical side away from the reference image will become longer and longer, and the cumulative error of perspective transformation will become larger and larger. Some cells are empty (have zero values). In order to speed up the counting of cell values, use integral graphs, where different cell values correspond to one of them. For the n th integral image, when the value of the image at this point is equal to the n th value, its value at point (x, y) is equal to 1. By comparing with traditional PID controllers, it can be seen that the output response time and steady-state performance of neural PID are better than those of the abovementioned PID controllers. Moreover, at the

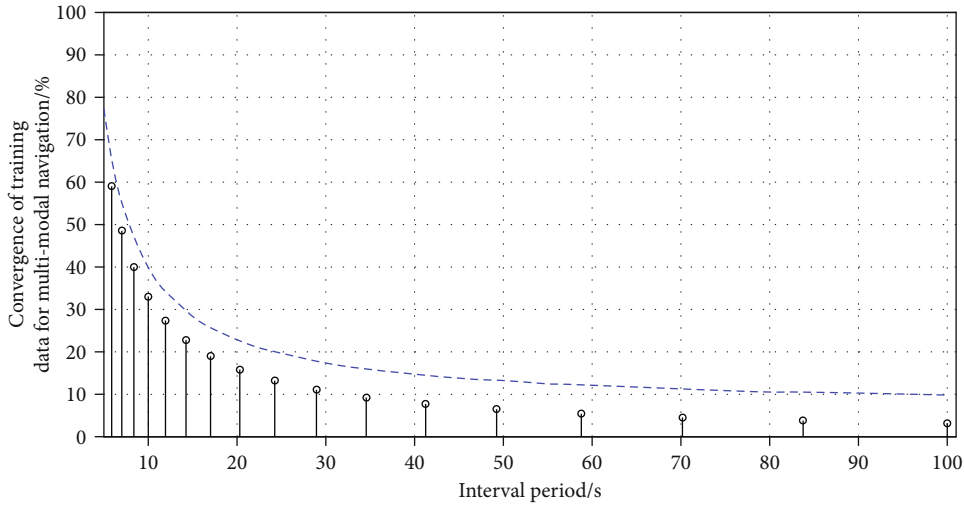


FIGURE 7: Convergence curve of multimodal street view navigation training data.

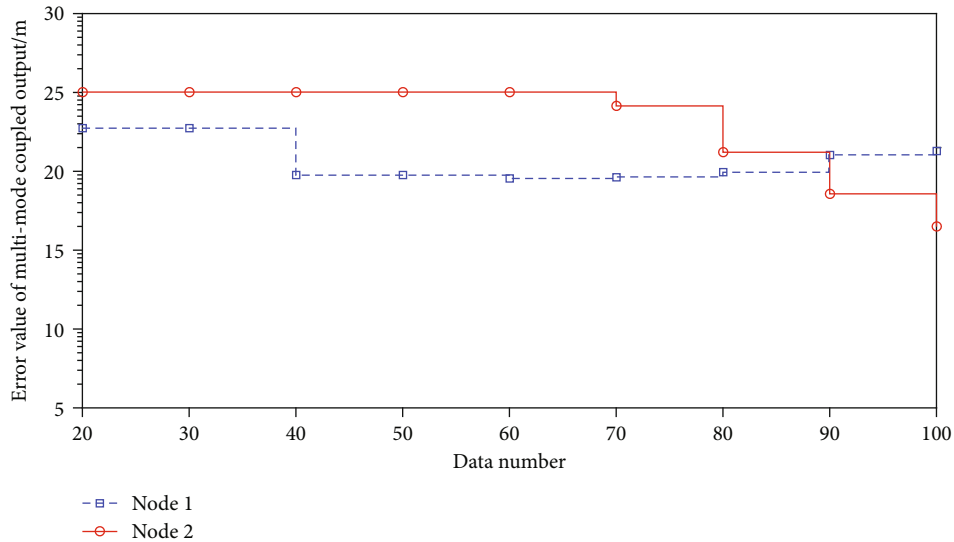


FIGURE 8: Root mean square error results of multimode coupled output.

initial stage, by adjusting the performance function of the neural PID, a good control effect that takes into account the proper output of the rudder angle can be achieved, and the neural network learning can be through online and offline learning methods, and at the same time through signal learning and no signal learning, and finally apply the integral image method to the calculation of HOG features: we count and group the number of cells with the same value in the area to form a histogram, and the length of the obtained histogram is equal to the number of cells.

Figure 8 shows the root mean square error results of the multimode coupled output. Gesture recognition through the Kinect and VGB schemes can avoid the influence of complex backgrounds and lighting, and the VGB scheme is more convenient because of higher recognition accuracy. However, the data provided by Kinect alone is used for recognition. To normalize it, this article uses Photoshop to complete the normalization of the positive sample images

and normalize the normal samples of different sizes into 80×80 training samples of uniform size. Then use the batch command to obtain the positive sample path list in preparation for generating the positive sample VEC file later. The low recognition rate of self-occlusion and tracking of small objects affects the experience of natural interaction in the project. Although the training process is more complicated based on deep neural network, it can quickly and effectively realize the recognition of gestures. In this paper, the combination of VGB and deep neural network solutions is well applied in the project, which not only improves the anti-interference ability of natural gesture interaction but also improves the accuracy of gesture recognition. At the same time, a scheme of adding a camera to capture gestures separately is adopted, and the user experience effect is better. Through many experimental tests, the result of the running positioning data output by the positioning system is compared with the true value of the actual path positioning,

and the root mean square error of the positioning is calculated. The results of the comparison show the effectiveness and robustness of the positioning system combined with inertial navigation and vision. In the course of the experiment, we found that any denoising processing of the original motion data may have a great impact on the final result, and it is very difficult to obtain a universal and effective denoising algorithm. On the other hand, it is inevitable that the raw data will produce noise during the collection process. The smoothness measurement algorithm based on curvature estimation proposed in this paper has good noise immunity, reduces the influence of noise on smoothness detection, and is stable in various situations.

5. Conclusion

The positioning system of the virtual reality device implemented in this paper needs to use the visual sensor to locate the mark on the target object in reality to confirm the position of the target object relative to the mixed reality device. For this reason, a bracket needs to be installed on the target, and the bracket is used to identify the recognition generated. The mark and the sphere mark placed irregularly in the space are fixed on the surface of the target object. By combining tourism products and smart terminals, this paper proposes a design and implementation of a smart tourism system based on a multimodal platform. The hardware of the system is based on the ARM platform, and the schematic diagrams of the main module circuits, such as the power supply, are given. SDRAM and other modules transplant the multimodal system to this platform. Then, we use Eelipse and multimodal SDK as the development environment to develop tourism application software to realize multimedia applications, electronic maps, travel notes, and other functions. This system has carried out a lot of experiments, comprehensively analyzed the experimental results, and proved that the improved moving target detection method proposed in this paper can filter out irrelevant targets, save the targets of interest, and improve the tracking accuracy and alarm correctness. The multimedia application supports audio playback, video playback, and picture browsing of tourist attractions; the electronic map module realizes the display of Google Maps on the terminal device and the positioning of the terminal device; the tourist recording module uses the multimodal SQLite database to realize travel notes input, modify, display, and other functions. Then, combined with positioning data, street view data, and scenic spot data, the improved wireless virtual reality model is used to calculate, and historical track points are introduced in the wireless virtual reality modeling process to analyze the characteristics of pedestrians in the scenic area and comprehensively consider the distance between the positioning point and the street view. By calculating the average positioning error data between the positioning system used in this article and the actual target object positioning and testing the average delay time of the positioning system tracking the target object identification, it can be determined that the mixed reality device positioning system implemented in this article is positioned in the real environment.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] D. Gavalas, V. Kasapakis, C. Konstantopoulos, G. Pantziou, and N. Vathis, "Scenic route planning for tourists," *Personal and Ubiquitous Computing*, vol. 21, no. 1, pp. 137–155, 2017.
- [2] A. Lakehal, S. Lepreux, L. Letalle, and C. Kolski, "From way-finding model to future context-based adaptation of HCI in urban mobility for pedestrians with active navigation needs," *International Journal of Human-Computer Interaction*, vol. 37, no. 4, pp. 378–389, 2021.
- [3] W. Min, S. Jiang, S. Wang et al., "A survey on context-aware mobile visual recognition," *Multimedia Systems*, vol. 23, no. 6, pp. 647–665, 2017.
- [4] C. H.-S. Di Feng, L. Rosenbaum, H. Hertlein et al., "Deep multi-modal object detection and semantic segmentation for autonomous driving: datasets, methods, and challenges," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 3, pp. 1341–1360, 2021.
- [5] M. Kragh, P. Christiansen, M. Laursen et al., "Fieldsafe: dataset for obstacle detection in agriculture," *Sensors*, vol. 17, no. 11, p. 2579, 2017.
- [6] S. Hall and A. I. Padela, "Engaging Muslim Americans for research on community health: lessons learned from a PCOR capacity-building program," *Progress in Community Health Partnerships: Research, Education, and Action*, vol. 15, no. 3, pp. 385–400, 2021.
- [7] J. Mollevik, "Evaluation of strategies for pedestrian guidance using commodity smartphones," in *2016 10th International Conference on Next Generation Mobile Applications, Security and Technologies (NGMAST)*, pp. 41–46, Cardiff, UK, 2016.
- [8] B. Hepp, M. Nießner, and O. Hilliges, "Plan3d: viewpoint and trajectory optimization for aerial multi-view stereo reconstruction," *ACM Transactions on Graphics (TOG)*, vol. 38, no. 1, pp. 1–17, 2019.
- [9] D. Dzemydienė, A. Burinskienė, and A. Miliauskas, "Integration of multi-criteria decision support with infrastructure of smart services for sustainable multi-modal transportation of freights," *Sustainability*, vol. 13, no. 9, p. 4675, 2021.
- [10] G. Benitez-Garcia, M. Haris, Y. Tsuda, and N. Ukita, "Finger gesture spotting from long sequences based on multi-stream recurrent neural networks," *Sensors*, vol. 20, no. 2, p. 528, 2020.
- [11] G. N. Ansah, A. J. B. Rubio, and Y. Bao, "Spatial cognition in a multimedia and intercultural world proceedings of the 7th International Conference on Spatial Cognition (ICSC 2018)," *Cognitive Processing*, vol. 19, no. S1, pp. 1–S76, 2018.
- [12] J. S. Bielo and C. Vaughn, "Materializing the bible: a digital scholarship project from the anthropology of religion," *Digital Humanities and Research Methods*, pp. 119–142, 2021.
- [13] J. Castagno and E. Atkins, "Roof shape classification from LiDAR and satellite image data fusion using supervised learning," *Sensors*, vol. 18, no. 11, p. 3960, 2018.

- [14] H. Blum, P. E. Sarlin, and J. Nieto, "The fishyscapes benchmark: measuring blind spots in semantic segmentation," *International Journal of Computer Vision*, vol. 129, no. 11, pp. 3119–3135, 2021.
- [15] A. Namoun, A. Tufail, N. Mehandjiev, A. Alrehaili, J. Akhlaghinia, and E. Peytchev, "An eco-friendly multimodal route guidance system for urban areas using multi-agent technology," *Applied Sciences*, vol. 11, no. 5, p. 2057, 2021.
- [16] D. Robben, S. Sunaert, and V. Thijs, "Determination of the cerebral perfusion territories using CT perfusion imaging," *Visions*, vol. 26, no. 1, pp. 62–64, 2016.
- [17] C. Willing, T. Brandt, and D. Neumann, "Electronic mobility market platforms—a review of the current state and applications of business analytics," *Electronic Markets*, vol. 27, no. 3, pp. 267–282, 2017.
- [18] J. Del Ser, E. Osaba, J. Sanchez-Medina, I. Fister, and I. Fister, "Bioinspired computational intelligence and transportation systems: a long road ahead," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 2, pp. 466–495, 2020.
- [19] C. Grün, H. Werthner, and B. Pröll, "Assisting tourists on the move—an evaluation of mobile tourist guides," *Mobile business*, pp. 171–180, 2018.
- [20] Y. T. Zheng, Z. J. Zha, and T. S. Chua, "Research and applications on georeferenced multimedia: a survey," *Multimedia Tools and Applications*, vol. 51, no. 1, pp. 77–98, 2011.
- [21] H. Jochen and Z. Alexander, "Supporting multi-modal interfaces for adaptive tour planning," *Geographic Information Sciences*, vol. 10, no. 2, pp. 117–128, 2020.