POSITIONING METHOD AND SYSTEM FOR INSPECTION SCENE

[0001] The present application claims priority to Chinese Patent Application No. 202411464133.4, filed with the China National Intellectual Property Administration (CNIPA) on October 21, 2024 and entitled "POSITIONING METHOD AND SYSTEM FOR INSPECTION SCENE", which is incorporated herein by reference in its entirety.

TECHNICAL FIELD

[0002] The present application relates to the technical field of intelligent transportation, and in particular, to a positioning method and system for an inspection scene.

BACKGROUND ART

[0003] As a peak of infrastructure construction passes, urban elevated roads, urban expressways, highways, tunnels, and other road infrastructure face great management and maintenance pressure. After a specific period of service, road surfaces, bridge structures, and other facilities have various defects and diseases due to a plurality of factors such as temperature, rain, and loads. Minor diseases affect appearances of the facilities, whereas severe diseases affect safety of slow traffic in addition to the appearances.

[0004] For a prolonged service life of road infrastructure, the most important thing is to improve an ability to detect diseases or incidents on the infrastructure, and to accurately describe positions, sizes, types, and other information of the diseases. Positioning data is a piece of important information of the diseases. After obtaining the positioning data, a maintenance team may find the corresponding diseases or incidents based on the positions for maintenance. In addition, the positioning data may facilitate further analytical application: The positioning data may be used to analyze a spatial distribution of the diseases or incidents, such as analyzing a clustering density and distribution pattern of the diseases or incidents in different spaces. It may also be used for temporal and categorical analysis of diseases or incidents at a same position, to perform spatial deduplication on repeatedly reported diseases at a same position; or used to analyze whether the position exhibits a pattern of frequent repairs and maintenance.

[0005] At present, a daily inspection of road facilities is mainly conducted by an inspector through visual observation while walking or driving. When a disease or an incident is found during the inspection, the inspector acquires an image of the disease or incident through an app on a mobile phone, and records and reports positioning data of this position through a

positioning function of the mobile phone. With technological advancements, some inspections have begun to use a professional inspection device to acquire images, an artificial intelligence (AI) algorithm to identify diseases and incidents, and a high-precision positioning sensor to obtain accurate position information, to reduce positioning data errors and improve positioning accuracy.

[0006] Currently, positioning data acquired during a daily inspection is usually two-dimensional longitude and latitude data, and may be displayed as scatter points on an electronic map. Road section positions are matched based on latitudes and longitudes. However, in special cases such as elevated road-ground road, tunnel-ground, and main line-ramp, vertical overlap occurs at a same longitude and latitude. Relying solely on the longitude and latitude makes it impossible to determine whether a current position belongs to an upper layer or a lower layer. Consequently, accurate positioning and road section matching cannot be achieved, which may result in identification and dispatch problems. Some existing ideas and methods include: (1) Manual recording: During an inspection and detection, an inspector observes and records a position of a current inspection scene on an app, or taps a switch added on an inspection device to record inspection scenes corresponding to different inspection times. This method mainly relies on manual subjective judgment and operations, making it susceptible to inaccurate data due to subjective influence of the inspector. (2) Track, altitude, and acceleration data acquisition: A high-precision navigation and positioning receiver and a triaxial acceleration sensor are added in an inspection device to record track, speed, altitude, and triaxial acceleration data of an inspection vehicle. A driving route and uphill and downhill time points of the vehicle are marked by using the track and triaxial acceleration data of the vehicle. A layer of overlapping roads at which the vehicle is likely traveling is determined based on an altitude and a driving status of the vehicle. This method relies heavily on sensor data and basic data acquisition: When a satellite signal is weak, acquired data such as a track, a speed, and an altitude has a large deviation and may even be invalid, resulting in inaccurate positioning. In addition, altitude differentiation thresholds vary for same facilities and are difficult to obtain for elevated roads and tunnels. These factors all interfere with accurate scene classification and subsequent analysis. (3) Data processing through a classification neural network: Some inspection devices are equipped with cameras. Images acquired by the cameras are input into the classification neural network. The neural network determines scenes to which the images belong. Although this method can make a judgment on the scenes of the images, the images can only be classified into limited classes, and it is impossible to further distinguish different objects. For example, if two auxiliary roads of a same bridge are very close, have a same direction, and both belong to a "bridge auxiliary road"

class, the two roads cannot be distinguished by relying solely on a result of an image classification model, leading to inaccurate road matching.

SUMMARY

[0007] An objective of the present application is to provide a positioning method and system for an inspection scene. An edge-cloud collaborative system architecture and an AI image analysis algorithm are used to implement accurate and rapid positioning for a scene of an inspection device. In addition, feature archiving is performed on road scenes through a vector database, to improve accuracy of image scene classification and facilitate subsequent further data matching and analysis.

[0008] To achieve the foregoing objective, the present application provides the following solutions:

[0009] According to a first aspect, the present application provides a positioning method for an inspection scene, including:

[0010] acquiring image data and positioning data of an inspection scene in real time;

[0011] performing semantic segmentation and feature extraction on the image data to obtain a scene classification result and a feature vector;

[0012] performing disease identification based on the image data, the scene classification result, and the feature vector, and if a disease is identified, performing a hybrid search in a preset scene vector database based on the positioning data and the feature vector to obtain a hybrid search result of the inspection scene; and

[0013] updating the hybrid search result to the scene vector database to complete disease positioning of the inspection scene.

[0014] In an exemplary embodiment, performing the semantic segmentation and the feature extraction on the image data to obtain the scene classification result and the feature vector includes:

[0015] inputting an original image into a semantic segmentation neural network model *M1* and

outputting a semantic segmentation image;

[0016] concatenating the semantic segmentation image and the original image to obtain a concatenated image; and

[0017] inputting the concatenated image into a neural network model *M2*, and obtaining the scene classification result and the feature vector.

[0018] In an exemplary embodiment, the semantic segmentation neural network model *M1* uses an image coding neural network and includes an encoder and a decoder. The encoder is configured to extract a high-dimensional feature coding vector from the original image. The decoder is configured to decode the high-dimensional feature coding vector into the semantic segmentation image with same dimensions as the original image.

[0019] In an exemplary embodiment, the neural network model *M2* uses a convolutional neural network and includes a backbone network and a head network. The backbone network processes the concatenated image through a plurality of convolutional layers and a pooling layer to obtain a high-dimensional feature map, and then processes the high-dimensional feature map through a global average pooling layer to obtain the feature vector. The head network processes the feature vector through a multilayer neural network containing a residual block to obtain the scene classification result.

[0020] In an exemplary embodiment, performing the hybrid search includes:

[0021] performing preliminary screening in the scene vector database based on a longitude, a latitude, and an azimuth in the positioning data to obtain a record data set; and

[0022] calculating a cosine similarity between the feature vector and the record data set, and performing secondary screening in the record data set based on the cosine similarity to obtain a hybrid search result set.

[0023] In an exemplary embodiment, updating the hybrid search result to the scene vector database includes:

[0024] if the hybrid search result set is empty, manually reviewing and then adding the image

data, the scene classification result, the feature vector, and the positioning data to the scene vector database;

[0025] if the hybrid search result set is not empty and object IDs and scene classes in the hybrid search result set are all the same, updating the image data, the scene classification result, the feature vector, and the positioning data to the scene vector database; or

[0026] if the hybrid search result set is not empty and different object IDs exist in the hybrid search result set, performing arithmetic averaging on feature vectors in the hybrid search result set to obtain a new feature vector, performing a hybrid search again in the scene vector database to obtain a new hybrid search result set, performing voting to obtain a corrected object ID and scene class, and updating the corrected object ID and scene class to the scene vector database.

[0027] In an exemplary embodiment, after updating the hybrid search result to the scene vector database, the method further includes: manually cleaning and reviewing the scene vector database on a regular basis.

[0028] According to a second aspect, the present application further provides a positioning system for an inspection scene, including an edge device and a cloud device. The edge device is configured to acquire image data and positioning data of an inspection scene in real time, and perform semantic segmentation and feature extraction on the image data to obtain a scene classification result and a feature vector. The cloud device is configured to process the image data, the positioning data, the scene classification result, and the feature vector to obtain a disease identification result and a disease position of the inspection scene.

[0029] In an exemplary embodiment, the edge device includes an industrial computer, a high-definition camera, and a navigation and positioning receiver. The high-definition camera and the navigation and positioning receiver are connected to the industrial computer. The industrial computer is configured to provide data transmission, image feature extraction, and power supply functions.

[0030] In an exemplary embodiment, the cloud device includes a disease identification module, a hybrid search module, and a dispatch module. The disease identification module is configured to perform disease identification based on the image data, the scene classification result, and the feature vector. The hybrid search module is configured to: if a disease is identified, perform a

hybrid search in a preset scene vector database based on the positioning data and the feature vector to obtain a hybrid search result of the inspection scene. The dispatch module is configured to update the hybrid search result to the scene vector database, and send the hybrid search result to corresponding maintenance personnel in combination with a disease identification result.

[0031] According to specific embodiments provided in the present application, the present application discloses the following technical effects:

[0032] (1) Enhanced efficiency through a combination of edge computing and cloud computing: Edge-cloud collaboration eliminates a need for manual on-site recording and measurement, making real-time data processing more efficient and rapid. The edge device is responsible for on-site data acquisition and preliminary processing, to reduce pressure of transmitting a large amount of raw data while ensuring data security and privacy protection.

[0033] (2) Complementarity of high-precision image recognition and manual review: The high-definition camera and an intelligent algorithm are integrated to perform semantic object extraction on a traffic facility in an image. An extraction result is combined with the original image such that a subsequent analysis model can obtain more information about on-site infrastructure. In addition, in the present application, manual review is retained to ensure result accuracy, particularly in an uncertain or ambiguous situation where manual intervention can provide a final decision.

[0034] (3) Multi-level data analysis: During scene and ID matching, the hybrid search fully utilizes diverse data such as the feature vector and the acquired longitude, latitude, and azimuth, to help improve search speed and accuracy. In comparison with relying solely on classes for discrimination, a difference between different objects of a same class is considered in the present application, enabling effective differentiation between objects that are physically close, share a same scene class, but belong to different facilities.

[0035] (4) Database optimization management: The vector database used in the present application is a composite relational database, which supports efficient query and storage functions. It does not need to record special field thresholds for classification during use, to facilitate management and maintenance of a large amount of historical data. It also provides a flexible data update mechanism to adapt to changing road conditions.

BRIEF DESCRIPTION OF THE DRAWINGS

[0036] To describe the technical solutions in the embodiments of the present application or in the related art more clearly, the following briefly describes the accompanying drawings required for the embodiments. Apparently, the accompanying drawings in the following description show merely some embodiments of the present application, and a person of ordinary skill in the art may still derive other accompanying drawings from these accompanying drawings without creative efforts.

[0037] FIG. 1 is a flowchart of a positioning method for an inspection scene according to an embodiment of the present application;

[0038] FIG. 2 is a schematic diagram of installation of an intelligent inspection device according to an embodiment of the present application;

[0039] FIG. 3 is a schematic diagram of a height from a vanishing point in an acquired image according to an embodiment of the present application; and

[0040] FIG. 4 is a flowchart of data processing according to an embodiment of the present application.

DETAILED DESCRIPTION OF THE EMBODIMENTS

[0041] The technical solutions in the embodiments of the present application are clearly and completely described below with reference to the accompanying drawings in the embodiments of the present application. Apparently, the described embodiments are merely some rather than all of the embodiments of the present application. All other embodiments obtained by a person of ordinary skill in the art based on the embodiments of the present application without creative efforts shall fall within the protection scope of the present application.

[0042] An objective of the present application is to provide a positioning method and system for an inspection scene. An edge-cloud collaborative system architecture and an AI image analysis algorithm are used to implement accurate and rapid positioning for a scene of an inspection device. In addition, feature archiving is performed on road scenes through a vector database, to improve accuracy of image scene classification and facilitate subsequent further data matching and analysis.

[0043] To make the foregoing objectives, features, and advantages of the present application clearer and more comprehensible, the present application is further described in detail below with reference to the accompanying drawings and specific implementations.

[0044] This embodiment provides a positioning method for an inspection scene, as shown in FIG. 1, including the following steps:

[0045] Acquire image data and positioning data of an inspection scene in real time.

[0046] Perform semantic segmentation and feature extraction on the image data to obtain a scene classification result and a feature vector.

[0047] Perform disease identification based on the image data, the scene classification result, and the feature vector, and if a disease is identified, perform a hybrid search in a preset scene vector database based on the positioning data and the feature vector to obtain a hybrid search result of the inspection scene.

[0048] Update the hybrid search result to the scene vector database to complete disease positioning of the inspection scene.

[0049] Specifically, in this embodiment, the image data and the positioning data of the inspection scene are acquired in real time through an intelligent inspection device installed in an inspection vehicle. The intelligent inspection device is configured with an intelligent algorithm, and can perform semantic segmentation, feature extraction, and image classification on the image data, and upload acquired and analyzed data to a cloud. After receiving the acquired data, the cloud performs a hybrid search through the feature vector, obtains a corresponding analysis result through post-processing, and stores relevant data in the database.

[0050] Composition and installation of the intelligent inspection device are shown in FIG. 2, including one or more high-definition cameras and an industrial computer. The industrial computer is also connected to a high-precision navigation and positioning receiver, is equipped with an Internet card, has specific AI computing power, and enables an identification algorithm program to be deployed. Regarding a power supply source, the intelligent inspection device uses a direct current (DC) power supply. In addition, the intelligent inspection device has good waterproof performance.

[0051] The industrial computer is installed inside the vehicle to power a vehicle-mounted device. It uses a fourth-generation (4G) or fifth-generation (5G) mobile communication technology and is connected to the high-precision navigation and positioning receiver and the high-definition cameras. It may send an image acquisition instruction to the high-definition cameras, and continuously receive, process, and upload acquired data. The industrial computer may further include a micro graphics processing unit (GPU). A deep neural network algorithm may accelerate image processing through the GPU.

[0052] The high-definition camera is powered by the industrial computer or vehicle-mounted device. It features a resolution of at least 1 megapixel and acquires a clear image with zero latency at a frame rate of 25 Hz or higher. It may be installed inside or outside the vehicle with a fixed field of view and configured to acquire an image of the vehicle's front or rear. The acquired image needs to be clear and has no obvious distortion and obstruction. As shown in FIG. 3, a ratio of a height h' from a vanishing point in the image to an upper edge of the image to a total height h of the image needs to fall within a specific range: $1/3 \leq h'/h \leq 3/4$.

[0053] A convolutional neural network classification model may be deployed on the industrial computer to vertically divide the image into n equal intervals and predict which interval has a highest probability of containing the vanishing point in the image.

[0054] The high-precision navigation and positioning receiver has a positioning error of at most 1 m, is configured with an inertial navigation system, and acquires the positioning data at a frequency of 10 Hz or higher.

[0055] Further, in this embodiment, data identification and processing are performed through a combination of perception and identification on an edge device and data post-processing in a cloud. A processing procedure and logic are shown in FIG. 4.

[0056] Performing semantic segmentation and feature extraction on the image data to obtain the scene classification result and the feature vector includes: inputting an original image into a semantic segmentation neural network model *M1* and outputting a semantic segmentation image; concatenating the semantic segmentation image and the original image to obtain a concatenated image; and inputting the concatenated image into a neural network model *M2*, and obtaining the scene classification result and the feature vector.

[0057] The semantic segmentation neural network model *M1* uses an image coding neural network and includes an encoder and a decoder. The encoder is configured to extract a high-dimensional feature coding vector from the original image. The decoder is configured to decode the high-dimensional feature coding vector into the semantic segmentation image with same dimensions as the original image.

[0058] The neural network model *M2* uses a convolutional neural network and includes a backbone network and a head network. The backbone network processes the concatenated image through a plurality of convolutional layers and a pooling layer to obtain a high-dimensional feature map, and then processes the high-dimensional feature map through a global average pooling layer to obtain the feature vector. The head network processes the feature vector through a multilayer neural network containing a residual block to obtain the scene classification result.

[0059] Performing the hybrid search includes: performing preliminary screening in the scene vector database based on a longitude, a latitude, and an azimuth in the positioning data to obtain a record data set; and calculating a cosine similarity between the feature vector and the record data set, and performing secondary screening in the record data set based on the cosine similarity to obtain a hybrid search result set.

[0060] Updating the hybrid search result to the scene vector database includes:

[0061] if the hybrid search result set is empty, manually reviewing and then adding the image data, the scene classification result, the feature vector, and the positioning data to the scene vector database;

[0062] if the hybrid search result set is not empty and object IDs and scene classes in the hybrid search result set are all the same, updating the image data, the scene classification result, the feature vector, and the positioning data to the scene vector database; or

[0063] if the hybrid search result set is not empty and different object IDs exist in the hybrid search result set, performing arithmetic averaging on feature vectors in the hybrid search result set to obtain a new feature vector, performing a hybrid search again in the scene vector database to obtain a new hybrid search result set, performing voting to obtain a corrected object ID and scene class, and updating the corrected object ID and scene class to the scene vector database.

[0064] Specifically, in this embodiment, during an inspection, the inspection vehicle acquires an original image m_i, and performs semantic segmentation on the image through the semantic segmentation neural network model *M1* to obtain a semantic segmentation image m_i'; concatenates the original image m_i and the semantic segmentation image m_i' to obtain a concatenated image m_i*; inputs the concatenated image m_i* into the neural network model *M2* to obtain a scene classification result c_i, and outputs a corresponding feature vector v_i; and after image processing is completed, uploads the acquired image, c_i, v_i, and a corresponding longitude, latitude, and azimuth to a cloud server.

[0065] The original image m_i usually has dimensions of [u, v, 3], that is, has three channels: RGB. An aspect ratio of the image is usually 16:9 or 4:3.

[0066] The azimuth, which may also be referred to as a heading angle, serves as a reference direction for image acquisition through the high-definition camera. With due north defined as 0°, the azimuth increases clockwise, and its value range is [0, 360).

[0067] The semantic segmentation neural network model *M1* is an image encoding-decoding neural network. First, a high-dimensional feature coding vector is extracted from the original image m_i through a feature extraction encoder neural network layer. Then, the feature vector is decoded into the semantic segmentation image m_i' with same pixel dimensions as the original image through a feature decoder neural network layer. Semantic segmentation is mainly to label a class of a road surface, a traffic safety facility such as a guardrail, a roadside facility such as a sign, and a road facility such as an overpass or an elevated structure in the image, and assign a specified value in [100, 255] based on the class. All unlabeled areas are classified as a "background" class with a value of 0. In this way, a class label is assigned to each pixel in the image, to generate the semantic segmentation image m_i' with dimensions of [u, v]. Assigning the specified value in [100, 255] based on the class means selecting j integers from [100, 255], corresponding to j types of facilities recognizable by an image semantic recognition network. One class corresponds to one integer.

[0068] Concatenating the original image m_i and the semantic segmentation image m_i' means concatenating the two images along a channel dimension to form the concatenated image m_i* with dimensions of [u, v, 4]. The 4 channels are an RGB image and the semantic segmentation image.

[0069] The neural network model *M2* includes two parts: the backbone network, which is a convolutional neural network trained based on re-identification (ReID), and a fully connected classification head network added on a basis of the backbone network for classification.

[0070] (1) Backbone network: The model first performs preprocessing such as downscaling or normalization on the concatenated image m_i*. The concatenated image m_i* is processed through a network containing a plurality of convolutional layers and a pooling layer to obtain a high-dimensional feature map [u', v', k]. Then, the high-dimensional feature map is processed through global average pooling to obtain a 512-dimensional feature vector v_i.

[0071] (2) Head network: v_i is input into a multilayer neural network containing a residual block for further processing. k classes and corresponding confidences are finally output from the neural network. The k classes correspond to k scene classes to be classified by the model, including but not limited to "ground road", "expressway", "auxiliary road", "tunnel", and "road under an elevated structure". A class with a highest confidence is used as a scene classification result determined by the model *M2* and denoted as c_i.

[0072] After analysis, the industrial computer uploads m_i, c_i, v_i, and a longitude and latitude [lng_i, lat_i] and an azimuth azi_i corresponding to a time to the cloud server. Numerical precision of the longitude and latitude [lng_i, lat_i] reaches 0.000001.

[0073] After receiving new acquired data, the cloud server performs a hybrid search in the scene vector database in the cloud by using the longitude and latitude [lng_i, lat_i], the azimuth azi_i, the feature vector v_i, and the class c_i, executes corresponding processing logic based on a hybrid search result to determine a scene in the acquired image, and updates the scene vector database in the cloud.

[0074] The scene vector database is a composite relational database and contains a plurality of records. Each record in the database includes a URL *url* of an image acquired, a longitude and latitude [*lng*, *lat*] of the image, an azimuth *azi* at which the image is acquired, a scene classification result *c* of the image, an object ID of the image, a feature vector *v* obtained after performing feature extraction on the image, a creation time *CREATETIME* of the record, whether the record is reviewed by a reviewer *CHECK*, and whether the record is marked for deletion *DELETE*. Data in the vector database dynamically changes. Reviewed data may be added to the

vector database. The reviewer may further clean and review the vector database on a regular basis to correct a scene class of an image.

[0075] The hybrid search is to perform preliminary screening on the data in the database based on the uploaded longitude, latitude, and azimuth to obtain a record data set $Vset_i$ with a longitude and latitude close to [lng_i, lat_i] and an azimuth similar to azi_i, calculate a similarity between the feature vector v_i and $Vset_i$ through a cosine distance, set a cosine similarity threshold T, and select a new vector set with a similarity greater than or equal to T from $Vset_i$ as a hybrid search result set $Vset_i'$.

[0076] Corresponding service post-processing is performed based on the hybrid search result set $Vset_i'$:

[0077] (1) If $Vset_i'$ is an empty set, that is, the vector database contains no vector similar to v_i at a longitude and latitude close to [lng_i, lat_i], data uploaded by the device, such as v_i, c_i, and m_i, is pushed to a pending confirmation queue. After being reviewed and confirmed by the reviewer, the data is added to the vector database V_{db}.

[0078] (2) If $Vset_i'$ is not an empty set, object IDs in $Vset_i'$ are all the same, and scene classes are all c_i, it is determined that a scene classification result corresponding to the acquired data is c_i, and data such as v_i, c_i, and the object ID is updated to V_{db}.

[0079] (3) If $Vset_i'$ is not an empty set and different object IDs exist in $Vset_i'$, arithmetic averaging is performed on feature vectors in $Vset_i'$ to obtain a new feature vector v''. A new feature search is performed in V_{db} by using [lng_i, lat_i], azi_i, and v'' to obtain e vectors with highest similarities, which form a new feature vector set $Vset_i''$. Voting is performed based on object IDs corresponding to the vectors. A scene class c'' corresponding to a most probable object ID in a voting result is used as a scene class of the acquired image. Data such as c'', [lng_i, lat_i], azi_i, and the most probable object ID is updated to V_{db}.

[0080] A voting manner includes but is not limited to:

[0081] i. Majority voting: A scene class with a highest count in the e vectors is used as the voting result.

[0082] ii. Weighted voting: Weighted voting is performed based on a cosine distance. A smaller distance indicates that more votes can be cast. An object ID with most votes in the e vectors is finally used.

[0083] Further, after the hybrid search result is updated to the scene vector database, the method further includes: manually cleaning and reviewing the scene vector database on a regular basis.

[0084] Specifically, the reviewer may clean the existing vector database V_{db} on a regular basis through an auxiliary tool, including but not limited to the following methods:

[0085] (1) The data in V_{db} is re-identified through a new neural network model *M2'* different from the neural network model *M2*. A data record with an identification result different from that in V_{db} is selected and pushed to the pending confirmation queue. The reviewer reviews the record, corrects a mismatched scene class, and updates a corrected scene class to the database.

[0086] (2) Images with a same object ID are extracted from V_{db}. Record data that does not belong to the ID is cleaned.

[0087] To further optimize the technical solutions, this embodiment further provides a positioning system for an inspection scene, including an edge device and a cloud device. The edge device is configured to acquire image data and positioning data of an inspection scene in real time, and perform semantic segmentation and feature extraction on the image data to obtain a scene classification result and a feature vector. The cloud device is configured to process the image data, the positioning data, the scene classification result, and the feature vector to obtain a disease identification result and a disease position of the inspection scene.

[0088] Further, the edge device includes an industrial computer, a high-definition camera, and a navigation and positioning receiver. The high-definition camera and the navigation and positioning receiver are connected to the industrial computer. The industrial computer provides data transmission, image feature extraction, and power supply functions.

[0089] Specifically, the edge device is mainly a vehicle-mounted intelligent inspection device. A description related to the device is as follows:

[0090] (1) The device is equipped with the industrial computer. The industrial computer is

equipped with an NVIDIA GT 1030 graphics card and can run a deep learning algorithm program. The industrial computer uses a DC power supply to ensure stable operation. The industrial computer has good waterproof performance and adapts to various outdoor conditions.

[0091] (2) A high-precision global positioning system (GPS) module and an Internet card are disposed in the industrial computer. It can acquire a longitude, a latitude, and an azimuth at a frame rate of 10 fps and transmit data to a cloud server in real time.

[0092] (3) The high-definition camera is installed on a roof of a vehicle and connected to the industrial computer. The camera acquires an image with a resolution of 2 megapixels at a frame rate of 25 fps. The image is clear and unobstructed. The camera faces the vehicle's front with a horizontal field of view.

[0093] A data acquisition and processing procedure of the edge device is as follows:

[0094] (1) After being powered on, the edge device acquires GPS data from a GPS sensor and determines, based on a GPS status, whether the vehicle is performing an inspection. When the vehicle is in motion, the industrial computer accumulates a travel distance of the vehicle. When the accumulated distance reaches 10 m, the industrial computer issues an image and GPS data acquisition instruction, and starts a new distance accumulation from 0.

[0095] (2) An acquired image m_i is input into a semantic segmentation neural network model *M1*. The model mainly performs semantic perception on a traffic safety facility, a road surface, a bridge, an elevated structure, or another facility in the image, and processes the image to generate a semantic segmentation image m_i' with same pixel dimensions as the original image m_i. A class is assigned to each pixel in the image m_i', including guardrail, anti-glare panel, road surface, bridge, traffic sign, sky, warning pole, background, or the like. In addition, a specific integer in [0, 255) is assigned based on the class.

[0096] (3) The image m_i and the semantic segmentation image m_i' are concatenated along a channel dimension to obtain a concatenated image m_i*. m_i* is scaled and input into a neural network model *M2*. The model outputs a 512-dimensional vector v_i and a k-dimensional confidence array corresponding to k scene classes, such as "expressway", "ground road", and "tunnel". A scene class with a highest confidence is a scene class inferred by the model and denoted as c_i.

[0097] (4) The image m_i, a longitude and latitude [lng_i, lat_i], and an azimuth azi_i that are acquired and analysis result data v_i and c_i are summarized and uploaded to the cloud device.

[0098] Further, the cloud device includes a disease identification module, a hybrid search module, and a post-processing module. The disease identification module is configured to perform disease identification based on the image data, the scene classification result, and the feature vector. The hybrid search module is configured to: if a disease is identified, perform a hybrid search in a preset scene vector database based on the positioning data and the feature vector to obtain a hybrid search result of the inspection scene. The post-processing module is configured to update the hybrid search result to the scene vector database, and send the hybrid search result to corresponding maintenance personnel in combination with a disease identification result.

[0099] Specifically, the cloud device is mainly a cloud server. A data receiving program, a scene vector database, a data processing and analysis program, a dispatch program, and the like are installed on the server.

[0100] (1) Disease identification module:

[0101] After receiving new data, including the image m_i, longitude and latitude [lng_i, lat_i], azimuth azi_i, and analysis result data v_i and c_i, the cloud server inputs the image data into a disease identification model. If a disease is identified in the image, the acquired data is sent to a post-processing program for further analysis. In this embodiment, the disease identification model uses an object recognition neural network and can mark information such as a type of the identified disease and a position of a bounding box in the image.

[0102] (2) Hybrid search module:

[0103] After data is received, a hybrid search is performed in a vector database Milvus by using the longitude and latitude [lng_i, lat_i], azimuth azi_i, and feature vector v_i. The vector database Milvus contains the following field information:

[0104] *lng*: a longitude.

[0105] *lat*: a latitude.

[0106] *azi*: an azimuth. Its value range is [0, 360).

[0107] *vec*: a feature vector.

[0108] *CHECK*: whether data is reviewed by a reviewer. 1 indicates reviewed. 0 indicates not reviewed.

[0109] *DELETE*: whether data is marked for deletion. 1 indicates marked for deletion. 0 indicates not marked for deletion.

[0110] *CREATETIME*: an acquisition time.

[0111] *c*: a scene class.

[0112] ID: an object ID.

[0113] During the hybrid search, the following search conditions are mainly satisfied:

[0114] i. $lng_i - 0.00001 \times 8 < lng_i + 0.00001 \times 8$.

[0115] ii. $lat_i - 0.00001 \times 8 < lat_i + 0.00001 \times 8$.

[0116] iii. $azi_i - 5 < azi < azi_i + 5$ ($azi_i \in [5, 355]$).

[0117] $0 < azi < azi_i + 5 \cup azi > azi_i - 5 + 360$ ($azi_i \in [0, 5)$).

[0118] $azi_i - 5 < azi < 360 \cup azi < azi_i + 5 - 360$ ($azi_i \in (355, 360)$).

[0119] iv. $DELETE = 0$ and $CHECK = 1$.

[0120] v. For a vector obtained through the search, a cosine distance is greater than 0.8.

[0121] A search result set $Vset_i$ is obtained based on constraint conditions of the hybrid search.

Analyzing the search result set $Vset_i$ includes:

[0122] i. If $Vset_i'$ is an empty set, that is, the vector database contains no vector similar to v_i at a longitude and latitude close to [lng_i, lat_i], data uploaded by the device, such as v_i, c_i, and m_i, is pushed to a pending confirmation queue. After being reviewed and confirmed by the reviewer, the data is added to the vector database V_{db}.

[0123] ii. If $Vset_i'$ is not an empty set, object IDs in $Vset_i'$ are all the same, and scene classes are all c_i, it is determined that a scene classification result corresponding to the acquired data is c_i, and data such as v_i, c_i, and the object ID is updated to V_{db}.

[0124] iii. If $Vset_i'$ is not an empty set and different object IDs exist in $Vset_i'$, arithmetic averaging is performed on feature vectors in $Vset_i'$ to obtain a new feature vector v''. A new feature search is performed in V_{db} by using [lng_i, lat_i], azi_i, and v'' to obtain 10 vectors with highest similarities, which form a new feature vector set $Vset_i''$. Voting is performed based on object IDs corresponding to the vectors. A scene class c'' corresponding to a most probable object ID in a voting result is used as a scene class of the acquired image. Data such as c'', [lng_i, lat_i], azi_i, and the most probable object ID is updated to V_{db}. The following voting logic is used: Weighted voting is performed based on a cosine distance. A smaller distance indicates that more votes can be cast. Quantities of votes that can be cast are [10, 9, 8, 7, 6, 5, 4, 3, 2, 1] in sequence. An object ID with most votes in the 10 vectors is finally used.

[0125] (3) Dispatch module:

[0126] If post-processed data does not need to be reviewed by the reviewer, it is directly pushed to the dispatch module. If it needs to be reviewed by the reviewer, relevant data is pushed to the dispatch module after the review by the reviewer succeeds. In the dispatch module, disease data is matched with a road section through a scene class and an object ID of a facility, and then pushed to a specific maintenance team and archived for subsequent further analysis.

[0127] Further, the reviewer may clean the existing scene vector database V_{db} on a regular basis through an auxiliary tool, including the following methods:

[0128] (1) The image data in V_{db} is re-identified through a deeper neural network model $M2'$. A data record with an identification result different from that in V_{db} is selected and pushed to the

pending confirmation queue. The reviewer reviews the record, corrects a mismatched scene class, and updates a corrected scene class to the database.

[0129] (2) Images with a same object ID are extracted from V_{db}. Record data that does not belong to the ID is cleaned.

[0130] Each embodiment of this specification is described in a progressive manner, each embodiment focuses on the difference from other embodiments, and the same and similar parts between the embodiments may refer to each other.

[0131] The foregoing embodiments are merely intended to describe the preferred implementations of the present application, but not to limit the scope of the present application. Various modifications and improvements made to the technical solutions of the present application by those of ordinary skill in the art without departing from the design spirit of the present application shall fall within the protection scope of the appended claims of the present application.

WHAT IS CLAIMED IS:

1. A positioning method for an inspection scene, comprising:

acquiring image data and positioning data of an inspection scene in real time;

performing semantic segmentation and feature extraction on the image data to obtain a scene classification result and a feature vector;

performing disease identification based on the image data, the scene classification result, and the feature vector, and if a disease is identified, performing a hybrid search in a preset scene vector database based on the positioning data and the feature vector to obtain a hybrid search result of the inspection scene; and

updating the hybrid search result to the scene vector database to complete disease positioning of the inspection scene; wherein

the scene vector database is a composite relational database and contains a plurality of records, and each record comprises a URL *url* of an image acquired, a longitude and latitude [*lng*, *lat*] at which the image is acquired, an azimuth *azi* at which the image is acquired, a scene classification result *c* of the image, an object ID of the image, a feature vector *v* obtained after performing feature extraction on the image, a creation time *CREATETIME* of the record, whether the record is reviewed by a reviewer *CHECK*, and whether the record is marked for deletion *DELETE*;

performing the semantic segmentation and the feature extraction on the image data to obtain the scene classification result and the feature vector comprises:

inputting an original image into a semantic segmentation neural network model *M1* and outputting a semantic segmentation image;

concatenating the semantic segmentation image and the original image to obtain a concatenated image; and

inputting the concatenated image into a neural network model *M2*, and obtaining the scene classification result and the feature vector;

the semantic segmentation neural network model *M1* uses an image coding neural network and comprises an encoder and a decoder; the encoder is configured to extract a high-dimensional feature coding vector from the original image; and the decoder is configured to decode the high-dimensional feature coding vector into the semantic segmentation image with same dimensions as the original image;

the neural network model *M2* uses a convolutional neural network and comprises a backbone network and a head network; the backbone network processes the concatenated image through a plurality of convolutional layers and a pooling layer to obtain a high-dimensional

feature map, and then processes the high-dimensional feature map through a global average pooling layer to obtain the feature vector; and the head network processes the feature vector through a multilayer neural network containing a residual block to obtain the scene classification result;

performing the hybrid search comprises:

performing preliminary screening in the scene vector database based on a longitude, a latitude, and an azimuth in the positioning data to obtain a record data set; and

calculating a cosine similarity between the feature vector and the record data set, and performing secondary screening in the record data set based on the cosine similarity to obtain a hybrid search result set; and

updating the hybrid search result to the scene vector database comprises:

if the hybrid search result set is empty, manually reviewing and then adding the image data, the scene classification result, the feature vector, and the positioning data to the scene vector database;

if the hybrid search result set is not empty and object IDs and scene classes in the hybrid search result set are all the same, updating the image data, the scene classification result, the feature vector, and the positioning data to the scene vector database; or

if the hybrid search result set is not empty and different object IDs exist in the hybrid search result set, performing arithmetic averaging on feature vectors in the hybrid search result set to obtain a new feature vector, performing a hybrid search again in the scene vector database to obtain a new hybrid search result set, performing voting to obtain a corrected object ID and scene class, and updating the corrected object ID and scene class to the scene vector database.

2. The positioning method for an inspection scene according claim 1, after updating the hybrid search result to the scene vector database, further comprising: manually cleaning and reviewing the scene vector database on a regular basis.

3. A positioning system for an inspection scene, comprising an edge device and a cloud device, wherein the edge device is configured to acquire image data and positioning data of an inspection scene in real time, and perform semantic segmentation and feature extraction on the image data to obtain a scene classification result and a feature vector; and the cloud device is configured to process the image data, the positioning data, the scene classification result, and the feature vector to obtain a disease identification result and a disease position of the inspection scene;

performing semantic segmentation and feature extraction on the image data to obtain the

scene classification result and the feature vector comprises:

inputting an original image into a semantic segmentation neural network model *M1* and outputting a semantic segmentation image;

concatenating the semantic segmentation image and the original image to obtain a concatenated image; and

inputting the concatenated image into a neural network model *M2*, and obtaining the scene classification result and the feature vector;

the semantic segmentation neural network model *M1* uses an image coding neural network and comprises an encoder and a decoder; the encoder is configured to extract a high-dimensional feature coding vector from the original image; and the decoder is configured to decode the high-dimensional feature coding vector into the semantic segmentation image with same dimensions as the original image;

the neural network model *M2* uses a convolutional neural network and comprises a backbone network and a head network; the backbone network processes the concatenated image through a plurality of convolutional layers and a pooling layer to obtain a high-dimensional feature map, and then processes the high-dimensional feature map through a global average pooling layer to obtain the feature vector; and the head network processes the feature vector through a multilayer neural network containing a residual block to obtain the scene classification result;

the cloud device comprises a disease identification module, a hybrid search module, and a dispatch module; the disease identification module is configured to perform disease identification based on the image data, the scene classification result, and the feature vector; the hybrid search module is configured to: if a disease is identified, perform a hybrid search in a preset scene vector database based on the positioning data and the feature vector to obtain a hybrid search result of the inspection scene; and the dispatch module is configured to update the hybrid search result to the scene vector database, and send the hybrid search result to corresponding maintenance personnel in combination with a disease identification result;

the scene vector database is a composite relational database and contains a plurality of records, and each record comprises a URL *url* of an image acquired, a longitude and latitude [*lng*, *lat*] at which the image is acquired, an azimuth *azi* at which the image is acquired, a scene classification result *c* of the image, an object ID of the image, a feature vector *v* obtained after performing feature extraction on the image, a creation time *CREATETIME* of the record, whether the record is reviewed by a reviewer *CHECK*, and whether the record is marked for deletion *DELETE*;

performing the hybrid search comprises:

performing preliminary screening in the scene vector database based on a longitude, a

latitude, and an azimuth in the positioning data to obtain a record data set; and

calculating a cosine similarity between the feature vector and the record data set, and performing secondary screening in the record data set based on the cosine similarity to obtain a hybrid search result set; and

updating the hybrid search result to the scene vector database comprises:

if the hybrid search result set is empty, manually reviewing and then adding the image data, the scene classification result, the feature vector, and the positioning data to the scene vector database;

if the hybrid search result set is not empty and object IDs and scene classes in the hybrid search result set are all the same, updating the image data, the scene classification result, the feature vector, and the positioning data to the scene vector database; or

if the hybrid search result set is not empty and different object IDs exist in the hybrid search result set, performing arithmetic averaging on feature vectors in the hybrid search result set to obtain a new feature vector, performing a hybrid search again in the scene vector database to obtain a new hybrid search result set, performing voting to obtain a corrected object ID and scene class, and updating the corrected object ID and scene class to the scene vector database.

4. The positioning system for an inspection scene according to claim 3, wherein the edge device comprises an industrial computer, a high-definition camera, and a navigation and positioning receiver; the high-definition camera and the navigation and positioning receiver are connected to the industrial computer; and the industrial computer to provides data transmission, image feature extraction, and power supply functions.