BEIDOU MULTI-SOURCE FUSION POSITIONING METHOD IN DISASTER ENVIRONMENT

TECHNICAL FIELD

The present invention relates to a multi-source fusion positioning method, and is particularly a Beidou multi-source fusion positioning method in a disaster environment.

BACKGROUND

In recent years, various natural disasters have occurred frequently. In disaster environments and emergency scenarios, observation environments are complex, and factors affecting positioning are highly uncertain, so that high-precision positioning for rescue vehicles is crucial for on-site disaster rescue operations. In various positioning methods, GNSS (Global Navigation Satellite System) can provide all-weather global reference positioning. However, the applicability and operability of navigation and positioning receiver devices in disaster emergency environments will be reduced, and accuracy is prone to reduction due to physical occlusion, electromagnetic interference and other problems in complex disaster environments. IMU (Inertial Measurement Unit) may provide high-precision pose derivation within a short period, but errors thereof may be gradually accumulated and increased over time. GNSS/IMU integrated navigation may achieve mutual complementation to a certain extent, but still cannot meet user requirements under conditions such as a long-duration GNSS outage. Visual sensors may be used for autonomous positioning and map construction, and are suitable for unknown and complex environments in disasters, thereby greatly compensating for the shortcomings of GNSS/IMU integrated navigation. Therefore, multi-source fusion positioning that integrates a camera with GNSS and IMU to improve system positioning accuracy has become a research hotspot in recent years.

Zhu et al. utilized a prior feature map to assist a GNSS/INS/Vision integrated navigation system, which incorporates visually matched landmarks and lane line constraints into an algorithm to achieve continuous high-precision positioning in complex urban environments. However, the system performs map construction and positioning sequentially, and neglects effects of map updates and sudden changes in road conditions on real-time positioning, leading to failure to guarantee the reliability and real-time performance of final positioning. Zhai and Yua proposed a semantic-based dynamic adaptive Kalman filtering fusion algorithm, which is used for vision-

assisted GNSS positioning to form an RTK/Vision loosely coupled system. Although incorporating semantic information into a filtering algorithm by deep learning during SLAM can improve positioning accuracy, positioning requirements under a long-duration GNSS outage remain unaddressed. Narula et al. proposed an algorithm that integrates a GNSS position estimate into a VSLAM framework for joint optimization positioning, which matches and updates pose results of an improved GNSS positioning model with Vision, and then constructs a high-precision map. The introduction of high-precision GNSS positioning does improve system positioning accuracy. However, the integrated navigation lacks consideration of a relationship between factors affecting GNSS positioning and visual and map points, resulting in failure to achieve comprehensive pose deviation reduction.

Gong et al. proposed a tightly integrated GNSS and VSLAM method on a manifold, which utilizes asynchronous tracking based on motion decomposition to fuse measured values of GNSS pseudoranges and visual feature points, so as to adapt to observed values of heterogeneous sensors. The fusion method can significantly improve system positioning accuracy and is superior to the loosely coupled method in performance. Chu et al. proposed an EKF-based GNSS/Vision/IMU integrated navigation algorithm, which may perform multi-state update solution. However, EKF only utilizes the most recent state estimate, and nonlinear factors representing measured values cannot be well re-linearized. Zhang and Wei proposed a factor graph-based GNSS/INS/OD integrated navigation algorithm, which improves system robustness by adding or deleting corresponding factor nodes based on the validity of sensors. Chen and Xu et al. proposed a factor graph-based GNSS/Vision/IMU integrated navigation algorithm, which takes GNSS-derived positions as prior information, introduces the information into factor graph nodes and provides constraints for factors of the VIO (Visual-Inertial Odometry) system. The algorithm can further constrain global trajectory accuracy. However, a situation where the GNSS outage or interference in complex environments leads to failure of positioning is not considered, thereby failing to guarantee robustness.

In summary, the prior art has the following shortcomings:

(1) In existing GNSS/IMU/Vision integrated navigation, open-source maps are commonly used. However, such maps have update lag and insufficient accuracy, and are particularly unable to reflect changes in surrounding scenes in disaster environments in a timely manner. Although the visual algorithm is continuously improved, the positioning characteristics of GNSS and the

effective information extracted from visual observations are not closely integrated, thereby ultimately failing to further provide reasonable and effective visual factors for global positioning.

(2) In existing GNSS/IMU/Vision integrated navigation, factor graph optimization is employed to improve positioning accuracy. However, disaster environments are highly uncertain, with frequent environmental changes. Existing algorithms do not take positioning failures or errors easily caused by dynamic object interference in changing environments into account. In recent years, factor graphs have been gradually used in fusion algorithms. However, existing factor graph-based fusion does not fully utilize a feedback effect of the global map on factor graph optimization, thereby ultimately resulting in low accuracy and robustness of the algorithm.

SUMMARY

Objective of invention: the technical problem to be solved by the present invention is to provide a Beidou multi-source fusion positioning method in a disaster environment aiming at the defects in the prior art.

In order to solve the above technical problem, the present invention discloses a Beidou multi-source fusion positioning method in a disaster environment, which includes the following steps:

step 1: acquiring images using a camera on a carrier to be positioned in the disaster environment, and constructing a VIO factor based on point and line feature extraction;

step 2: constructing a map and performing refined correction on the map under a GNSS/IMU constraint to obtain a global map; and

step 3: aaugmenting visual factors using the global map, and performing global pose optimization for a GNSS/INS/vision integrated navigation system to complete Beidou multi-source fusion positioning in the disaster environment.

The constructing the VIO factor based on the point and line feature extraction in the step 1, specifically includes:

step 1-1: performing the point and line feature extraction and keyframe determination;

step 1-2: constructing a visual feature reprojection residual factor;

step 1-3: constructing an IMU pre-integration residual factor; and

step 1-4: constructing the VIO factor.

The performing the point and line feature extraction and the keyframe determination in the step 1-1, specifically includes:

step 1-1-1: point and line feature extraction: according to the acquired images, obtaining point and line features using FAST point feature extraction and LSD line feature extraction; and matching the point and line features by a BRIEF descriptor, and eliminating mismatched results of the point and line features by a RANSAC method;

step 1-1-2: keyframe determination: setting the acquired images as a set of consecutive frames $F = \{F_1, F_2, F_3, \ldots, F_i, \}$, wherein F_i represents an $i\text{-}th$ frame, and each frame of image F_i includes the following parameters: a corresponding timestamp T_i and a number b_i of point and line features contained in the frame of image, and K_n represents an $n\text{-}th$ keyframe, determining whether the $i\text{-}th$ frame of image F_i is a keyframe according to the following criteria:

a first criterion, in which a time interval between the $i\text{-}th$ frame of image F_i and an adjacent keyframe is greater than a specified threshold, which means that $T_i - T_{K_n} > T_0$, T_{K_n} represents a timestamp of a latest keyframe, and T_0 is a set time-difference threshold;

a second criterion, in which the number of point and line features contained in the keyframe is greater than a specified threshold, which means that $b_i > b_0$, and b_0 is a set number threshold of point and line features; and

directly discarding frames determined as non-keyframes; and

step 1-1-3: feature matching and filtering: after identifying the keyframe, performing feature matching using a KNN algorithm, which means that: for one feature in one frame of image, a most similar feature is searched for in another frame of image, and then the features are regarded as a matching pair; and during the above matching process, eliminating mismatches using a RANSAC algorithm.

Further, the constructing the visual feature reprojection residual factor in the step 1-2, refers to constructing a visual feature reprojection residual model, and taking a distance between a projection point of a map point in an image and a corresponding observation point as a visual feature reprojection residual, and specifically includes:

step 1-2-1: point feature reprojection residual, which is specifically as follows:

for an $l\text{-}th$ feature point L, normalized coordinates thereof observed in an $i\text{-}th$ camera coordinate system corresponding to the $i\text{-}th$ frame F_i are set as $\hat{L}_l^{c_i}$, normalized coordinates of the same feature point observed in a $j\text{-}th$ camera coordinate system corresponding to a $j\text{-}th$ frame F_j are obtained as $\hat{L}_l^{c_j}$ by an image feature tracking method, normalized coordinates of the feature

point L transformed from the *i-th* camera coordinate system in which the feature point is first observed to the camera coordinate system *j* in which the feature point is observed again are $L_l^{c_j}$, referring to a predicted observed value, and then a visual observation residual term is a difference $r_l(\hat{z}_l^{c_j}, \mathcal{X}_{VIO})$ between the predicted value and an actual observed value, which is as follows:

$$r_l(\hat{z}_l^{c_j}, \mathcal{X}_{VIO}) = \hat{L}_l^{c_j} - L_l^{c_j}$$

wherein, $\mathcal{X}_{VIO}$ represents a global state vector of a VIO local pose in a world coordinate system, and $\hat{z}_l^{c_j}$ represents a point feature observation of the camera;

$$\mathcal{X}_{VIO} = [x_i, x_{i+1}, \dots, x_{i+f}, \lambda_i, \lambda_{i+1}, \dots, \lambda_{i+p}, o_i, o_{i+1}, \dots, o_{i+l}]$$

$$x_n = [p_n^\omega, q_n^\omega, v_n^\omega, b_{an}, b_{gn}], n \in [i, i+f]$$

wherein, x_n includes a position p_n^ω, an orientation q_n^ω, a velocity v_n^ω, an accelerometer bias b_{an} and a gyroscope bias b_{gn} of a carrier provided by the an *n-th* IMU; f, p and l represent total numbers of image frames, point features and line features in a sliding window respectively; λ_{i+m} represents an inverse depth between an $m-th$ feature point and an initial feature point; and o_{i+w} represents an orthogonal representation of a *w-th* feature line in the world coordinate system; and

step 1-2-2: line feature reprojection residual, referring to taking a distance from a line endpoint observed in a camera plane to a reprojected line as the residual, which is specifically as follows:

an *i-th* camera, which refers to a camera capturing the *i-th* frame F_i, is allowed to observe a line L_k, and the line L_k is projected onto a normalized plane of the *i-th* keyframe to obtain a projected line equation $l_k^{c_i}$; and two line endpoints observed in the camera plane are $\hat{s}_k^{c_i} = [u_s \quad v_s \quad 1]^T$ and $\hat{e}_k^{c_i} = [u_e \quad v_e \quad 1]^T$ respectively, $d(x, L)$ is used to represent a distance from a point x to a line L, and then the line feature reprojection residual $r_{L_i}(\hat{s}_i^{c_j}, \hat{e}_i^{c_j}, \mathcal{X}_{VIO})$ is:

$$r_{L_i}(\hat{s}_i^{c_j}, \hat{e}_i^{c_j}, \mathcal{X}_{VIO}) = \begin{bmatrix} d(\hat{s}_k^{c_i}, l_k^{c_i}) \\ d(\hat{e}_k^{c_i}, l_k^{c_i}) \end{bmatrix}$$

.

Further, the constructing the IMU pre-integration residual factor in the step 1-3, specifically includes:

step 1-3-1: constructing an IMU pre-integration residual model, which is specifically as

follows:

IMU integration between two adjacent keyframes is performed in an IMU carrier coordinate system of a previous keyframe F_i to obtain a relative motion increment independent of a state of the keyframe F_i, wherein a change in an IMU bias is handled by first-order linear approximation; and in the IMU coordinate system, IMU measurement equations are:

$$\hat{a}_t = a_t + b_{a_t} + R_w^{b_t} g^w + n_a$$

$$\hat{w}_t = w_t + b_{w_t} + n_w$$

wherein, $\hat{a}_t$ is a measured value of acceleration at a time t, a_t is a noise-free acceleration at the time t, b_{a_t} is an accelerometer bias at the time t, $R_w^{b_t}$ is a coordinate transformation matrix, g^w is a gravitational acceleration, and n_a is accelerometer white noise; and $\hat{w}_t$ is a measured value of gyroscope at the time t, w_t is a noise-free angular velocity at the time t, b_{w_t} is a gyroscope bias at the time t, and n_w is gyroscope white noise;

integration is performed according to the above IMU measurement equations and coordinate system transformation is performed, and then IMU pre-integration is as follows:

$$\begin{cases} \alpha_{b_j}^{b_i} = \iint_{t\in[i,j]} \left(R_{b_t}^{b_i} \left(\hat{a}_t - b_{a_t} - n_a \right) \right) \mathrm{dt}^2 \\ \beta_{b_j}^{b_i} = \int_{t\in[i,j]} \left(R_t \left(\hat{a}_t - b_{a_t} - n_a \right) \right) \mathrm{dt} \\ \gamma_j^i = \int_{t\in[i,j]} \left(\frac{1}{2} \Omega \left(\hat{w}_t - b_{w_t} - n_w \right) q_t^{b_i} \right) \mathrm{dt} \end{cases}$$

wherein, $\alpha_{b_j}^{b_i},\ \beta_{b_j}^{b_i}\ and\ \gamma_j^i$ are relative position, velocity and rotation integrated from t_i to t_j, and an intermediate term $\Omega(w)$ is calculated as follows:

$$\Omega(w) = \begin{bmatrix} -[w]_\times & w \\ -w^T & 0 \end{bmatrix}$$

wherein, w represents the measured value of gyroscope, $\times$ represents an operator $(\cdot)\times$ mapping a vector to a skew-symmetric matrix thereof, and T represents a matrix transpose operation; and

step 1-3-2: constructing the IMU pre-integration residual factor, which is specifically as follows:

after IMU pre-integration, a residual $r_B \left(\hat{z}_{b_j}^{b_i}, \mathcal{X}_{VIO} \right)$ between a system state and IMU measurement is as follows:

$$r_B\left(\hat{z}_{b_j}^{b_i}, \mathcal{X}_{VIO}\right) = \left[\delta\alpha_{b_j}^{b_i} \quad \delta\beta_{b_j}^{b_i} \quad \delta\gamma_j^i \quad \delta b_a \quad \delta b_g\right]^T$$

wherein, δ represents a difference between a measured value and a predicted value, and $\hat{z}_{b_j}^{b_i}$ represents an IMU observation.

Further, the constructing the VIO factor in the step 1-4, specifically includes:

according to the point feature reprojection residual, the line feature reprojection residual and the IMU pre-integration residual, performing joint optimization by a least squares method, which is specifically as follows:

$$\mathcal{X}_{VIO}{}^* = \underset{\mathcal{X}_{VIO}}{argmin}\{\left\|r_p - H_p\mathcal{X}_{VIO}\right\|^2 + \sum_{(i,j)\in B}\left\|r_B\left(\hat{z}_{b_j}^{b_i}, \mathcal{X}_{VIO}\right)\right\|_{\Sigma_{b_i}^{b_i}}^2 + \sum_{(i,j)\in C_1}\left\|r_l\left(\hat{z}_l^{c_j}, \mathcal{X}_{VIO}\right)\right\|_{\Sigma_l^{c_j}}^2 + \sum_{(i,j)\in C_2}\left\|r_{L_i}\left(\hat{z}_{L_i}^{c_j}, \mathcal{X}_{VIO}\right)\right\|_{\Sigma_{L_i}^{c_j}}^2\}$$

wherein, $\mathcal{X}_{VIO}{}^*$ represents a corresponding state value at an optimal solution, r_p and H_p are prior information, r_B is the IMU pre-integration residual, r_l is the point feature reprojection residual, and r_{L_k} is the line feature reprojection residual; and B is a measured value set of IMU, C_1 and C_2 are measured value sets of point and line features, and $\Sigma_{b_j}^{b_i}$, $\Sigma_l^{c_j}$ and $\Sigma_{L_k}^{c_j}$ are corresponding covariance matrices; and

according to the above optimization result, constructing a model using a VIO position increment and an attitude increment from a time t_{i-1} to a time t_i to obtain a measurement model as follows:

$$y_{\mathrm{i,VO}} = \begin{Bmatrix} y_{\delta R,i} \\ y_{\Delta p,i} \end{Bmatrix} = \begin{Bmatrix} exp\left(n_{\delta R,i}\right) R_{i-1}^T R_i \\ R_{i-1}^T(P_i - P_{i-1}) + n_{\Delta p,i} \end{Bmatrix}$$

wherein, $y_{\delta R,i}$ and $y_{\Delta p,i}$ are measured values of attitude increment and position increment of visual odometry, $n_{\delta R,i}$ is measurement noise of attitude increment, R_{i-1} and R_i are measured values of attitudes at adjacent times, P_{i-1} and P_i are measured values of positions at adjacent times, and $n_{\Delta p,i}$ is measurement noise of position increment; and then the VIO factor $r_{VIO}\left(\hat{z}_{i,\mathrm{VIO}}, \chi\right)$ is:

$$r_{VIO}\left(\hat{z}_{i,\mathrm{VIO}}, \chi\right) = \begin{Bmatrix} r_{\delta R,i} \\ r_{\Delta p,i} \end{Bmatrix} = \begin{Bmatrix} log\left(y_{\delta R,i} - R_{i-1}^T R_i\right) \\ y_{\Delta p,i} - R_{i-1}^T(P_i - P_{i-1}) \end{Bmatrix}$$

wherein, $\hat{z}_{i,\mathrm{VIO}}$ represents a measured value of VIO system, and $r_{\delta R,i}$ and $r_{\Delta p,i}$ represent an attitude-increment residual and a position-increment residual respectively.

Further, the constructing the map and performing the refined correction on the map under the GNSS/IMU constraint in the step 2, specifically includes:

step 2-1: preliminary map construction, which is specifically as follows:

after keyframe determination, a keyframe selected from a camera image sequence is transformed into a three-dimensional space by an LSS method, and a pixel point (u, v) in the image is mapped to a corresponding global coordinate system point (X_w, Y_w, Z_w) by a method as follows:

$$[u \quad v \quad 1]^T = KRT[X_w \quad Y_w \quad Z_w \quad 1]^T$$

wherein, R is a camera rotation matrix, T is a camera translation vector, and K is a camera intrinsic parameter matrix;

$$R = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & cos\,\theta & -sin\,\theta & 0 \\ 0 & sin\,\theta & cos\,\theta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$T = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -\dfrac{h}{sin\,\theta} \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

wherein, h is a camera height, and θ is a camera pitch angle;

step 2-2: open region identification, wherein, by a convolutional neural network semantic segmentation method, a mapped element, which refers to the map point in the three-dimensional space, and also refers to the global coordinate system point in the step 2-1), is assigned a label, and a density of objects with the same label is determined, which is specifically as follows:

the open region is defined, which means that a density of tall buildings in the region with an aspect ratio greater than 1: 1 and buildings meeting a height requirement in the same region is less than a preset threshold;

when a road is identified in the open region, a point on the road is selected as a center of the open region and assigned a new label: an open region reference point; and when the open region belongs to a non-traversable region, a building closest to a traversable road in the open region is assigned the above new label: the open region reference point; and

after using the SSD convolutional neural network semantic segmentation method, a detected object in the image is assigned a classification label and two-dimensional position information; and according to the LSS method in the step 2-1, the above classification label is mapped to a

three-dimensional position;

step 2-3: refinement under GNSS/IMU constraint, which means that, according to prior training results of the convolutional neural network semantic segmentation algorithm, during real-time positioning, the open region identification is performed simultaneously with the map construction, and according to the label of the open area reference point, when passing through the region or approaching the region which refers to being impassable, constraint correction is performed using a loosely coupled GNSS/IMU navigation solution;

step 2-4: global high-precision map optimization construction, which is specifically as follows:

whether continuously generated local maps overlap is determined through loop closure detection, and when the continuously generated local maps overlap, the local maps are updated according to nonlinear optimization, which is implemented by solving the following nonlinear least squares problem via a Gauss-Newton method:

$$min_{x_1,x_2,\ldots,x_n} \sum_{i,j} \rho\left(e_{ij}(x_i, x_j, z_{ij})^T \Omega_{ij} e_{ij}(x_i, x_j, z_{ij})\right)$$

wherein, $e_{ij}(x_i, x_j, z_{ij})$ is an error between observation data z_{ij} from a pose x_i to a pose x_j and pose estimation, Ω_{ij} is an information matrix of the observation data, and $\rho(\cdot)$ is a robust kernel function;

when traversed regions do not overlap, a co-visibility relationship is obtained according to perceptual similarity during BEV perception, the co-visibility relationship is represented by a matrix V, and V_{ij} represents a co-visibility degree between the pose x_i and the pose x_j; and feature points P_i and P_j under similar frames F_i and F_j are searched for according to the co-visibility relationship, and corresponding feature points representing the same landmark are searched for by calculating a Euclidean distance $d(P_i, P_j)$ between the feature points, as follows:

$$d(P_i, P_j) = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2}$$

when the Euclidean distance is less than a specified threshold, it is determined that the feature points P_i and P_j are identical, and a global map is generated according to the corresponding feature points representing the same landmark found based on the co-visibility relationship; and

step 2-5: global factor expansion based on feature matching, which means that visual feature state vectors are expanded using the generated or updated map points, and new visual reprojection

residual factors are generated as auxiliary positioning information for a global factor graph optimization process; and when performing visual factor expansion using the map, only the feature point reprojection residuals are expanded.

Further, the refinement under GNSS/IMU constraint in the step 2-3, is specifically as follows:

when passing through the open region reference point, GNSS and IMU integrated navigation yields a high-precision pose to correct a previously predicted pose of the point, and further provides a new constraint for map construction after the time: position and attitude information provided by GNSS/IMU is transformed from the world coordinate system to the camera coordinate system, and further transformed to a pixel coordinate system through a camera intrinsic matrix, and then jointly optimized with a previously predicted visual target point at the same position, which refers to minimizing a value η of the following formula:

$$\eta = argmin\frac{1}{2}\sum_{i=1}^{n}\left\|u_i - \frac{1}{z}K\left(R_w^c(P - P_c)\right)\right\|^2$$

wherein, P is target point coordinates for joint optimization between the GNSS/IMU observation point and the map point, K is a camera intrinsic parameter, z is a feature point depth, R_w^c and P_c are prior transformation matrix information and map feature point position information during visual image processing and map point generation respectively, and u_i is an observed value composed of coordinates positioned by GNSS/IMU and subjected to coordinate system transformation to the pixel coordinate system and a disparity of left and right images.

Further, the expanding the visual factors using the global map in the step 3, specifically includes:

step 3-1: construction of pseudorange factor and Doppler velocity measurement factor, which is specifically as follows:

a pseudorange measurement model is obtained during single-point positioning solution, a pseudorange factor residual refers to an error between a predicted value and an observed value of a pseudorange, and according to the pseudorange measurement model, a pseudorange residual of a satellite s at the time t_i is calculated as r_{Pse};

GNSS velocity measurement results are introduced to constrain position variables of adjacent epochs to obtain a Doppler velocity measurement model; and a Doppler velocity factor residual refers to an error between a predicted value and an observed value of a velocity, and a residual of Doppler velocity measurement at the time t_i is calculated as r_D;

step 3-2: global factor fusion and initialization, which is specifically as follows:

whether a current frame is the keyframe is determined, which refers to determining the keyframe according to the method in the step 1; and when GNSS satellite observation data are input, the data are constructed as the keyframe and added to the sliding window; and

a coordinate system of the visual-inertial odometry is initialized and aligned with a GNSS system;

step 3-3: sliding window optimization, wherein, after the initialization in the step 3-2 is successful, subsequent observation is added to a factor graph for overall optimization based on the factor graph; and the sliding window is introduced to limit a graph size, which is specifically as follows:

when a new frame of observation arrives, a corresponding system state-variable node is constructed therefor first and added to the factor graph, and a corresponding error factor is constructed according to a measurement type thereof and connected with a relevant state-variable node; and when the window is full at the time, a node corresponding to an oldest frame in the window is marginalized, the node is removed and a prior constraint factor is constructed for a node related to the node in the window to participate in subsequent global state optimization of the system; and

step 3-4: global optimal pose, wherein, after the above steps, a high-precision pose of a carrier to be positioned is obtained, and according to the high-precision pose, the global map is further corrected to achieve real-time update of the map, and used for calculating the next epoch.

Further, the sliding window optimization in the step 3-3, means that an optimization problem constructed by the factors in the sliding window is:

$$\chi^* = \underset{\chi}{argmin}\{ \left\|\boldsymbol{r_p} - \boldsymbol{H_p}\chi\right\|^2 + \sum_{s\in \boldsymbol{P_{se}}} \left\|\boldsymbol{r_{Pse}}\left(\hat{\boldsymbol{z}}^s_{r,i}, \boldsymbol{\chi}\right)\right\|^2_{\boldsymbol{\Sigma}^s_{r,i}} + \sum_{i\in D}\left\|\boldsymbol{r_D}\left(\hat{\boldsymbol{z}}_{i,\mathrm{DV}}, \boldsymbol{\chi}\right)\right\|^2_{\boldsymbol{\Sigma}_{i,\mathrm{DV}}}$$

$$+ \sum_{i\in V}\left\|\boldsymbol{r}_{\mathrm{VIO}}\left(\hat{\boldsymbol{z}}_{i,\mathrm{VIO}}, \boldsymbol{\chi}\right)\right\|^2_{\Sigma_{i,\mathrm{VIO}}} + \sum_{(i,j)\in M}\left\|\boldsymbol{r}_i\left(\hat{\boldsymbol{z}}_i^{\boldsymbol{cj}}, \boldsymbol{\chi}\right)\right\|^2_{\boldsymbol{\Sigma}_i^{cj}} \}$$

$$\chi = \left[x'_i, x'_{i+1}, \dots, x'_{i+f}, \lambda_i, \lambda_{i+1}, \dots, \lambda_{i+p}, o_i, o_{i+1}, \dots, o_{i+l}, t_{r,s}, T\right]$$

$$x'_n = \left[p^e_{b_n}, q^e_{b_n}, v^e_{b_n}\right], n \in [i, i+f]$$

wherein, χ^* represents a corresponding state at the optimal solution obtained, χ represents the global state vector in the sliding window, which is relative to a state quantity in the VIO system, x'_n represents a system state at the time t_i in the sliding window, and $p^e_{b_n}$, $q^e_{b_n}$ and $v^e_{b_n}$ represent a

position, an attitude and a velocity at the time t_i in a coordinate system e respectively; $t_{r,s}$ represents **r**eceiver clock offsets of different satellite systems at the time t_i, r and s are a receiver and a satellite system respectively, T represents a tropospheric delay at the time t_i, and r_i is a visual point feature reprojection residual of map expansion; Pse is a set of measured values of pseudoranges in the sliding window; D is a set of measured values of Doppler velocities in the sliding window; V is a set of measured values of VIOs in the sliding window, and M is a set of selected point features of map points in the sliding window; and $\Sigma_{r,i}^{s}$, $\Sigma_{i,\mathrm{DV}}$, $\Sigma_{\mathrm{i,VIO}}$ and $\Sigma_{i}^{c_j}$ represent covariance matrices corresponding to a measure value of pseudorange $\hat{\boldsymbol{z}}_{r,i}^{s}$, a measured value of Doppler velocity $\hat{\boldsymbol{z}}_{i,\mathrm{DV}}$, a measured value of visual-inertial odometry $\hat{\boldsymbol{z}}_{i,\mathrm{VIO}}$ and a measured value of map point feature $\hat{\boldsymbol{z}}_{i}^{\boldsymbol{c_j}}$ respectively.

Beneficial effects:

(1) Aiming at the problems of lag and low accuracy in high-precision map construction by the GNSS/IMU/Vision integrated navigation system in the disaster environment, according to the strong perception capability and environmental adaptability of the visual sensor, the preliminary map construction is performed by the LSS perception algorithm, the surrounding information is identified through the convolutional neural network, and according to semantic identification and detection results thereof, the GNSS/IMU constraints are introduced during open region map construction, thereby achieving real-time performance and refinement of disaster scene map construction.

(2) Aiming at the problems of low robustness and accuracy caused by dynamic object interference in the disaster environment, the factor graph is generally optimized under the assumption of static environment; however, various forms of dynamic changes in the disaster environment degrade the performance of the optimization algorithm, and meanwhile, the data mismatches in the dynamic environment may also affect the accuracy of optimization results; and in the present invention, through global map prediction generated in real time, appropriate visual points are selected for expansion, and factors in a factor graph are added and removed to adapt to dynamic environmental changes under natural disasters, thereby improving robustness and accuracy of a rescue-vehicle localization system.

BRIEF DESCRIPTION OF THE DRAWINGS

The present invention is further described in detail hereinafter with reference to the drawings and specific embodiments, and the advantages of the above and/or other aspects of the present invention will be clearer.

FIG. 1 is a general flow chart of the present invention.

DETAILED DESCRIPTION

Existing open-source electronic maps have update lag, and significant environmental changes in natural disaster scenarios result in untimely updates. Although visual sensors may be used to perceive surrounding environments and generate maps, a relationship between GNSS and IMU positioning and map points in a global map has not been fully exploited. Therefore, further research is required to improve system positioning accuracy. The present invention aims to solve the problems of map update lag in disaster environments and low positioning accuracy of vehicles in complex disaster environments. The core content of the present invention consists of 3 stages: (1) construction of VIO factor based on point and line feature extraction; (2) high-precision map construction and refined correction under GNSS/IMU constraints; and (3) expansion of visual factors using global map and global pose optimization for GNSS/INS/Vision integrated navigation system.

The present invention provides a Beidou multi-source fusion positioning method in a disaster environment, a general flow chart is shown in FIG. 1, and a specific technical solution is as follows.

(1) Construction of VIO factor based on point and line feature extraction

1) Point and line feature extraction and keyframe determination

First step: point and line feature extraction. After a camera acquires images, the present invention performs point feature extraction using FAST (Features From Accelerated Segment Test) and performs line feature extraction using LSD (Line Segment Detector). Extracted point and line features are matched using a BRIEF (Binary Robust Independent Elementary Features) descriptor, and mismatches of point and line features are eliminated using RANSAC (Random Sample Consensus).

Second step: keyframe determination. In the case of a large number of images to be identified, to reduce a computational burden for a real-time performance, keyframe selection is performed. The present invention involves a set of consecutive frames $F = \{F_1, F_2, F_3, \ldots, F_n, \}$, each frame of image has a corresponding timestamp T_i and a number b_i of point and line features contained

therein, wherein i represents an i-*th* frame, and K_n represents an n-*th* keyframe.

In the present invention, the keyframe determination is based on three criteria: first, a time interval between two adjacent keyframes is greater than a specified threshold, which means that $T_i - T_{K_n} > T_0$, wherein T_{K_n} is a timestamp of the most recent keyframe, and T_0 is a preset time difference threshold; second, sliding window graph optimization executed by a global map construction and optimization module has been completed, and a Boolean value is used to mark whether the optimization is finished, so that the system constructs as many keyframes as possible to improve global map accuracy; and third, the number of point line features contained therein is greater than a specified threshold, which means that $b_i > b_0$, wherein b_0 is a preset threshold, and image frames that are not keyframes are directly discarded to reduce a computational load.

Third step: feature matching and filtering. After finding the keyframes, to find a matching relationship of corresponding image features between the frames, the present invention introduces a KNN (K-Nearest Neighbors) algorithm for matching. For one feature in one image, the present invention searches for the most similar feature (the nearest neighbor) in another image, and then regards the features as a matching pair. Considering possible mismatches, a RANSAC (Random Sample Consensus) algorithm is introduced to eliminate the mismatches. The RANSAC algorithm randomly selects data subsets (samples), constructs a model using these samples, and performs consistency checking on an entire dataset, which refers to setting an original dataset as D, a sample data subset as S, and a model to be tested as M. The sample S is randomly selected from the dataset D, and then model parameters are computed using the sample S, which are represented as $M(S)$. When applied to the entire dataset, the present invention calculates the consistency of non-sample data points d with the model $M(S)$, which is represented as $C(d, M(S))$, and then determines which data points are consistent with the model based on a threshold, thereby identifying and eliminating the mismatches.

2) Visual feature reprojection residual factor.

A visual feature reprojection residual model is constructed. A distance between a projection point of a map point in an image and a corresponding observation point is taken as a visual feature reprojection residual.

First step: point feature reprojection residual. For an 1-*th* feature point L, normalized coordinates thereof observed in an i-*th* camera coordinate system are $\hat{L}_l^{c_i}$, normalized coordinates of the same feature point observed in a j-*th* camera coordinate system are obtained as $\hat{L}_l^{c_j}$ according

to feature matching, normalized coordinates of the feature point L transformed from the i-*th* camera coordinate system in which the feature point is first observed to the j-*th* camera coordinate system in which the feature point is observed again according to a principle of projection and back-projection are $L_l^{c_j}$ (referring to a predicted observed value), and then a visual observation residual term is a difference $r_l(\hat{z}_l^{c_j}, \mathcal{X}_{VIO})$ between the predicted value and an actual observed value, which is as follows:

$$r_l(\hat{z}_l^{c_j}, \mathcal{X}) = \hat{L}_l^{c_j} - L_l^{c_j}$$

Second step: line feature reprojection residual. The line feature reprojection residual of the present invention refers to taking a distance from a line endpoint observed in a camera plane to a reprojected line as the residual. An i-*th* camera observes a line L_k, and the line L_k is projected onto a normalized plane of an i-*th* keyframe to obtain a projected line equation $l_k^{c_i}$. Two line endpoints observed in the camera plane are $\hat{s}_k^{c_i} = [u_s \quad v_s \quad 1]^T$ and $\hat{e}_k^{c_i} = [u_e \quad v_e \quad 1]^T$ respectively, a distance from a point x to a line L is represented by $d(x, L)$, and then the line feature reprojection residual is:

$$\boldsymbol{r}_{L_i}(\hat{\boldsymbol{s}}_{\boldsymbol{i}}^{\boldsymbol{c_j}}, \hat{\boldsymbol{e}}_{\boldsymbol{i}}^{\boldsymbol{c_j}}, \boldsymbol{\mathcal{X}}) = \begin{bmatrix} \boldsymbol{d}(\hat{\boldsymbol{s}}_{\boldsymbol{k}}^{\boldsymbol{c_i}}, \boldsymbol{l}_{\boldsymbol{k}}^{\boldsymbol{c_i}}) \\ \boldsymbol{d}(\hat{\boldsymbol{e}}_{\boldsymbol{k}}^{\boldsymbol{c_i}}, \boldsymbol{l}_{\boldsymbol{k}}^{\boldsymbol{c_i}}) \end{bmatrix}$$

3) IMU pre-integration residual factor.

First step: construction of IMU pre-integration residual model. IMU integration between two adjacent keyframes is performed in an IMU carrier coordinate system of a previous keyframe to obtain a relative motion increment independent of a state of the keyframe , wherein a change in an IMU bias is handled by first-order linear approximation. In an IMU coordinate system, IMU measurement equations are:

$$\hat{a}_t = a_t + b_{a_t} + R_w^{b_t} g^w + n_a$$

$$\hat{w}_t = w_t + b_{w_t} + n_w$$

wherein, $\hat{a}_t$ is a measured value of acceleration at a time t, a_t is a noise-free acceleration at the time t, b_{a_t} is an accelerometer bias at the time t, $R_w^{b_t}$ is a coordinate transformation matrix, g^w is a gravitational acceleration, and n_a is accelerometer white noise; and $\hat{w}_t$ is a measured value of gyroscope at the time t, w_t is a noise-free angular velocity at the time t, b_{w_t} is a gyroscope bias at the time t, and n_w is gyroscope white noise.

Integration is performed according to the above equations and coordinate system

transformation is performed, and IMU pre-integration equations may be written as follows:

$$\begin{cases} \boldsymbol{\alpha}_{b_j}^{b_i} = \iint_{t\in[i,j]} \left(R_{b_t}^{b_i}\left(\hat{a}_t - b_{a_t} - n_a\right)\right) \mathrm{dt}^2 \\ \boldsymbol{\beta}_{b_j}^{b_i} = \int_{t\in[i,j]} \left(R_t\left(\hat{a}_t - b_{a_t} - n_a\right)\right) \mathrm{dt} \\ \boldsymbol{\gamma}_j^i = \int_{t\in[i,j]} \left(\frac{1}{2}\Omega\left(\hat{w}_t - b_{w_t} - n_w\right) q_t^{b_i}\right) \mathrm{dt} \end{cases}$$

wherein, α、β、γ are relative position, velocity and rotation integrated from a time i to a time j, and Ω is:

$$\Omega(w) = \begin{bmatrix} w \\ 0 \end{bmatrix}_R = \begin{bmatrix} -[w]_\times & w \\ -w^T & 0 \end{bmatrix}$$

Second step: construction of IMU pre-integration residual factor.

After IMU pre-integration, a residual formula between a system state and IMU measurement is:

$$r_{\boldsymbol{B}}\left(\hat{\boldsymbol{z}}_{b_j}^{b_i}, \boldsymbol{\chi}\right) = \begin{bmatrix} \delta\boldsymbol{\alpha}_{b_j}^{b_i} & \delta\boldsymbol{\beta}_{b_j}^{b_i} & \delta\boldsymbol{\gamma}_j^i & \delta b_a & \delta b_g \end{bmatrix}^T$$

wherein, δ represents a difference between a measured value and a predicted value.

4) Construction of VIO factor.

The present invention utilizes the point feature observation residual, the line feature observation residual and the IMU residual obtained above, and employs the least squares method for joint optimization:

$$\chi^* = \underset{\chi}{argmin}\{ \left\|\boldsymbol{r_p} - \boldsymbol{H_p}\chi\right\|^2 + \sum_{(i,j)\in B} \left\|\boldsymbol{r_B}\left(\hat{\boldsymbol{z}}_{b_j}^{b_i}, \boldsymbol{\chi}\right)\right\|_{\boldsymbol{\Sigma}_{b_i}^{b_i}}^2 + \sum_{(i,j)\in C_1} \left\|\boldsymbol{r_l}\left(\hat{\boldsymbol{z}}_{\boldsymbol{l}}^{\boldsymbol{c_j}}, \boldsymbol{\chi}\right)\right\|_{\boldsymbol{\Sigma}_l^{c_j}}^2 + \sum_{(i,j)\in C_2} \left\|\boldsymbol{r}_{\boldsymbol{L}_i}\left(\hat{\boldsymbol{z}}_{\boldsymbol{L}_i}^{\boldsymbol{c_j}}, \boldsymbol{\chi}\right)\right\|_{\boldsymbol{\Sigma}_{L_i}^{c_j}}^2 \}$$

wherein, $\boldsymbol{r_p}$ and $\boldsymbol{H_p}$ are prior information, $\boldsymbol{r_B}$ is the IMU pre-integration residual, $\boldsymbol{r_l}$ is the visual point feature reprojection residual, and $\boldsymbol{r}_{\boldsymbol{L_k}}$ is the visual line feature reprojection residual; $\boldsymbol{B}$ is a measured value set of IMU, and $\boldsymbol{C_1}$ and $\boldsymbol{C_2}$ are measured value sets of point and line features; and $\boldsymbol{\Sigma}_{b_j}^{b_i}$, $\boldsymbol{\Sigma}_l^{c_j}$ and $\boldsymbol{\Sigma}_{L_k}^{c_j}$ are corresponding covariance matrices thereof.

Since visual-inertial odometry is accurate within a small range, according to the above optimization result, a model is constructed using a VIO position increment and an attitude increment from a time i-1 to a time i to obtain a measurement model:

$$\boldsymbol{y}_{\mathrm{i,VO}} = \begin{Bmatrix} \boldsymbol{y}_{\delta R,i} \\ \boldsymbol{y}_{\Delta p,i} \end{Bmatrix} = \begin{Bmatrix} exp\left(n_{\delta R,i}\right) R_{i-1}^{T} R_i \\ R_{i-1}^{T}(P_i - P_{i-1}) + n_{\Delta p,i} \end{Bmatrix}$$

wherein, $\boldsymbol{y}_{\delta R,i}$ and $\boldsymbol{y}_{\Delta p,i}$ are measured values of attitude increment and position increment of visual odometry, $n_{\delta R,i}$ is measurement noise of attitude increment, R_{i-1} and R_i are measured values of attitudes at adjacent times, P_{i-1} and P_i are measured values of positions at adjacent times, and $n_{\Delta p,i}$ is measurement noise of position increment; and

then the VIO factor is:

$$\boldsymbol{r}_{VIO}\left(\hat{\boldsymbol{z}}_{i,\mathrm{VIO}}, \boldsymbol{\chi}\right) = \begin{Bmatrix} \boldsymbol{r}_{\delta R,i} \\ \boldsymbol{r}_{\Delta p,i} \end{Bmatrix} = \begin{Bmatrix} log\left(\boldsymbol{y}_{\delta R,i} - R_{i-1}^{T} R_i\right) \\ \boldsymbol{y}_{\Delta p,i} - R_{i-1}^{T}(P_i - P_{i-1}) \end{Bmatrix}$$

(2) High-precision map construction and refined correction under GNSS constraints

First step: preliminary map construction. After camera keyframe selection, the camera images are transformed into a three-dimensional space by an LSS (a 2D-to-3D algorithm, Lift-Splat-Shoot) method, which means that two-dimensional images are subjected to inverse perspective transformation by introducing depth estimation, and 2D pixel points (u, v) are mapped to corresponding global coordinate system points (X_w, Y_w, Z_w):

$$[u \quad v \quad 1]^T = KRT[X_w \quad Y_w \quad Z_w \quad 1]^T$$

wherein, R is a camera rotation matrix, T is a camera translation vector, and K is a camera intrinsic parameter matrix.

$$R = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & cos\,\theta & -sin\,\theta & 0 \\ 0 & sin\,\theta & cos\,\theta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$T = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -\dfrac{h}{sin\,\theta} \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

wherein, H is a camera height, and θ is a camera pitch angle.

Camera features are associated with depth by the LSS method, thereby ultimately achieving three-dimensional perception from a spatial perspective.

Second step: open region identification. By a convolutional neural network semantic segmentation algorithm, a predicted mapped element (3D map point) is assigned a label, and a density of objects with the same label is determined.

The present invention employs an SSD convolutional neural network, and an SSD algorithm may simultaneously detect and classify objects in a single image. Firstly, datasets (images under different scenes are acquired via a vehicle-mounted camera in advance to construct tall building and road datasets) are used to train an SSD convolutional neural network model, which refers to performing convolution on feature maps at different scales, generating a series of fixed-size candidate regions, and directly performing classification and position regression on the regions to obtain information about targets to be detected (which refer to tall buildings, roads and open regions). Each feature map has a different size, which also results in candidate regions of different sizes. Specifically, SSD may generate multiple default boxes at each position. The default boxes correspond to different aspect ratios and scales, such as aspect ratios of 1: 1, 1: 2 and 2: 1, to describe the diversity of sizes and shapes of objects. For the tall buildings, boxes with greater heights are generally required.

The open region defined in the present invention is characterized by low-rise buildings and low density. Therefore, a density of tall buildings in the open region with the aspect ratio of 1: 1 in the present invention and buildings meeting a height requirement in the same region is less than a certain threshold. When a road is identified in the open region, a point on the road is selected as a center of the open region and assigned a new label “an open region reference point”; and when the open region belongs to a non-traversable region, a building closest to a traversable road in the open region is assigned the new label. After training the SSD convolutional neural network, final results obtained from visual image processing are a set of predicted boxes, and each predicted box has a clear classification label (tall building, road or open region reference point) and position information (center coordinates, width and height of the box). According to the LSS algorithm in the previous step, the labels obtained in the step may be assigned to corresponding three-dimensional positions.

Third step: refinement under GNSS/IMU constraint. According to prior training results, during real-time positioning, the open region identification is performed simultaneously with the map construction, and according to the label of “the open area reference point”, when passing through the region or approaching the region (in the case of being impassable), constraint correction is performed using a loosely coupled GNSS/IMU navigation solution. That is, when passing through the open region reference point, GNSS and IMU integrated navigation may yield a high-precision pose to correct a previously predicted pose of the point, and further provide a new

constraint for map construction after the time: position and attitude information provided by GNSS/IMU is transformed from the world coordinate system to the camera coordinate system, and further transformed to a pixel coordinate system through a camera intrinsic matrix, and then jointly optimized with a previously predicted visual target point at the same position, which refers to minimizing a value of the following formula:

$$\eta = argmin\frac{1}{2}\sum_{i=1}^{n}\left\|\boldsymbol{u}_i - \frac{1}{z}\boldsymbol{K}\left(\boldsymbol{R}_w^c(\boldsymbol{P}-\boldsymbol{P}_c)\right)\right\|^2$$

wherein, $\boldsymbol{P}$ is target point coordinates for joint optimization between the GNSS/IMU observation point and the map point, $\boldsymbol{K}$ is a camera intrinsic parameter, z is a feature point depth, $\boldsymbol{R}_w^c$ and $\boldsymbol{P}_c$ are prior transformation matrix information and map feature point position information during visual image processing and map point generation respectively, and $\boldsymbol{u}_i$ is an observed value composed of coordinates positioned by GNSS/IMU and subjected to coordinate system transformation to the pixel coordinate system and a disparity of left and right images.

By minimizing the loss function via a gradient descent method, a semi-sparse high-precision map composed of feature points with high registration accuracy may be obtained, which provides favorable observational constraints for the positioning algorithm. Through the above processing, a high-precision map containing geometric information and feature information, which is subjected to correction under the GNSS/IMU constraints, is finally obtained.

Fourth step: global high-precision map optimization construction. Whether continuously generated local maps overlap may be determined through loop closure detection, and when the continuously generated local maps overlap, the local maps may be updated according to nonlinear optimization, which may be implemented by solving the following nonlinear least squares problem via a Gauss-Newton method:

$$min_{x_1,x_2,\ldots,x_n}\sum_{i,j}\rho\left(e_{ij}(x_i,x_j,z_{ij})^T\Omega_{ij}e_{ij}(x_i,x_j,z_{ij})\right)$$

wherein, $e_{ij}(x_i,x_j,z_{ij})$ is an error between observation data z_{ij} from a pose x_i to a pose x_j and pose estimation, Ω_{ij} is an information matrix of the observation data, and $\rho(\cdot)$ is a robust kernel function, which is used to reduce an influence of outliers.

When traversed regions do not overlap, a co-visibility relationship is obtained according to perceptual similarity in a BEV (Bird's-Eye View) perception process, and the co-visibility relationship is represented by a matrix V, and V_{ij} represents a degree of co-visibility between a

pose i and a pose j. According to the co-visibility relationship, feature points P_i and P_j under similar frames F_i and F_j are searched for, and corresponding feature points representing the same landmark are searched for by calculating a Euclidean distance $d(P_i, P_j)$ between the feature points:

$$d(P_i, P_j) = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2}$$

When the Euclidean distance is less than a specified threshold, it is determined that the feature points P_i and P_j are identical, and a global map is generated according to the corresponding feature points representing the same landmark found based on the co-visibility relationship.

Fifth step: global factor expansion based on feature matching. Global map feedback lies in expanding visual feature state vectors using the generated/updated map points, and generating new visual reprojection residual factors as auxiliary positioning information for a global factor graph optimization process. Considering the real-time performance of the algorithm, when performing visual factor expansion using the map, only the feature point reprojection residuals are considered, while the line features are ignored.

$$r_i = f(X, P_i) - O_i$$

wherein, X is a visual feature state vector, P_i is a feature of a certain point in a three-dimensional map that is similar to that in an observed frame after matching via the RANSAC algorithm, $f()$ is a visual reprojection function, which maps the state vector and the map point onto the image plane, and O_i is a corresponding feature point observed in the image.

When new visual reprojection residual factors are expanded, the present invention considers two criteria to ensure the quality of factor expansion: first, a difference between a projected point of the map point on the image and an actual observation point, which refers to a reprojection error, is less than a threshold; and second, whether geometric relationships of map points observed from different viewpoints are consistent, which means that whether the geometric relationships of feature points observed from different viewpoints or at different time points are consistent.

(3) Global pose optimization for GNSS/INS/Vision integrated navigation system under global factor expansion.

The high-precision global map generated in the previous section stores high-precision geometric information of perceived environment. With the aid of the map, the present invention may construct a collinearity equation through an optical center of the camera using feature points and matched landmarks, thereby providing additional information factors to constrain a full-

system factor optimization process.

First step: construction of pseudorange factor and Doppler velocity measurement factor. During single-point positioning solution, a pseudorange measurement model may be obtained as:

$$\boldsymbol{P}_r^s = \left\|\boldsymbol{p}_s^{\mathrm{ECI}} - \boldsymbol{p}_r^{\mathrm{ECI}}\right\| + c(\boldsymbol{\zeta}_s^T \delta\boldsymbol{t} - \Delta t^s) + \boldsymbol{T}_r^s + \boldsymbol{I}_r^s + \boldsymbol{M}_r^s + \varepsilon_r^s$$

wherein, $\boldsymbol{P}_r^s$ is a pseudorange, $\boldsymbol{p}_s^{\mathrm{ECI}}$ is coordinates of a satellite s in an Earth-centered inertial coordinate system, $\boldsymbol{p}_r^{\mathrm{ECI}}$ is coordinate of a receiver r in the Earth-centered inertial coordinate system, $\boldsymbol{\zeta}_s^T$ is a 4×1 vector, in which a corresponding satellite constellation element is 1 and the other three elements are 0, Δt^s is **a satellite clock offset**, $\boldsymbol{T}_r^s$ and $\boldsymbol{I}_r^s$ represent a tropospheric delay and an ionospheric delay respectively, $\boldsymbol{M}_r^s$ is a delay caused by multipath effects, and ε_r^s is measurement noise of the pseudorange.

A pseudorange factor residual refers to an error between a predicted value and an observed value of the pseudorange. Therefore, according to the pseudorange measurement model, a pseudorange residual of the satellite s at the time i may be written as:

$$\boldsymbol{r}_{Pse}\left(\hat{\boldsymbol{z}}_{r_i}^s, \chi\right) = \left\|\boldsymbol{p}_s^{\mathrm{ECI}} - \boldsymbol{p}_{r_i}^{\mathrm{ECI}}\right\| + c(\boldsymbol{\zeta}_s^T \delta\boldsymbol{t} - \Delta t^s) + \boldsymbol{T}_{r_i}^s + \boldsymbol{I}_{r_i}^s - \boldsymbol{P}_{r_i}^s$$

To establish a temporal correlation of an observed value of GNSS in a factor graph, GNSS velocity measurement results are introduced to constrain position variables of adjacent epochs to obtain a Doppler velocity measurement model:

$$\boldsymbol{v}_i^{\mathrm{DV}} = h_i^{\mathrm{DV}}(x_i, x_{i+1}) + \boldsymbol{w}_i^{\mathrm{DV}}$$

$$h_i^{\mathrm{DV}}(x_i, x_{i+1}) = \begin{bmatrix} \dfrac{p_{i+1,x} - p_{i,x}}{\Delta t} \\ \dfrac{p_{i+1,y} - p_{i,y}}{\Delta t} \\ \dfrac{p_{i+1,z} - p_{i,z}}{\Delta t} \end{bmatrix}$$

wherein, $\boldsymbol{v}_i^{\mathrm{DV}}$ is a Doppler velocity at the time i, $p_{i,x}$ is a position at the time i, $\boldsymbol{w}_i^{\mathrm{DV}}$ is noise of measured value of the velocity, and Δt is a time difference between adjacent epochs.

GNSS velocity measurement results are introduced to constrain position variables of adjacent epochs to obtain a Doppler velocity measurement model; and a Doppler velocity factor residual refers to an error between a predicted value and an observed value of the velocity, and a residual of Doppler velocity measurement at the time i may be written as:

$$\boldsymbol{r}_D\left(\hat{z}_{i,\mathrm{DV}}, \boldsymbol{\chi}\right) = \boldsymbol{v}_i^{\mathrm{DV}} - h_i^{\mathrm{DV}}(x_i, x_{i+1})$$

Second step: global factor fusion and initialization. During initialization, it is first necessary

to determine whether a current frame is a keyframe. A main basis for keyframe determination in the GNSS/INS/Vision algorithm proposed herein has been described during the construction of the VIO factor. For GNSS observations, since the convergence of parameters during the single-point positioning of the pseudorange requires a certain amount of time, the continuity of observation of the satellite must be ensured. Therefore, when the GNSS satellite observations are input, the system may construct the observations as keyframes and add the keyframes to a sliding window.

Since a coordinate system of the visual-inertial odometry is a local coordinate system, the coordinate system needs to be initialized and aligned with a GNSS system before information fusion. The initialization and alignment between the GNSS system and the visual-inertial odometry may be regarded as a position trajectory alignment problem. A reference coordinate system of the GNSS system is an ENU system, while a coordinate system of VIO and expanded visual point information is a local world coordinate system. Before data fusion, a transformation matrix between the two coordinate systems needs to be calculated. If the ENU coordinate system is transformed into an ECEF coordinate system, which is the world coordinate system used in the present invention, the coordinate transformation matrix is $\boldsymbol{R}_n^e$,

$$\boldsymbol{R}_n^e = \begin{bmatrix} -\sin\beta & -\sin\alpha\cos\beta & \cos\alpha\cos\beta \\ \cos\beta & -\sin\alpha\sin\beta & \cos\alpha\sin\beta \\ 0 & \cos\alpha & \sin\alpha \end{bmatrix}$$

wherein, n is the ENU coordinate system; e is the ECEF coordinate system, and α and β are a latitude and a longitude of an origin of the ENU coordinate system respectively.

Third step: sliding window optimization. After the initialization is successful, subsequent observation may be added to the factor graph for overall optimization based on the factor graph. In navigation and positioning problems, a size of the factor graph grows as an amount of observation information increases. Therefore, to improve a real-time performance of a graph optimization algorithm, the present invention generally introduces the sliding window to limit the graph size.

When a new frame of observation arrives, a corresponding system state-variable node is constructed therefor first and added to the factor graph, and a corresponding error factor is constructed according to a measurement type thereof and connected with a relevant state-variable node. When the window is full at the time, a node corresponding to an oldest frame in the window requires to be marginalized, the node is removed and a prior constraint factor is constructed for a node related to the node in the window to participate in subsequent global state optimization of the

system.

An optimization problem constructed by the factors in the sliding window is:

$$\chi^* = \underset{\chi}{argmin}\{ \left\|\boldsymbol{r_p} - \boldsymbol{H_p}\chi\right\|^2 + \sum_{s\in \boldsymbol{P_{se}}} \left\|\boldsymbol{r_{Pse}}\left(\hat{\boldsymbol{z}}_{r,i}^{s}, \boldsymbol{\chi}\right)\right\|_{\boldsymbol{\Sigma}_{r,i}^{s}}^{2} + \sum_{i\in D}\left\|\boldsymbol{r_D}\left(\hat{\boldsymbol{z}}_{i,\mathrm{DV}}, \boldsymbol{\chi}\right)\right\|_{\boldsymbol{\Sigma}_{i,\mathrm{DV}}}^{2}$$

$$+ \sum_{i\in V}\left\|\boldsymbol{r}_{\mathrm{VIO}}\left(\hat{\boldsymbol{z}}_{i,\mathrm{VIO}}, \boldsymbol{\chi}\right)\right\|_{\Sigma_{i,\mathrm{VIO}}}^{2} + \sum_{(i,j)\in M}\left\|\boldsymbol{r}_i\left(\hat{\boldsymbol{z}}_i^{\boldsymbol{c_j}}, \boldsymbol{\chi}\right)\right\|_{\boldsymbol{\Sigma}_i^{c_j}}^{2} \}$$

$$\chi = \left[x_i', x_{i+1}', \ldots, x_{i+f}', \lambda_i, \lambda_{i+1}, \ldots, \lambda_{i+p}, o_i, o_{i+1}, \ldots, o_{i+l}, t_{r,s}, T\right]$$

$$\boldsymbol{x}_n' = \left[p_{b_n}^e, q_{b_n}^e, v_{b_n}^e\right], n \in [i, i+f]$$

wherein, χ^* represents a corresponding state at the optimal solution obtained, χ represents the global state vector in the sliding window, which is relative to a state quantity in the VIO system, $\boldsymbol{x}_n'$ represents a system state at the time t_i in the sliding window, and $p_{b_n}^e$, $q_{b_n}^e$ and $v_{b_n}^e$ represent a position, an attitude and a velocity at the time t_i in **a** coordinate system e respectively; $t_{r,s}$ represents receiver clock offsets of different satellite systems at the time t_i, r and s are a receiver and a satellite system respectively, and T represents a tropospheric delay at the time t_i.

$\boldsymbol{r_p}$ and $\boldsymbol{H_p}$ are a residual and a coefficient matrix of a prior constraint generated by marginalization; $\boldsymbol{r_{Pse}}$ is a pseudorange residual, $\boldsymbol{r_D}$ is a Doppler velocity residual, $\boldsymbol{r}_{\mathrm{VIO}}$ is a visual-inertial odometry residual, and $\boldsymbol{r}_i$ is a visual point feature reprojection residual of map expansion; $\boldsymbol{Pse}$ is a set of measured values of pseudoranges in the sliding window, $\boldsymbol{D}$ is a set of measured values of Doppler velocities in the sliding window, V is a set of measured values of VIOs in the sliding window, and $\boldsymbol{M}$ is a set of selected point features of map points in the sliding window; and $\boldsymbol{\Sigma}_{r,i}^{s}$, $\boldsymbol{\Sigma}_{i,\mathrm{DV}}$, $\boldsymbol{\Sigma}_{\mathrm{i,VIO}}$ and $\boldsymbol{\Sigma}_i^{c_j}$ are corresponding covariance matrices thereof.

Fourth step: global optimal pose. After GNSS/IMU/Vision global factor expansion and semi-tightly integrated navigation solution, a high-precision pose of a rescue vehicle may be obtained. According to the calculated high-precision pose of the vehicle, the global map in the disaster environment is further corrected (by the same steps as the high-precision map construction), thereby achieving real-time map update, and is used for calculating the next epoch.

In a specific implementation, the present application provides a computer storage medium and a corresponding data processing unit, wherein the computer storage medium is capable of storing a computer program, and the computer program, when executed by the data processing unit, can run the inventive contents of the Beidou multi-source fusion positioning method in the

disaster environment provided by the present invention and some or all steps in various embodiments. The storage medium may be a magnetic disk, an optical disk, a read-only memory (ROM) or a random access memory (RAM), and the like.

Those skilled in the art can clearly understand that the technical solutions in the embodiments of the present invention can be realized by means of a computer program and a corresponding general hardware platform thereof. Based on such understanding, the essence of the technical solutions in the embodiments of the present invention or the part contributing to the prior art, may be embodied in the form of a computer program, i.e., a software product. The computer program, i.e., the software product is stored in a storage medium including a number of instructions such that a device (which may be a personal computer, a server, a singlechip, an MCU or a network device, and the like) including the data processing unit executes the methods described in various embodiments or some parts of the embodiments of the present invention.

The present invention provides an idea and a method for a Beidou multi-source fusion positioning method in a disaster environment, with many methods and ways to realize the technical solution specifically. Those described above are merely the preferred embodiments of the present invention, and it should be pointed out that those of ordinary skills in the art may further make improvements and decorations without departing from the principle of the present invention, and these improvements and decorations should also be regarded as the scope of protection of the present invention. All the unspecified components in the embodiments can be realized by the prior art.

CLAIMS

1. A Beidou multi-source fusion positioning method in a disaster environment, comprising the following steps:

step 1: acquiring images using a camera on a carrier to be positioned in the disaster environment, and constructing a VIO factor based on point and line feature extraction;

step 2: constructing a map and performing refined correction on the map under a GNSS/IMU constraint to obtain a global map; and

step 3: augmenting visual factors using the global map, and performing global pose optimization for a GNSS/INS/vision integrated navigation system to complete Beidou multi-source fusion positioning in the disaster environment.

2. The Beidou multi-source fusion positioning method in the disaster environment according to claim 1, wherein the constructing the VIO factor based on the point and line feature extraction in the step 1, specifically comprises:

step 1-1: performing the point and line feature extraction and keyframe determination;

step 1-2: constructing a visual feature reprojection residual factor;

step 1-3: constructing an IMU pre-integration residual factor; and

step 1-4: constructing the VIO factor.

3. The Beidou multi-source fusion positioning method in the disaster environment according to claim 2, wherein the performing the point and line feature extraction and the keyframe determination in the step 1-1, specifically comprises:

step 1-1-1: point and line feature extraction: according to the acquired images, obtaining point and line features using FAST point feature extraction and LSD line feature extraction; and matching the point and line features by a BRIEF descriptor, and eliminating mismatched results of the point and line features by a RANSAC method;

step 1-1-2: keyframe determination: setting the acquired images as a set of consecutive frames $F = \{F_1, F_2, F_3, \ldots, F_i, \}$, wherein F_i represents an $i\text{-}th$ frame, and each frame of image F_i

comprises the following parameters: a corresponding timestamp T_i and a number b_i of point and line features contained in the frame of image, allowing K_n to represent an $n\text{-}th$ keyframe, and determining whether the $i\text{-}th$ frame of image F_i is a keyframe according to the following criteria:

a first criterion, in which a time interval between the $i\text{-}th$ frame of image F_i and an adjacent keyframe is greater than a specified threshold, which means that $T_i - T_{K_n} > T_0$, T_{K_n} represents a timestamp of a latest keyframe, and T_0 is a set time-difference threshold;

a second criterion, in which the number of point and line features contained in the keyframe is greater than a specified threshold, which means that $b_i > b_0$, and b_0 is a set number threshold of point and line features; and

directly discarding frames determined as non-keyframes; and

step 1-1-3: feature matching and filtering: after identifying the keyframe, performing feature matching using a KNN algorithm, which means that: for one feature in one frame of image, a most similar feature is searched for in another frame of image, and then the features are regarded as a matching pair; and during the above matching process, eliminating mismatches using a RANSAC algorithm.

4. The Beidou multi-source fusion positioning method in the disaster environment according to claim 3, wherein the constructing the visual feature reprojection residual factor in the step 1-2, refers to constructing a visual feature reprojection residual model, and taking a distance between a projection point of a map point in an image and a corresponding observation point as a visual feature reprojection residual, and specifically comprises:

step 1-2-1: point feature reprojection residual, which is specifically as follows:

for an $l\text{-}th$ feature point L, normalized coordinates thereof observed in an $i\text{-}th$ camera coordinate system corresponding to the $i\text{-}th$ frame F_i are set as $\hat{L}_l^{c_i}$, normalized coordinates of the same feature point observed in a $j\text{-}th$ camera coordinate system corresponding to a $j\text{-}th$ frame F_j are obtained as $\hat{L}_l^{c_j}$ by an image feature tracking method, normalized coordinates of the feature point L transformed from the $i\text{-}th$ camera coordinate system in which the feature point is first observed to the camera coordinate system j in which the feature point is observed again are $L_l^{c_j}$, referring to a predicted observed value, and then a visual observation residual term is a

difference $r_l\left(\hat{z}_l^{c_j}, \mathcal{X}_{VIO}\right)$ between the predicted value and an actual observed value, which is as follows:

$$r_l\left(\hat{z}_l^{c_j}, \mathcal{X}_{VIO}\right) = \hat{L}_l^{c_j} - L_l^{c_j}$$

wherein, $\mathcal{X}_{VIO}$ represents a global state vector of a VIO local pose in a world coordinate system, and $\hat{z}_l^{c_j}$ represents a point feature observation of the camera;

$$\mathcal{X}_{VIO} = \left[x_i, x_{i+1}, \ldots, x_{i+f}, \lambda_i, \lambda_{i+1}, \ldots, \lambda_{i+p}, o_i, o_{i+1}, \ldots, o_{i+l}\right]$$

$$x_n = \left[p_n^{\omega}, q_n^{\omega}, v_n^{\omega}, b_{an}, b_{gn}\right], n \in [i, i+f]$$

wherein, x_n comprises a position p_n^{ω}, an orientation q_n^{ω}, a velocity v_n^{ω}, an accelerometer bias b_{an} and a gyroscope bias b_{gn} of a carrier provided by an n-*th* IMU; f, p and l represent total numbers of image frames, point features and line features in a sliding window respectively; λ_{i+m} represents an inverse depth between an m-*th* feature point and an initial feature point; and o_{i+w} represents an orthogonal representation of a w-*th* feature line in the world coordinate system; and

step 1-2-2: line feature reprojection residual, referring to taking a distance from a line endpoint observed in a camera plane to a reprojected line as the residual, which is specifically as follows:

an i-*th* camera, which refers to a camera capturing the i-*th* frame F_i, is allowed to observe a line L_k, and the line L_k is projected onto a normalized plane of the i-*th* keyframe to obtain a projected line equation $l_k^{c_i}$; and two line endpoints observed in the camera plane are $\hat{s}_k^{c_i} = [u_s \quad v_s \quad 1]^T$ and $\hat{e}_k^{c_i} = [u_e \quad v_e \quad 1]^T$ respectively, $d(x, L)$ is used to represent a distance from a point x to a line L, and then the line feature reprojection residual $r_{L_i}\left(\hat{s}_i^{c_j}, \hat{e}_i^{c_j}, \mathcal{X}_{VIO}\right)$ is:

$$r_{L_i}\left(\hat{s}_i^{c_j}, \hat{e}_i^{c_j}, \mathcal{X}_{VIO}\right) = \begin{bmatrix} d\left(\hat{s}_k^{c_i}, l_k^{c_i}\right) \\ d\left(\hat{e}_k^{c_i}, l_k^{c_i}\right) \end{bmatrix}$$

.

5. The Beidou multi-source fusion positioning method in the disaster environment according to claim 4, wherein the constructing the IMU pre-integration residual factor in the step 1-3, specifically comprises:

step 1-3-1: constructing an IMU pre-integration residual model, which is specifically as follows:

IMU integration between two adjacent keyframes is performed in an IMU carrier coordinate system of a previous keyframe F_i to obtain a relative motion increment independent of a state of the keyframe F_i, wherein a change in an IMU bias is handled by first-order linear approximation; and in the IMU coordinate system, IMU measurement equations are:

$$\hat{a}_t = a_t + b_{a_t} + R_w^{b_t} g^w + n_a$$

$$\hat{w}_t = w_t + b_{w_t} + n_w$$

wherein, $\hat{a}_t$ is a measured value of acceleration at a time t, a_t is a noise-free acceleration at the time t, b_{a_t} is an accelerometer bias at the time t, $R_w^{b_t}$ is a coordinate transformation matrix, g^w is a gravitational acceleration, and n_a is accelerometer white noise; and $\hat{w}_t$ is a measured value of gyroscope at the time t, w_t is a noise-free angular velocity at the time t, b_{w_t} is a gyroscope bias at the time t, and n_w is gyroscope white noise;

integration is performed according to the above IMU measurement equations and coordinate system transformation is performed, and then IMU pre-integration is as follows:

$$\begin{cases} \alpha_{b_j}^{b_i} = \iint_{t\in[i,j]} \left(R_{b_t}^{b_i} \left(\hat{a}_t - b_{a_t} - n_a \right) \right) \mathrm{dt}^2 \\ \beta_{b_j}^{b_i} = \int_{t\in[i,j]} \left(R_t \left(\hat{a}_t - b_{a_t} - n_a \right) \right) \mathrm{dt} \\ \gamma_j^i = \int_{t\in[i,j]} \left(\frac{1}{2} \Omega \left(\hat{w}_t - b_{w_t} - n_w \right) q_t^{b_i} \right) \mathrm{dt} \end{cases}$$

wherein, $\alpha_{b_j}^{b_i}$, $\beta_{b_j}^{b_i}$ and γ_j^i are relative position, velocity and rotation integrated from t_i to t_j, and an intermediate term $\Omega(w)$ is calculated as follows:

$$\Omega(w) = \begin{bmatrix} -[w]_\times & w \\ -w^T & 0 \end{bmatrix}$$

wherein, w represents the measured value of gyroscope, × represents an operator $(\cdot)\times$ mapping a vector to a skew-symmetric matrix thereof, and T represents a matrix transpose operation; and

step 1-3-2: constructing the IMU pre-integration residual factor, which is specifically as follows:

after IMU pre-integration, a residual $r_B\left(\hat{z}_{b_j}^{b_i}, \mathcal{X}_{VIO}\right)$ between a system state and IMU measurement is as follows:

$$r_B\left(\hat{z}_{b_j}^{b_i}, \mathcal{X}_{VIO}\right) = \left[\delta\alpha_{b_j}^{b_i} \quad \delta\beta_{b_j}^{b_i} \quad \delta\gamma_j^i \quad \delta b_a \quad \delta b_g\right]^T$$

wherein, δ represents a difference between a measured value and a predicted value, and $\hat{z}_{b_j}^{b_i}$ represents an IMU observation.

6. The Beidou multi-source fusion positioning method in the disaster environment according to claim 5, wherein the constructing the VIO factor in the step 1-4, specifically comprises:

according to the point feature reprojection residual, the line feature reprojection residual and the IMU pre-integration residual, performing joint optimization by a least squares method, which is specifically as follows:

$$\mathcal{X}_{VIO}{}^* = \underset{\mathcal{X}_{VIO}}{argmin}\{\left\|r_p - H_p\mathcal{X}_{VIO}\right\|^2 + \sum_{(i,j)\in B}\left\|r_B\left(\hat{z}_{b_j}^{b_i}, \mathcal{X}_{VIO}\right)\right\|_{\Sigma_{b_i}^{b_i}}^2 + \sum_{(i,j)\in C_1}\left\|r_l\left(\hat{z}_l^{c_j}, \mathcal{X}_{VIO}\right)\right\|_{\Sigma_l^{c_j}}^2 + \sum_{(i,j)\in C_2}\left\|r_{L_i}\left(\hat{z}_{L_i}^{c_j}, \mathcal{X}_{VIO}\right)\right\|_{\Sigma_{L_i}^{c_j}}^2\}$$

wherein, $\mathcal{X}_{VIO}{}^*$ represents a corresponding state value at an optimal solution, r_p and H_p are prior information, r_B is the IMU pre-integration residual, r_l is the point feature reprojection residual, and r_{L_k} is the line feature reprojection residual; and B is a measured value set of IMU, C_1 and C_2 are measured value sets of point and line features, and $\Sigma_{b_j}^{b_i}$, $\Sigma_l^{c_j}$ and $\Sigma_{L_k}^{c_j}$ are covariance

matrices corresponding to a measured value $\hat{z}_{b_j}^{b_i}$ of IMU, a measured value $\hat{z}_l^{c_j}$ of point feature and a measured value $\hat{z}_{L_i}^{c_j}$ of line feature respectively; and

according to the above optimization result, constructing a model using a VIO position increment and an attitude increment from a time t_{i-1} to a time t_i to obtain a measurement model as follows:

$$y_{\text{i,VO}} = \begin{Bmatrix} y_{\delta R,i} \\ y_{\Delta p,i} \end{Bmatrix} = \begin{Bmatrix} exp\left(n_{\delta R,i}\right) R_{i-1}^T R_i \\ R_{i-1}^T (P_i - P_{i-1}) + n_{\Delta p,i} \end{Bmatrix}$$

wherein, $y_{\delta R,i}$ and $y_{\Delta p,i}$ are measured values of attitude increment and position increment of visual odometry, $n_{\delta R,i}$ is measurement noise of attitude increment, R_{i-1} and R_i are measured values of attitudes at adjacent times, P_{i-1} and P_i are measured values of positions at adjacent times, and $n_{\Delta p,i}$ is measurement noise of position increment; and then the VIO factor $r_{VIO}\left(\hat{z}_{i,\text{VIO}}, \chi\right)$ is:

$$r_{VIO}\left(\hat{z}_{i,\text{VIO}}, \chi\right) = \begin{Bmatrix} r_{\delta R,i} \\ r_{\Delta p,i} \end{Bmatrix} = \begin{Bmatrix} log\left(y_{\delta R,i} - R_{i-1}^T R_i\right) \\ y_{\Delta p,i} - R_{i-1}^T (P_i - P_{i-1}) \end{Bmatrix}$$

wherein, $\hat{z}_{i,\text{VIO}}$ represents a measured value of VIO system, and $r_{\delta R,i}$ and $r_{\Delta p,i}$ represent an attitude-increment residual and a position-increment residual respectively.

7. The Beidou multi-source fusion positioning method in the disaster environment according to claim 6, wherein the constructing the map and performing the refined correction on the map under the GNSS/IMU constraint in the step 2, specifically comprises:

step 2-1: preliminary map construction, which is specifically as follows:

after keyframe determination, a keyframe selected from a camera image sequence is transformed into a three-dimensional space by an LSS method, and a pixel point (u, v) in the image is mapped to a corresponding global coordinate system point (X_w, Y_w, Z_w) by a method as follows:

$$[u \quad v \quad 1]^T = KRT[X_w \quad Y_w \quad Z_w \quad 1]^T$$

wherein, R is a camera rotation matrix, T is a camera translation vector, and K is a camera intrinsic parameter matrix;

$$R = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & cos\,\theta & -sin\,\theta & 0 \\ 0 & sin\,\theta & cos\,\theta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$T = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -\dfrac{h}{sin\,\theta} \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

wherein, h is a camera height, and θ is a camera pitch angle;

step 2-2: open region identification, wherein, by a convolutional neural network semantic segmentation method, a mapped element, which refers to the map point in the three-dimensional space, and also refers to the global coordinate system point in the step 2-1), is assigned a label, and a density of objects with the same label is determined, which is specifically as follows:

the open region is defined, which means that a density of tall buildings in the region with an aspect ratio greater than 1: 1 and buildings meeting a height requirement in the same region is less than a preset threshold;

when a road is identified in the open region, a point on the road is selected as a center of the open region and assigned a new label: an open region reference point; and when the open region belongs to a non-traversable region, a building closest to a traversable road in the open region is assigned the above new label: the open region reference point; and

after using the SSD convolutional neural network semantic segmentation method, a detected object in the image is assigned a classification label and two-dimensional position information; and according to the LSS method in the step 2-1, the above classification label is mapped to a three-dimensional position;

step 2-3: refinement under GNSS/IMU constraint, which means that, according to prior training results of the convolutional neural network semantic segmentation algorithm, during real-time positioning, the open region identification is performed simultaneously with the map construction, and according to the label of the open area reference point, when passing through

the region or approaching the region which refers to being impassable, constraint correction is performed using a loosely coupled GNSS/IMU navigation solution;

step 2-4: global high-precision map optimization construction, which is specifically as follows:

whether continuously generated local maps overlap is determined through loop closure detection, and when the continuously generated local maps overlap, the local maps are updated according to nonlinear optimization, which is implemented by solving the following nonlinear least squares problem via a Gauss-Newton method:

$$min_{x_1,x_2,\ldots,x_n} \sum_{i,j} \rho\left(e_{ij}(x_i, x_j, z_{ij})^T \Omega_{ij} e_{ij}(x_i, x_j, z_{ij})\right)$$

wherein, $e_{ij}(x_i, x_j, z_{ij})$ is an error between observation data z_{ij} from a pose x_i to a pose x_j and pose estimation, Ω_{ij} is an information matrix of the observation data, and $\rho(\cdot)$ is a robust kernel function;

when traversed regions do not overlap, a co-visibility relationship is obtained according to perceptual similarity during BEV perception, the co-visibility relationship is represented by a matrix V, and V_{ij} represents a co-visibility degree between the pose x_i and the pose x_j; and feature points P_i and P_j under similar frames F_i and F_j are searched for according to the co-visibility relationship, and corresponding feature points representing the same landmark are searched for by calculating a Euclidean distance $d(P_i, P_j)$ between the feature points, as follows:

$$d(P_i, P_j) = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2}$$

when the Euclidean distance is less than a specified threshold, it is determined that the feature points P_i and P_j are identical, and a global map is generated according to the corresponding feature points representing the same landmark found based on the co-visibility relationship; and

step 2-5: global factor expansion based on feature matching, which means that visual feature state vectors are expanded using the generated or updated map points, and new visual reprojection residual factors are generated as auxiliary positioning information for a global factor

graph optimization process; and when performing visual factor expansion using the map, only the feature point reprojection residuals are expanded.

8. The Beidou multi-source fusion positioning method in the disaster environment according to claim 7, wherein the refinement under GNSS/IMU constraint in the step 2-3, is specifically as follows:

when passing through the open region reference point, GNSS and IMU integrated navigation yields a high-precision pose to correct a previously predicted pose of the point, and further provides a new constraint for map construction after the time: position and attitude information provided by GNSS/IMU is transformed from the world coordinate system to the camera coordinate system, and further transformed to a pixel coordinate system through a camera intrinsic matrix, and then jointly optimized with a previously predicted visual target point at the same position, which refers to minimizing a value η of the following formula:

$$\eta = argmin\frac{1}{2}\sum_{i=1}^{n}\left\|u_i - \frac{1}{z}K\left(R_w^c(P - P_c)\right)\right\|^2$$

wherein, P is target point coordinates for joint optimization between the GNSS/IMU observation point and the map point, K is a camera intrinsic parameter, z is a feature point depth, R_w^c and P_c are prior transformation matrix information and map feature point position information during visual image processing and map point generation respectively, and u_i is an observed value composed of coordinates positioned by GNSS/IMU and subjected to coordinate system transformation to the pixel coordinate system and a disparity of left and right images.

9. The Beidou multi-source fusion positioning method in the disaster environment according to claim 8, wherein the augmenting the visual factors using the global map in the step 3, specifically comprises:

step 3-1: construction of pseudorange factor and Doppler velocity measurement factor, which is specifically as follows:

a pseudorange measurement model is obtained during single-point positioning solution, a pseudorange factor residual refers to an error between a predicted value and an observed value of

a pseudorange, and according to the pseudorange measurement model, a pseudorange residual of a satellite s at the time t_i is calculated as r_{Pse};

GNSS velocity measurement results are introduced to constrain position variables of adjacent epochs to obtain a Doppler velocity measurement model; and a Doppler velocity factor residual refers to an error between a predicted value and an observed value of a velocity, and a residual of Doppler velocity measurement at the time t_i is calculated as r_D;

step 3-2: global factor fusion and initialization, which is specifically as follows:

whether a current frame is the keyframe is determined, which refers to determining the keyframe according to the method in the step 1; and when GNSS satellite observation data are input, the data are constructed as the keyframe and added to the sliding window; and

a coordinate system of the visual-inertial odometry is initialized and aligned with a GNSS system;

step 3-3: sliding window optimization, wherein, after the initialization in the step 3-2 is successful, subsequent observation is added to a factor graph for overall optimization based on the factor graph; and the sliding window is introduced to limit a graph size, which is specifically as follows:

when a new frame of observation arrives, a corresponding system state-variable node is constructed therefor first and added to the factor graph, and a corresponding error factor is constructed according to a measurement type thereof and connected with a relevant state-variable node; and when the window is full at the time, a node corresponding to an oldest frame in the window is marginalized, the node is removed and a prior constraint factor is constructed for a node related to the node in the window to participate in subsequent global state optimization of the system; and

step 3-4: global optimal pose, wherein, after the above steps, a high-precision pose of a carrier to be positioned is obtained, and according to the high-precision pose, the global map is further corrected to achieve real-time update of the map, and used for calculating the next epoch.

10. The Beidou multi-source fusion positioning method in the disaster environment according to claim 9, wherein the sliding window optimization in the step 3-3, means that an optimization problem constructed by the factors in the sliding window is:

$$\chi^* = \underset{\chi}{argmin}\{ \left\|\boldsymbol{r_p} - \boldsymbol{H_p}\chi\right\|^2 + \sum_{s\in \boldsymbol{P_{se}}} \left\|\boldsymbol{r_{Pse}}\left(\hat{\boldsymbol{z}}_{r,i}^{s}, \boldsymbol{\chi}\right)\right\|_{\boldsymbol{\Sigma}_{r,i}^{s}}^2 + \sum_{i\in D} \left\|\boldsymbol{r_D}\left(\hat{\boldsymbol{z}}_{i,\mathrm{DV}}, \boldsymbol{\chi}\right)\right\|_{\boldsymbol{\Sigma}_{i,\mathrm{DV}}}^2$$
$$+ \sum_{i\in V} \left\|\boldsymbol{r}_{\mathrm{VIO}}\left(\hat{\boldsymbol{z}}_{i,\mathrm{VIO}}, \boldsymbol{\chi}\right)\right\|_{\Sigma_{i,\mathrm{VIO}}}^2 + \sum_{(i,j)\in M} \left\|\boldsymbol{r}_i\left(\hat{\boldsymbol{z}}_i^{\boldsymbol{cj}}, \boldsymbol{\chi}\right)\right\|_{\boldsymbol{\Sigma}_i^{cj}}^2 \}$$

$$\chi = \left[x_i', x_{i+1}', \ldots, x_{i+f}', \lambda_i, \lambda_{i+1}, \ldots, \lambda_{i+p}, o_i, o_{i+1}, \ldots, o_{i+l}, t_{r,s}, T\right]$$

$$x_n' = \left[p_{b_n}^e, q_{b_n}^e, v_{b_n}^e\right], n \in [i, i+f]$$

wherein, χ^* represents a corresponding state at the optimal solution obtained, χ represents the global state vector in the sliding window, which is relative to a state quantity in the VIO system, x_n' represents a system state at the time t_i in the sliding window, and $p_{b_n}^e$, $q_{b_n}^e$ and $v_{b_n}^e$ represent a position, an attitude and a velocity at the time t_i in a coordinate system e respectively; $t_{r,s}$ represents receiver clock offsets of different satellite systems at the time t_i, r and s are a receiver and a satellite system respectively, T represents a tropospheric delay at the time t_i, and r_i is a visual point feature reprojection residual of map expansion; Pse is a set of measured values of pseudoranges in the sliding window, D is a set of measured values of Doppler velocities in the sliding window, V is a set of measured values of VIOs in the sliding window, and M is a set of selected point features of map points in the sliding window; and $\Sigma_{r,i}^{s}$, $\Sigma_{i,\mathrm{DV}}$, $\Sigma_{\mathrm{i,VIO}}$ and $\Sigma_i^{c_j}$ represent covariance matrices corresponding to a measure value $\hat{\boldsymbol{z}}_{r,i}^{s}$ of pseudorange, a measured value $\hat{\boldsymbol{z}}_{i,\mathrm{DV}}$ of Doppler velocity, a measured value $\hat{\boldsymbol{z}}_{i,\mathrm{VIO}}$ of visual-inertial odometry and a measured value $\hat{\boldsymbol{z}}_i^{\boldsymbol{cj}}$ of map point feature respectively.