# Research on real time 6DOF robot localization based on visual and inertial fusion

Xinfang.Zheng,Yongxiang.Guo
*School of electronic and information engineering, Beijing Jiaotong University, Beijing 100044,China*
*School of electronic and information engineering, Beijing Jiaotong University, Beijing 100044,China)*

**Abstract**

In this paper, first of all, how to realize the real-time positioning of the robot by the inertial sensor was discussed. Secondly, how to realize the real-time positioning of robot by visual sensor was discussed and researched deeply. Finally, combined with low-cost inertial measurement unit (IMU) in the image sampling interval time of the short-term stationarity, the visual odometry method integrating the inertial measurement information was studied in depth. In the study of theory at the same time, experiments on visual robot localization by LIBVISO2 were carried out and the experiments of visual and inertial fusion were carried out.The experimental results are displayed by the visualization tool RVIZ in the Robot Operating System (ROS).

***Keyword*：*Robot Location; IMU;Visual Odometry;***

## 1. Introduction

The future development of robots tends to be more mobile, and mobile robots will require navigation and positioning in both dynamic and unknown environments.It is very difficult to realize precise and real-time positioning of mobile robot by a simple navigation or positioning method, so that fusion is very important.At present, integrated navigation system[1] is divided into two kinds of loose coupling and tight coupling.In the loosely coupled system, the position and attitude information are outputted by the respective localization methods, and then the information being optimized by some algorithm, finally the optimal result is given.Loose coupling method is simple and easy to implement, andits cost is low.In a tightly coupled system, each positioning method has a close relationship, and one location method will appear in another positioning process, and finally the position and attitude information is output, rather than each method can get a positioning information.

In recent years, with the rapid development of hardware devices, making the current robot visual navigation[2] is quite active.Visual sensor can get a lot of information, and visual navigation is closer to the human navigation mechanism.Binocular vision is like the person's eyes, through which you can get outside of the image information.Stereoscopic parallax is formed by imaging the same environmental characteristics in the left and right cameras.Binocular stereoscopic vision system uses this parallax to calculate the depth information so as to realize three-dimensional reconstruction.For visual odometry, the process of feature detection, matching and RANSAC out-of-site filtering[3] requires more time, affecting real-time performance, limiting the sampling frequency of the vision system, resulting in the inability of the robot to accelerate and to carry out stable pose tracking.Inertial navigation is similar to the human brain, having a better ability to perceive the rotation of information, and simultaneously having the advantages of short-term stability.Since inertial navigation is done by way of integrationandthere is the phenomenon of error drift,it does not work for long periods of time.

Just like the combination of human eyes and brain,using the low-cost IMU (Inertial Measurement Unit) which has short-term stationarity in the image sampling interval, we aim to address the robot location problems by blending the visual odometry with inertial navigation, so that the 6DOF (Six Degrees of Freedom) mobile robot localization[4] is realized with high accuracy and real-time performance.

## 2. Inertial navigation theory

In order to realize the localization of mobile robot, multiple sets of data of three-axis accelerometer and three-axis gyroscope are collected firstly, then the mean value of the multiple groups of data is used to eliminate the matching error. The quaternion is updated by the second-order iterative algorithm of the Biker approximation method, simultaneously the rotation matrix and the optimal attitude angle are calculated for the current state. Finally, the acceleration is integrated twice in the navigation coordinate system to obtain the position information.In order to obtain more precise location information, it is possible to determine whether the average value of the triaxial output of the gyroscope is zero in the case that the attitude angle of the carrier does not change,so as to realize the simple optimization of the algorithm.Ideally, if the attitude angle does not change, the output average should be zero, but the output value is slightly higher than zero due to vibration and other reasons.The process is as follows:

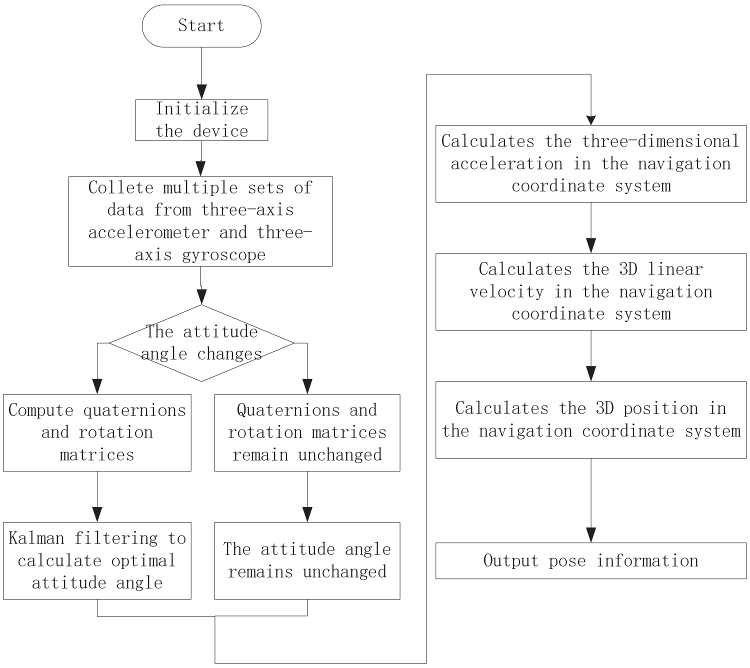

Start
Initialize the device
Collete multiple sets of data from three-axis accelerometer and three-axis gyroscope
The attitude angle changes
Compute quaternions and rotation matrices
Quaternions and rotation matrices remain unchanged
Kalman filtering to calculate optimal attitude angle
The attitude angle remains unchanged
Calculates the three-dimensional acceleration in the navigation coordinate system
Calculates the 3D linear velocity in the navigation coordinate system
Calculates the 3D position in the navigation coordinate system
Output pose information

Figure 2.1 Flow chart of inertial navigation

(1) the update of quaternion Q

Suppose that the vector coordinate system is represented by a rotating quaternion[5] with respect to the navigation coordinate system (geographic coordinate system),the updating of the quaternion can be realized by the rotation quaternion differential method.

$$\dot{Q}=\begin{bmatrix}\dot{q}_0\\ \dot{q}_1\\ \dot{q}_2\\ \dot{q}_3\end{bmatrix}=\frac{1}{2}\omega_{nb}^{b}Q=\frac{1}{2}\begin{bmatrix}0 & -\omega_{nbx}^{b} & -\omega_{nby}^{b} & -\omega_{nbz}^{b}\\ \omega_{nbx}^{b} & 0 & \omega_{nbz}^{b} & -\omega_{nby}^{b}\\ \omega_{nby}^{b} & -\omega_{nbz}^{b} & 0 & \omega_{nbx}^{b}\\ \omega_{nbz}^{b} & \omega_{nby}^{b} & -\omega_{nbx}^{b} & 0\end{bmatrix}\begin{bmatrix}q_0\\ q_1\\ q_2\\ q_3\end{bmatrix} \quad (1)$$

(2) The calculation of the attitude matrix R

Attitude matrix R can be obtainedaccording to the quaternion of the four elements.

$$R=\begin{bmatrix}q_0^2+q_1^2-q_2^2-q_3^2 & 2\left(q_1q_2+q_0q_3\right) & 2\left(q_1q_3-q_0q_2\right)\\ 2\left(q_1q_2-q_0q_3\right) & q_0^2-q_1^2+q_2^2-q_3^2 & 2\left(q_2q_3+q_0q_1\right)\\ 2\left(q_1q_3+q_0q_2\right) & 2\left(q_2q_3-q_0q_1\right) & q_0^2-q_1^2-q_2^2+q_3^2\end{bmatrix} \quad (2)$$

(3) The best normalization of quaternion Q.

The best normalization of quaternion with Euclidean norm as the index is shown below.

$$Q = \frac{\hat{Q}}{\sqrt{\hat{q}_0^2 + \hat{q}_1^2 + \hat{q}_2^2 + \hat{q}_3^2}} \quad (3)$$

(4) Coordinate transformation of acceleration

Acceleration $a^b$ measured by the accelerometer can be converted to $a^n$ by the conversion matrix R.

$$a^n = \begin{bmatrix} a_x^n \\ a_y^n \\ a_z^n \end{bmatrix} = R \begin{bmatrix} a_x^b \\ a_y^b \\ a_z^b \end{bmatrix} = \begin{bmatrix} R_{11} & R_{12} & R_{13} \\ R_{21} & R_{22} & R_{23} \\ R_{31} & R_{32} & R_{33} \end{bmatrix} \begin{bmatrix} a_x^b \\ a_y^b \\ a_z^b \end{bmatrix} \quad (4)$$

(5) Correction speed

$$\begin{bmatrix} \dot{v}_x^n \\ \dot{v}_y^n \\ \dot{v}_z^n \end{bmatrix} = \begin{bmatrix} a_x^n \\ a_y^n \\ a_z^n \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \\ g \end{bmatrix} + \begin{bmatrix} 0 & 2\omega_{iez}^n + \omega_{enz}^n & -\left(2\omega_{iey}^n + \omega_{eny}^n\right) \\ -\left(2\omega_{iez}^n + \omega_{enz}^n\right) & 0 & 2\omega_{iex}^n + \omega_{enx}^n \\ 2\omega_{iey}^n + \omega_{eny}^n & -\left(2\omega_{iex}^n + \omega_{enx}^n\right) & 0 \end{bmatrix} \begin{bmatrix} v_x^n \\ v_y^n \\ v_z^n \end{bmatrix}$$

(5)

(6) Calculate the attitude rate

$$\bar{\omega}_{nb}^b = \bar{\omega}_{ib}^{b-1} - \bar{\omega}_{in}^b = \bar{\omega}_{ib}^b - R\left(\bar{\omega}_{en}^n + \bar{\omega}_{ie}^n\right) \quad (6)$$

(7) Calculate the attitude angle

The attitude matrix R can also be calculated from the relationship between the pitch angle $\theta$, the roll angle $\varphi$ and the heading angle $\psi$.

$$\begin{aligned} R &= R(\psi)R(\theta)R(\varphi) \\ &= \begin{bmatrix} \cos(\psi)\cos(\varphi) - \cos(\theta)\sin(\psi)\sin(\varphi) & -\cos(\theta)\cos(\varphi)\sin(\psi) - \cos(\psi)\sin(\varphi) & \sin(\psi)\sin(\theta) \\ \cos(\varphi)\sin(\psi) + \cos(\psi)\cos(\theta)\sin(\varphi) & \cos(\psi)\cos(\theta)\cos(\varphi) - \sin(\psi)\sin(\varphi) & -\cos(\psi)\sin(\theta) \\ \sin(\theta)\sin(\varphi) & \cos(\varphi)\sin(\theta) & \cos(\theta) \end{bmatrix} \\ &= \begin{bmatrix} R_{11} & R_{12} & R_{13} \\ R_{21} & R_{22} & R_{21} \\ R_{31} & R_{32} & R_{22} \end{bmatrix} \end{aligned}$$

(7)

The main values of the pitch angle $\theta$, the roll angle $\varphi$ and the heading angle $\psi$ can be obtained from the elements of the attitude matrix.

$$\theta_{\text{主}} = \sin^{-1}\left(R_{32}\right)$$

$$\varphi_{\text{主}} = \tan^{-1}\left(\frac{-R_{31}}{R_{33}}\right) \quad (8)$$

$$\psi_{\text{主}} = \tan^{-1}\left(\frac{-R_{12}}{R_{22}}\right)$$

The pitch angle $\theta$ is defined in the interval $\left(-90^o, 90^o\right)$, the roll angle $\varphi$ is defined in the interval $\left(-180^o, 180^o\right)$, and the heading angle $\psi$ is defined in the interval $\left(0^o, 360^o\right)$. Since the definition interval of the pitch angle $\theta$ coincides with the main value interval of the arcsine function, the main value is the true value; The definition range of the roll angle $\varphi$ and the heading angle $\psi$ is not consistent with the value range of the arc tangent function, so the true value of the roll angle $\varphi$ and heading angle $\psi$ also need to be determined according to the positive and negative of the $R_{33}$ and $R_{22}$. The true values of the three attitude angles are calculated as follows.

$$\theta = \theta_{主}$$

$$\varphi = \begin{cases} \varphi_{主} & \left(R_{33} > 0\right) \\ \varphi_{主} + 180^o & \left(R_{33} < 0, \varphi_{主} < 0\right) \\ \varphi_{主} - 180^o & \left(R_{33} < 0, \varphi_{主} > 0\right) \end{cases} \qquad (9)$$

$$\psi = \begin{cases} \psi_{主} & \left(R_{22} > 0, \psi_{主} > 0\right) \\ \psi_{主} + 360^o & \left(R_{22} > 0, \psi_{主} < 0\right) \\ \psi_{主} + 180^o & \left(R_{22} < 0\right) \end{cases}$$

When the pitch angle is $\theta = \pm 90^o$, the roll axis and the heading axis coincide. The three attitude angle degrees of freedom become two attitude angle degrees of freedom. At this time, the attitude is represented by the pitch angle $\theta$ and the sum of the roll angle and the heading angle $\varphi + \psi$ .Based on experience, when the pitch angle is $\theta = \pm 90^o$ and the roll angle $\varphi = 0^o$ , $\psi$ can be calculated according to equation (7).

## 3. Visual odometry

The visual odometer [6] estimates the camera's posture and position informationbyusing the characteristics of the images acquired by the camera during motion and according to the positional relationship of these features in three-dimensional space.Binocular camera used in the visual odometer, on the one hand they can track the image feature points in space, on the other hand they can restore the feature point of the scale of information,and they can effectively compute the camera's current location information and attitude information.In this paper, we take the LIBVISO2 as an example to introduce the working principle of visual odometry, the most important part of the motion estimation.

The camera motion parameters are calculated by minimizing the re-projection error.First, bucketing is used to reduce the number of features (in practice we retain between 200 and 500 features) and spread them uniformly over the image domain. Next, we project feature points from the previous frame into 3d via triangulation using the calibration parameters of the stereo camera rig. Assuming squared pixels and zero skew, the reprojection into the current image is given by(10).

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \begin{pmatrix} f & 0 & c_u \\ 0 & f & c_v \\ 0 & 0 & 1 \end{pmatrix} \left[ \begin{pmatrix} R(t) & T \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} - \begin{pmatrix} s \\ 0 \\ 0 \end{pmatrix} \right]$$

(10)

with

- homogeneous image coordinates (u v 1)T
- focal length f
- principal point (cu,cv),(cu,cv)
- rotation matrix R(r)=Rx(rx)Ry(ry)Rz(rz)
- translation vector t=(tx ty tz)T
- 3d point coordinates X=(x y z)T
- and shift s=0 (left image), s = baseline (right image).

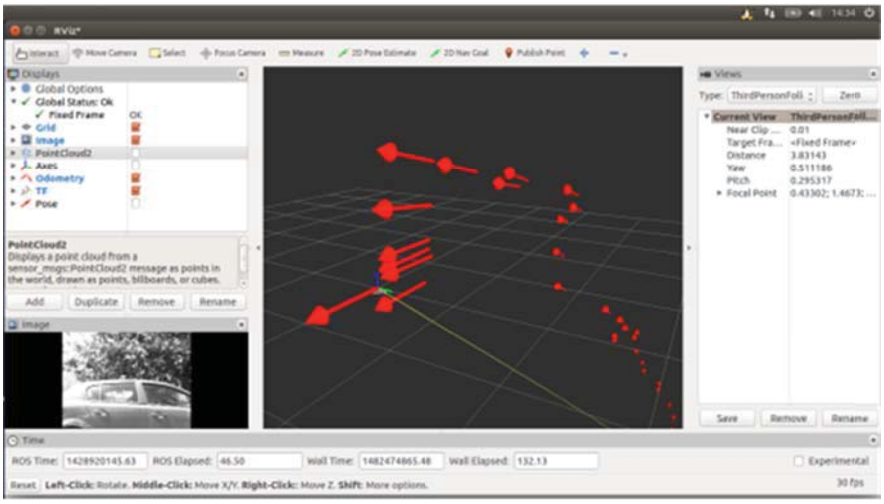
Let now π(l)(X;r,t):R3→R2 denote the projection implied by Eq, 1, which takes a 3d point X and maps it to a pixel x(l)i∈R2 on the left image plane. Similarly, let π(r)(X;r,t) be the projection onto the right image plane. Using Gauss-Newton optimization, we iteratively minimize

$$\sum_{i=1}^{N} \left\| x_i^{(l)} - \pi^{(l)}(X;r,t) \right\|^2 + \left\| x_i^{(r)} - \pi^{(r)}(X;r,t) \right\|^2$$

(11)

with respect to the transformation parameters$\begin{pmatrix} R(t) & T \end{pmatrix}$. Here $x_i^{(l)}$ and $x_i^{(r)}$ denote the feature locations in the current left and right images respectively. The required Jacobians $J_\pi(l)$ and $J_\pi(r)$ are readily derived from Eq. (10). To be robust against outliers, we wrap our estimation approach into a RANSAC scheme, by first estimating $\begin{pmatrix} R(t) & T \end{pmatrix}$ for 50 times independently using 3 randomly drawn correspondences. All inliers of the winning iteration are then used for refining the parameters, yielding the final transformation $\begin{pmatrix} R(t) & T \end{pmatrix}$.

This paper conducts experiments in the ros system terminal,and the experimental results are displayedin the visualization tool RVIZ.



PointCloud2
Displays a point cloud from a sensor_msgs::PointCloud2 message as points in the world, drawn as points, billboards, or cubes.

Type: ThirdPersonFoll

| Current View | ThirdPersonFoll... |
|---|---|
| Near Clip ... | 0.01 |
| Target Fra... | <Fixed Frame> |
| Distance | 3.83143 |
| Yaw | 0.511186 |
| Pitch | 0.295317 |
| Focal Point | 0.43302; 1.4673; ... |

ROS Time: 1428920145.63 ROS Elapsed: 46.50 Wall Time: 1482474865.48 Wall Elapsed: 132.13

Left-Click: Rotate. Middle-Click: Move X/Y. Right-Click: Move Z. Shift: More options.

Figure 3.1 Visual odometer operation display

Finally, the mutual communication between nodes is observed through the ROS operating system node rqt_graph.As shown in Figure 3.2:

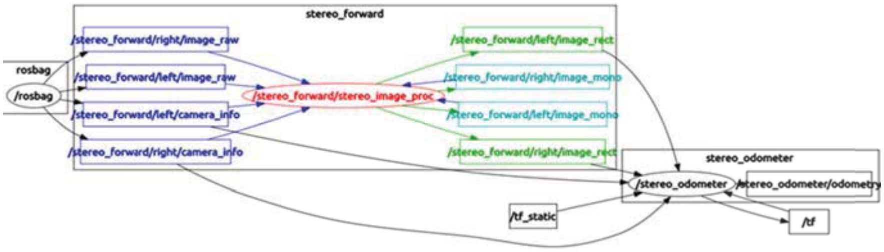

stereo_forward
/stereo_forward/right/image_raw
/stereo_forward/left/image_raw
/stereo_forward/left/camera_info
/stereo_forward/right/camera_info
rosbag
/rosbag
/stereo_forward/stereo_image_proc
/stereo_forward/left/image_rect
/stereo_forward/right/image_mono
/stereo_forward/left/image_mono
/stereo_forward/right/image_rect
stereo_odometer
/stereo_odometer
/stereo_odometer/odometry
/tf_static
/tf

Figure 3.2 Node communication diagram

## 4. Fusion of visual navigation and inertial navigation

In this paper, the visual navigation and inertial navigation are fused by the method of loose coupling. The method used is similar to the weighting method, but not the weighting method.That is not a channel on the visual odometer processing to get the rotation matrix and the translation vector, another channel for inertial navigation processing to get the rotation matrix and translation vector, and then the weights are assigned to both, finallywecalculate the average.But we propose a new fusion method, and the inertia processing function is used to process the measured values of the gyroscope and accelerometer in the inertial measurement element. This function outputs one-dimensional array of six elements. The output value of the visual odometer processing function is compared with the element of the array. And then the four residuals are calculated and merged into the Gauss-Newton algorithm. Then, the rotation matrix and the translation vector after the fusion are calculated.The weights of inertial navigation and visual parts can be given separately. The weights can be set empirically or experimentally according to the experimental results.The method has the advantages of simplicity and small calculation amount, and it is suitable for occasions where the real-time requirement is high.

Refering to the above method to study the fusion method, the structure of the visual inertial navigation fusion system is shown in Figure 4.1:

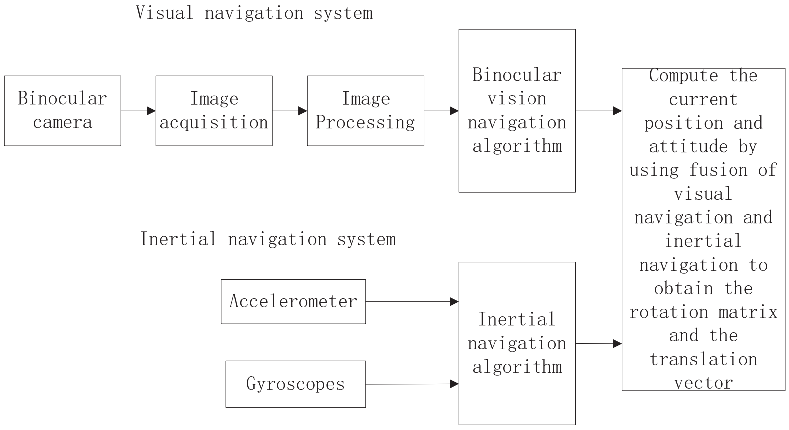

Visual navigation system
Binocular camera
Image acquisition
Image Processing
Binocular vision navigation algorithm
Compute the current position and attitude by using fusion of visual navigation and inertial navigation to obtain the rotation matrix and the translation vector
Inertial navigation system
Accelerometer
Gyroscopes
Inertial navigation algorithm

Figure 4-1 The fusion system of visual and inertial

The IMU has high reliability, short-term accuracy, and it is not affected by distance factors. It can output a variety of navigation information such as position, speed, and attitude at high frequencies, but the error is cumulative over time , The rotation matrix R and the translation vector T obtained from the short-term operation of the IMU can be referred to as the rotation matrix R and the translation vector T obtained by the visual odometer.If the rotation matrix R and the translation vector T calculated by the visual odometer at one time are too different from the rotation matrix R and the translation vector T calculated by the inertial navigation, there is a problem in the output of the visual odometer R and T.

Combined with the above principle, the R and T calculated by the IMU are integrated into the objective function of the visual odometer, and the new objective function is as follows:

$$F(\mathrm{R},\mathrm{T}) = \underset{R,T}{\operatorname{argmin}}\, k\sum_{i=1}^{n}\left\| m_i^{'} - \left(Rm_i + T\right)\right\|^2 + \left(1-k\right)\left\|(\mathrm{R},\mathrm{T})_{visual} - (\mathrm{R},\mathrm{T})_{imu}\right\|^2 \tag{12}$$

In the above formula, k represents a weight.In the program, there is an array called tr and named imu_tr, respectively.Tr represents a six-element array obtained by visual odometry, and imu_tr represents a six-element array obtained by inertial navigation.The six-element array is the key to calculating the rotation matrix, the Jacobian matrix and calculating the residuals.R and T in the objective function are computed by the six-element array, and R and T show the rotation matrix and the translation vector.Let the vector $r_i$ be the i-th row vector of the matrix R, and for the objective function, R needs to satisfy the constraint condition:

$$
\begin{aligned}
&R=\left[r_1^T, r_2^T, r_3^T\right], \\
&r_i^T r_i=1, i=1,2,3, \qquad (13) \\
&r_i^T r_j=0, i, j=1,2,3, i \neq j.
\end{aligned}
$$

we choose the Gauss-Newton method to solve it in the least squares method for solving the above objective function.After the visual and inertial fusion program is written, we run the program primarily by launching the launch file. The Launch file contains several parts as shown in Figure 4.2:

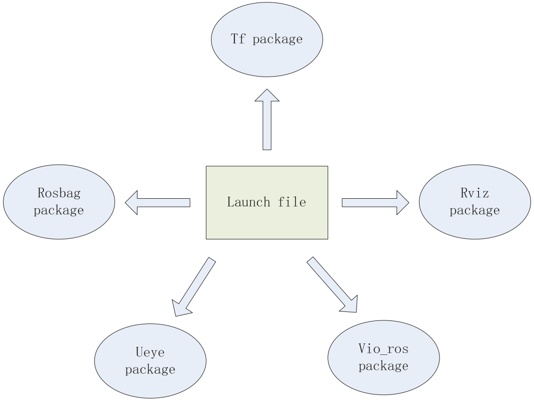
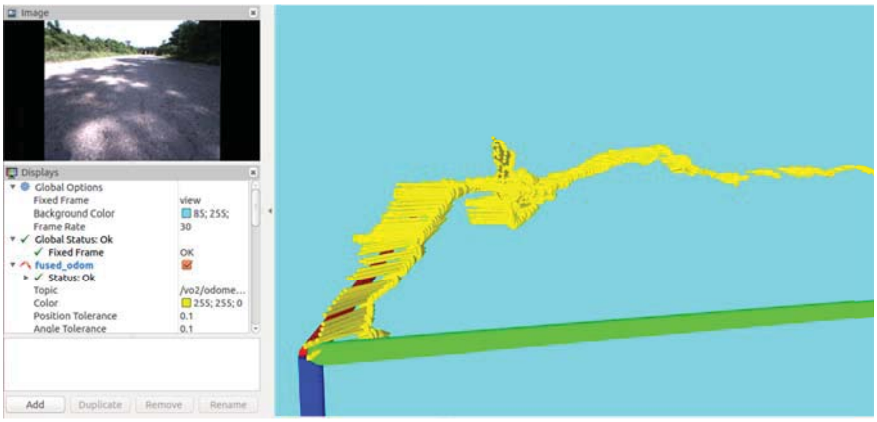


Figure 4.2 Packages involved in the program

Among them, tf package for setting some variables, so that nodes can publish messages to the subject, so as to achieve the purpose of communication between nodes and nodes. At the same time as both the camera and the IMU sensor, you need to set the operating frequency of both set to the same value, the experiment set to 10HZ.Rosbag package is mainly used to call the data, and ROS data is stored in the bag format, and we need to set the path of the packet.The Ueye package mainly contains the image information captured by the camera and the calibration result of the camera.Vio_ros package contains mainly the algorithm of visual inertia fusion algorithm source, but also it sets up some of the variables required for the path information and simultaneously it sets the target function of the value of the weight and so on.Rviz package is mainly used for visual displaying of the final results.



```
▾ Global Options
    Fixed Frame              view
    Background Color         🟦 85; 255;
    Frame Rate               30
▾ ✓ Global Status: Ok
    ✓ Fixed Frame            OK
▾ fused_odom                 ☑
  ▸ ✓ Status: Ok
    Topic                    /vo2/odome...
    Color                    🟨 255; 255; 0
    Position Tolerance       0.1
    Angle Tolerance          0.1
```

Figure 4.3Rviz displays the running results

## 5. Conclusion

In this paper, three localization algorithms are studied: inertial navigation based localization algorithm, visual navigation based localization algorithm and visual inertial navigation fusion robot localization algorithm.In the inertial navigation part, we mainly study the inertial navigation positioning process and the inertia navigation attitude calculation.Then, the method of stereo vision localization is discussed, and the robot localization is realized by LIBVISO2 odometer library.Using the low-cost IMUwhich has short-term stationarity in the image sampling interval, the rotation matrix R and the translation vector T obtained with the IMU limit the R and T obtained by the visual odometer.At last, the experiment is carried out, and the experimental result is displayed by the visualization tool RVIZ in the ROS operating system.

## References

[1] YuhongJI, RanghuiWang.The Technology and Application of Global Navigation Positioning System GNSS [J] .Global Positioning System,2010,05:69-75.

[2] YunweiJia, Tiegen Liu, LilanGao. Robot Localization and Navigation System Based on Monocular Vision[J]. Transactions of Tianjin University,2012,05:335-342.

[3]Xinwu Ma, Fujia Sun. Research on Extraction of Bottom of Shoe Pattern Based on Binocular Stereo Vision[J]. International Journal of Plant Engineering and Management,2016,01:20-34.

[4]Yongjie Zhao. Inverse displacement analysis of the general six degree-of-freedom serial robot based on optimization method[J]. Journal of Chongqing University (English Edition),2011,02:60-67.

[5]WeiChen. Based on Quaternion and Kalman Filter Attitude Angle Estimation

Algorithm Research and Application [D]. Yanshan University,2015.
[6]LingnanXia, BoZhang, YingguanWang, JianmingWei. Robot Localization Based on Inertial Sensor and Vision Odometer [J] .ActaAutomaticaSinica, 2013,01: 166-172.
[7] Han Lihua. Research on Mobile Robot Localization based on Inertial Navigation and Vision Sensor Information Fusion [D]. Chang'an University, 2012.