Image Tracking of Rotating Scene Based on Fusion of MEMS-IMU Data and Vision Data

Zhengzhong Huang1,*, Xiangcheng Tang1, Shuoyi Zou1, Zili Huang1, Linghui Li1, Jianrong Shang1
1Information Processing Center Southwest Institute of Technical Physics Chengdu, China
2Office in Chengdu Area Rocket Force Military Deputy Chengdu, China

*Corresponding author e-mail: tomhzz@126.com

Abstract. Image tracking is a research field in constant evolution due to the new technological advances. The limited-space and low-cost features of missiles restrict the application and development of image tracking. Missile-borne image guidance system is still facing a series of difficult issues of anti-interference. This paper constructs a practical image tracking model using Kalman Filter based on fusion of MEMS-IMU Data and Vision Data after deriving the transformation matrix between the image sequences and MEMS-IMU data. The tracking model is tested by experiments. The result shows tracking based on fusion of MEMS-IMU Data and Vision Data is an effective solution of rotating scene and tracking accuracy is about 4 times better than vision only. The achievements can be not only used in military field such as Precision Guided Weapons and Photoelectric Reconnaissance Device, but also broadly applied in commercial products such as Virtual Reality, Intelligent Robot, Unmanned Airborne Vehicle and so on.

1. Introduction

Image guidance is one of the most important and commonly used guidance modes of precision guided weapons [1]. As the core technology of image guidance, image tracking is becoming mature. The existing technical system of image tracking can basically support the guidance mode of fire-and-forget, but it is still facing a series of difficult issues of anti-interference: the image tracking accuracy is limited by the resolution of the image detector; the latency of tracking loop mostly depends on the processor's ability; the relative motion between the missile and target makes the image scene present a wide range of scale and rotation changes[2]; other anti-interference problems such as illumination change, shadow, target shelter, targets cross-movement and etc[3]. Image tracking can hardly overcome these existing difficulties use vision data only. A bottleneck is in the way of the development of image tracking field [4].

Apart from the image sensor, the precision guided weapons is equipped with other non-imaging sensor such as MEMS-IMU, GPS, radar, laser ranging etc. MEMS-IMU is a tiny IMU. Its short-time measurement data is more accurate; small data size, mini memory resources, low processing algorithm complexity etc. MEMS-IMU is an active measurement sensor, which is not affected by the external environment. Image tracking and MEMS-IMU have many complementary features. Because the
inertial sensor is accurate in short time, it can be used to replace the approximate search process of image tracking. MEMS-IMU data assisted image tracking can save memory resources and processor resources, promote the application of complex tracking algorithm, and solve the conflict between real-time and complexity of the algorithm [5]. MEMS-IMU will not affected by the external environment, and the MEMS-IMU data can be used to predict the shape and position of the target, which can greatly improve the anti-interference performance of image tracking. At the same time, the accumulated error of the image data is small, which can correct the long-time drift of the MEMS-IMU in return.

Recently, the fusion of vision and IMU has received considerable attention in the research community. Corke et al. (2007) discussed how to integrate information from vision and IMU to provide a robust and non-ambiguous representation of robotic motion back in 2007. They cover the fundamentals of these two sensing modalities systematically from the perspectives of physical principles and the engineering and biological implementations [6]. Carrillo et al. (2012) used Kalman Filter for stereo visual odometry and IMU measurements fusion to provide accurate estimates of the UAV position and velocity [7]. Alatise et al. (2017) presented a fusion of an IMU of six degrees of freedom (6-DoF), and a vision to determine a low-cost and accurate position for an autonomous mobile robot [8]. WANG, Y. J. et al. (2015) integrated the

Sequence matching algorithm and IMU/visual odometry to improve the performance of image sequence matching and reduce the navigation error of inertial/visual odometry[9].Li, X. Y. et al. (2011) presented a new tightly-coupled hybrid tracking approach combining vision-based systems with an inertial sensor and applied in AR tracking[10]. Yang, X. et al. (2015) presented a hybrid tracking method using fusion of vision and inertial sensing for accurate and efficient pose tracking of planar targets on modern smartphones [11].

Prior work has documented the effectiveness of fusion of vision and IMU. The study of image tracking technology based on fusion of MEMS-IMU Data and Vision Data can make full use of the advantages both vision sensor and inertial sensor. It will help solving the complicated difficulties image tracking field is facing, and would contribute to the further development of image guidance.

2. Theory and Model

2.1. Transformation of Image Sequences

In the process of image tracking, the transformation relation of image \( I \) between two adjacent frames is usually called frame-to-frame transformation \( C \), and the formula is expressed as:

\[
I_{i+1} = C \cdot I_i
\]  

Suppose the target is a stable one won’t move, the pixel position of the target in the next frame is the frame-to-frame transformation of the pixel position in the current frame. For the missile image tracker, image transformation is mainly caused by missile attitude transformation. In other words, frame-to-frame transformation \( C \) is related to the rotation and translation of the missile. And the rotation and translation transformation of the missile can be measured by MEMS-IMU. So the
problem of establishing the bond between the image and MEMS-IMU turns to deduce the frame-to-frame transformation expression $C^e_r$.

The expression of $C^e_r$ is derived as below:

Firstly, the target is located in the geographic coordinate system $e$, which is converted to the image sensor coordinate system $u$ through translation and rotation transformation. Then $u$ is converted to the screen coordinate system $s$ through perspective transformation, so as to complete the conversion from target to pixel. Shown in Fig. 1.

The transformation matrix $C^e_r$ from the geographic coordinate $e$ to the camera coordinate $u$ is the inverse transformation from the camera coordinate $u$ to the geographic coordinate $e$,

$$ C^e_r = (T \cdot R)^{-1} = R^{-1} \cdot T^{-1} $$(2)

$R$ is orthogonal matrix, which can be obtained from the property of the orthogonal matrix $R^{-1} = R^T$. The expression of transformation from geographic coordinate system $e$ to camera coordinate system $u$ is:

$$
\begin{bmatrix}
X_u \\
Y_u \\
Z_u
\end{bmatrix} = 
\begin{bmatrix}
ux & uy & uz & 0 \\
vx & vy & vz & 0 \\
nx & ny & nz & 0
\end{bmatrix}
\begin{bmatrix}
1 & 0 & 0 & -tx \\
0 & 1 & 0 & -ty \\
0 & 0 & 1 & -tz
\end{bmatrix}
\begin{bmatrix}
X_e \\
Y_e \\
Z_e
\end{bmatrix}
$$

(3)

Rewrite the expression in Euler transformation form:

$$
\begin{bmatrix}
X_e \\
Y_e \\
Z_e
\end{bmatrix} = 
\begin{bmatrix}
\cos \alpha & -\sin \alpha & 0 \\
\sin \alpha & \cos \alpha & 0 \\
0 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
\cos \beta & 0 & -\sin \beta \\
0 & 1 & 0 \\
\sin \beta & 0 & \cos \beta
\end{bmatrix}
\begin{bmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
X_e \\
Y_e \\
Z_e
\end{bmatrix}
$$

(4)

Screen coordinate transformation is the process of mapping a 3D object in the camera space to the perspective of a 2D screen. The transformation matrix $C^s_u$ from the detector coordinate $u$ to the screen coordinate $s$ is

$$
\begin{bmatrix}
X_s \\
Y_s \\
Z_s
\end{bmatrix} = 
\begin{bmatrix}
\alpha & 0 & 0 & 0 \\
0 & \beta & 0 & 0 \\
0 & 0 & 1 & 0
\end{bmatrix}
\begin{bmatrix}
X_u \\
Y_u \\
Z_u
\end{bmatrix} = C^s_u \cdot C^e_r
$$

(5)

$\alpha$ and $\beta$ are the scale factors of the image sensor means $X_u$ axis and $Y_u$ axis. They are the internal parameters of the image sensor [12].

$$
\begin{bmatrix}
X_s \\
Y_s \\
Z_s
\end{bmatrix} = C^s_u \cdot C^e_r \cdot C^e_u
$$

(6)

Since target is a stable that means $(X_e, Y_e, Z_e)$ is fixed. So from equation (6), a detailed expression of equation (1) is:
Expression of frame-to-frame transformation $C$ is:

$$ C = C_r^u C_r^{(i+1)} \left( C_r^{(i)} \right)^{-1} \left( C_r^{(i+1)} \right)^{-1} $$  \hfill (8)

The transformation matrix $C_r$ is determined by image sensor. The transformation matrix $C_r^u$ is related to $\alpha, \beta, \gamma, tx, ty, tz$ which can be measured by MEMS-IMU.

2.2. Direct-Trans Model

Direct-Trans Model directly using the transformation between two adjacent frames. From equation (8), next frame $I_{i+1}$ can be predicted by current frame $I_i$ and $\alpha, \beta, \gamma, tx, ty, tz$. The schematic diagram is shown in Fig. 2.

![Figure 2. Schematic Diagram of Direct-Trans Model](image)

Construct Kalman Filtering model as below.

Select $x = [I_i, q_i, \omega_i]^T$ as the state vector, where $I_i$ means frame, $q_i$ means quaternion, $\omega_i$ means angular velocity. The state equation is:

$$ \hat{x}_i = F_i \hat{x}_{i-1} + B_i \bar{u}_i = $$

$$ \begin{bmatrix} C_r^u C_r^{(i)} \left( C_r^{(i)} \right)^{-1} \left( C_r^{(i+1)} \right)^{-1} \cdot I_{i-1} \\ \exp \left( \frac{\omega_{i-1} \cdot t}{2} \right) \otimes q_{i-1} + B_i \bar{u}_i \end{bmatrix} $$  \hfill (9)

Equation (9) is a nonlinear filtering problem, which needs to be solved by EKF [13].

As the frame frequency of the missile-borne image sensor is above 50Hz, the single image processing time is only 20ms. In addition to the image tracking algorithm time and servo control time, the fusion of MEMS-IMU and image tracking solution time is less than 3ms. Therefore, in practical application, the Direct-Trans Model cannot meet the real-time requirement. Combined with the practical application of image guidance, we have done an appropriate approximation and linearization.
2.3. Diff-Pixel Model

Diff-pixel means the difference pixel between current frame $I_t$ and next frame $I_{t+1}$. The essence of target image tracking is to count the change pixels between two frames and predict the position of the target in the next frame. The main principle of diff-pixel model is to optimize and fuse the complementary information of MEMS-IMU and image sensor in temporal dimension and spatial dimension with the equivalent difference pixel deviation as the intermediate variable, establish the mapping relationship between inter-frame motion and inter-frame pixel deviation, and achieve accurate target tracking. The schematic diagram is shown in Fig. 3.

![Figure 3. Schematic Diagram of Diff-Pixel Model](image)

Since the high frame frequency and fast image updating speed, use the last frame image as the state estimation. The MEMS-IMU data can used as the observation value. Construct Kalman Filtering model as below.

Select $x_k = [I_t]^T$ as the state vector, where $I_t$ means frame. The state equation is:

$$\hat{x}_k = F_k \hat{x}_{k-1} + B_k u_k = \hat{x}_{k-1} + u_k$$  \hspace{1cm} (10)

Select $z_k = H_k \hat{x}_k + \gamma_k$ as observation vector, $\gamma_k$ is the MEMS-IMU error. Since the image period time is only 20ms, $\gamma_k$ is Short-Time error, approximately the Angular Random Walk[14].

From equation (8):

$$H_k = C^u_{y(i)} C^u_{x(i+1)}^{-1} \left( C^u_x \right)^{-1}$$  \hspace{1cm} (11)

The observation equation is:

$$\hat{x}_k = \hat{x}_k + K' (z_k - \hat{z}_k)$$  \hspace{1cm} (12)

3. Experiment and Analysis

Experiment is designed to test tracking model using Kalman Filter based on fusion of MEMS-IMU data and vision data. Vision data and MEMS-IMU data captured simultaneously. The data processing and estimation process were analyzed offline separately by fusion of MEMS-IMU data and vision data versus by vision data only.

3.1. Experimental Setup

Fig. 4, shows the major hardware used to carry out the experiment. The camera is a FOV28° CCD with a resolution of 1280×1024 at 50 fps. The MEMS-IMU is ADIS16445 which is produced by ADI whose In-Run Bias
Stability is $12^\circ/\text{hr}$ and Angular Random Walk is $0.56^\circ/\sqrt{\text{hr}}$. The axis of the camera is in alignment with MEMS-IMU’s reference direction. Camera and MEMS-IMU are fixed in a capture PCB. The capture PCB is mounted in a three-axis platform. Image and MEMS-IMU data is captured and stored in a computer. While data capturing, the platform is rotated to generate an image sequence with rotation. The data processing and estimation process were analyzed offline using MATLAB software.

3.2. Experimental Results

![Figure 4. Experimental Platform](image)

**Figure 4.** Experimental Platform

![Figure 5. Experimental Results of Rotating Scene](image)

**Figure 5.** Experimental Results of Rotating Scene. (a) show results of tracking with vision data only; (b) show results of tracking with fusion of MEMS-IMU data and vision data.
Fig. 5 shows image tracking results of rotating scene. Experiment tested separately tracking with vision data only—or vision mode—and tracking with fusion of MEMS-IMU data and vision data—or fusion mode. Two tests started at the same position and carried in the same condition. Fusion mode is quite stable till the end frame 486. While vision mode stray from the target at frame 130 and lost target at frame 486. In fact vision mode completely lost target at frame 423 already.

Fig. 6 shows tracking trajectory results. At first view vision mode’s trajectory has obvious fluctuations, but fusion mode’s trajectory suits actual target trajectory very well.

Fig. 7 shows tracking deviation results. The experiment has set a limitation of 8 pixels. Vison mode’s tracking error always above 2 pixels and up to 8 pixels, the RMSE_x= 3.51 pixels, RMSE_y=4.84 pixels. Vision mode completely lost target at frame 423. Fusion mode’s tracking error is about 1 pixel, the RMSE_x= 0.94 pixels, RMSE_y=0.68 pixels. But at about frame 410 fusion mode’s tracking error increases to 4~5 pixels. Tracking stability should pay more attention and further study need to follow.

Experimental results proved that the tracking accuracy of fusion mode is about 4 times better than vision mode, the tracking model is reasonable and tracking based on fusion of MEMS-IMU Data and Vision Data is an effective solution of rotating scene.
4. Conclusion
Our study embarks from the actual problem of missile-borne image guidance system. Concluded a series of difficult issues which missile-borne image guidance system is facing by system analysis missiles’ features. After discussed the advantage complementary between vision and MEMS-IMU, a practical image tracking model using Kalman Filter based on fusion of MEMS-IMU Data and Vision Data has been proposed. Experiment is designed to test the tracking model. The results proved that tracking based on fusion of MEMS-IMU Data and vision data is an effective solution of rotating scene. The theory can be used to deal with other difficulties missile-borne image guidance system is facing, such as scale changes, illumination change, shadow, target shelter, targets cross-movement and etc. Additionally, with the help of MEMS-IMU to decrease the operand, more complexity image tracking algorithm can be applied in image guidance system. Further work will be required to study tightly coupled fusion system and introduce more other sensors such as Radar, Lidar and GPS to seek new breakthroughs.

References
[1] Yin, X. M. , Zhang, H. . (2017). An introduction to the application and development of guidance technology. 2017 Weapon engineering conference. OSEC.
[2] Cui, P. G. . (2018). Research on Robust Tracking Technology for Blocking Interference in Complex Scenes[D]. University of Chinese Academy of Sciences.
[3] Wang, B. X. . (2015). Research of target detection and tracking algorithm based on online learning[D]. China Research and Development Academy of Machinery Equipment.
[4] Tang, T. C. . (2015). Technology Research of Image Tracking Based on MEMS-IMU Information Fusion[D]. China Research and Development Academy of Machinery Equipment.
[5] Tang X. C., Tian J. W., Huang Z. Z. & Liu D. S. (2018). Subpixel Correlation Tracking Method Based on MEMS-IMU Information Fusion. INFRARED, VOL. 39, NO. 3, 29-35.
[6] Corke, P., Lobo, J. and Dias, J. (2007). An introduction to inertial and visual sensing. The International Journal of Robotics Research, 26(6), 519–35.
[7] Carrillo, L.R.G., López, A.E.D., Lozano, R. and Pégard, C. (2012). Combining Stereo Vision and Inertial Navigation System for a Quad-Rotor UAV. Journal of Intelligent Robot Systems, 65, 373–87.
[8] Alatise, M. B., & Hancke, G. P. (2017). Pose estimation of a mobile robot based on fusion of imu data and vision data using an extended kalman filter. Sensors, 17(10), 2164-. 
[9] Wang, Y. J. , Hu, X. P. , Lian, J. X. , Zhang, L. L. , He, X. F. . (2015) Inertial/visual odometry aided sequence image matching method. 7th Conference of Chinese Society of Inertial Technology.
[10] Li, X. Y. , Chen, D. Y. . (2011). Tightly-coupled multi-sensor hybrid tracking algorithm. Journal of Image & Graphics.
[11] Yang, X. , Si, X. , Xue, T. , & Cheng, K. T. T. , (2015). [POSTER] Fusion of Vision and Inertial Sensing for Accurate and Efficient Pose Tracking on Smartphones. 2015 IEEE International Symposium on Mixed and Augmented Reality (ISMAR). IEEE Computer Society.
[12] Lobo, J. , & Dias, J. . (2007). Relative Pose Calibration Between Visual and Inertial Sensors. Sage Publications, Inc.
[13] Sadeghzadeh-Nokhodberiz, N. , & Poshtan, J. . (2015). Interconnected maximum likelihood estimator and extended kalman filter for inertial measurement unit calibration fusing three-dimensional camera information. Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering, 229(2), 106-117.
[14] Zhou, S. , Fei, F. , Zhang, G. , Liu, Y. , & Li, W. . (2014). Hand-writing motion tracking with vision-inertial sensor fusion: calibration and error correction. Sensors, 14(9), 15641-15657.