Research on Virtual–Real Occlusion Fusion Technology Based on Terrain Information

Yang Lu*, Deyou Xu, Haibo Lu
Army Arms University, Nanjing, Jiangsu, China
**Corresponding Author*

Abstract: To address the core problem that, in augmented reality training scenarios, existing technologies can hardly balance the accuracy of terrain-occlusion determination and computational efficiency when large virtual targets are fused with real environments, this paper proposes a virtual–real occlusion fusion method based on terrain information. This method constructs a complete technical pipeline of "data preprocessing-model optimization-spatial alignment-occlusion determination." Through key algorithms such as feature-preserving lightweight processing of terrain models, high-precision spatial alignment based on multi-source pose fusion, and rapid occlusion determination driven by terrain parameters, a systematic solution suitable for wide-area geographic scenarios is formed. Experimental results show that, while preserving critical terrain features and ensuring the accuracy of occlusion determination, the proposed method greatly reduces model computational overhead and significantly improves the real-time rendering performance and visual consistency of large virtual target fusion, thereby providing reliable technical support for the application of wide-area scenarios in simulation training.

Keywords: Augmented Reality; Virtual-Real Occlusion; Terrain Information; Large-Scale Terrain; Real-Time Rendering

1. Introduction

Augmented reality technology can accurately integrate virtual information with real scenes, thereby providing highly immersive visual support for simulation training. In wide-area training scenarios, the rationality of occlusion in the virtual-real fusion of large virtual targets is mainly determined by the spatial structure of wide-area terrain, which is essentially different from that of near-field small targets focusing on detailed occlusion. Existing virtual–real occlusion technologies are mostly designed for small-scale targets and near-field scenarios. When applied to large-scale terrain scenes, they are prone to problems such as excessive model loading, insufficient spatial alignment accuracy, and low occlusion-computation efficiency, making it difficult to simultaneously achieve occlusion-determination accuracy and real-time operational performance. Therefore, focusing on the application requirements of virtual–real fusion for large training targets, this paper proposes a virtual–real occlusion fusion method based on terrain information, with the aim of solving the trade-off between occlusion accuracy and computational efficiency in complex wide-area scenarios and providing a feasible technical solution for the further application of augmented reality technology in wide-area training scenarios.

2. Application Value of Virtual-Real Occlusion Fusion Technology Based on Terrain Information

2.1 Enhancing the Realism of Large Virtual Target Fusion

Visual consistency is a core factor determining the immersion and credibility of simulation training. For large virtual training targets, the rationality of their occlusion relationship with real terrain directly determines the visual realism of the scene. Once the occlusion relationship is misaligned, phenomena violating physical laws may occur, such as a runway "penetrating" a hill or fortifications "floating" above high ground, which will seriously undermine immersion and scene credibility. Different from near-field small targets that need to focus on detailed occlusion, the visual realism of large virtual targets is entirely determined by the spatial shielding relationship of wide-area terrain structures. High-precision terrain parameters such as elevation and slope are the key basis for accurately determining the visible and occluded regions of virtual targets. They can reproduce terrain-occlusion relationships that conform to real-world rules and provide highly credible visual support for simulation training.

2.2 Ensuring the Spatial Accuracy of Virtual–Real Scene Fusion

In wide-area scenarios, the spatial accuracy of virtual-real fusion is the core prerequisite for occlusion determination, and the key lies in realizing high-precision spatial alignment between the field of view of the observation-and-aiming device and the terrain model. The matching accuracy among the pose of the observation-and-aiming device, the spatial position of the virtual target, and the terrain model directly determines the rationality of the occlusion relationship. Even if the occlusion-determination logic is complete, any misalignment in spatial mapping will directly lead to a global disorder of the occlusion relationship. By accurately binding the terrain model to a standard geographic coordinate system and combining training-level coordinate transformation and error-correction mechanisms, frame-by-frame precise mapping between the real-time observation image of the observation-and-aiming device and the geographic space of the terrain model can be achieved, thus completing strict spatial alignment between the observation-and-aiming device and the terrain data and providing a reliable spatial reference for occlusion determination in wide-area scenarios.

2.3 Improving the Timeliness of Training-Scenario Construction

Training and exercise areas often span tens to hundreds of square kilometers. Traditional high-precision terrain-modeling methods tend to generate excessively large models, resulting in overly high loading latency at the terminal side and making it difficult to meet the application requirements of rapid-response scenarios such as contingency training and cross-regional maneuver exercises. A lightweight processing workflow for wide-area terrain can substantially compress model size while preserving the key terrain structures that determine occlusion relationships, thereby adapting to the limited computing power of training observation- and-aiming devices and effectively avoiding model-loading stutter and computational delay. At the same time, it can support the rapid stitching and local updating of terrain models in multiple regions, significantly improving the efficiency of constructing and deploying wide-area training and exercise scenarios.

3. Technical Difficulties of Virtual–Real Occlusion Fusion in Wide-Area Scenarios

3.1 The Difficulty of Lightweight Modeling While Preserving Wide-Area Terrain Features

To adapt to the limited computing power of training observation terminals, it is necessary to simplify high-precision terrain models in a lightweight manner. However, conventional mesh-simplification algorithms tend to over-smooth terrain, resulting in distortion of key terrain features that directly determine occlusion relationships and consequently causing incorrect occlusion judgments. Meanwhile, it is difficult to balance the simplification weights between large-scale gentle terrain and key occluding terrain: excessive simplification will cause the loss of core occlusion structures, whereas insufficient simplification will lead to excessively large model size and bring about loading and rendering delays. Therefore, how to accurately preserve the key terrain structures that determine occlusion relationships during the simplification process while at the same time realizing efficient model lightweighting is the core trade-off problem that this study needs to solve.

3.2 The Difficulty of Ensuring the Robustness of Spatial Alignment between Observation-and-Aiming Devices and Terrain Models

In wide-area scenarios, virtual–real spatial alignment involves wide-area geographic coordinate mapping and multi-sensor data fusion, and it is the core prerequisite for occlusion determination. In training environments, positioning signals are prone to attenuation or multipath effects in mountainous areas and in environments with strong electromagnetic interference. Angle encoders and attitude sensors are likely to accumulate errors under conditions of long endurance and high maneuverability. In addition, the numerical precision and algorithmic stability of wide-area coordinate transformation also face stringent application requirements. Even a slight error in spatial alignment may lead to an overall misalignment of occlusion relationships over a wide area. Therefore, how to ensure the alignment accuracy and robustness between observation-and-aiming devices and terrain models in complex battlefield environments is a fundamental issue that must be resolved for the practical implementation of this technology.

3.3 The Difficulty of Balancing Accuracy and Real-Time Performance in the Occlusion Determination of Large Virtual Targets

Traditional virtual–real occlusion algorithms mostly employ techniques such as pixel-by-pixel depth detection and ray casting. These methods are mainly designed for detailed occlusion scenarios involving small targets, and they incur high computational overhead. In wide-area terrain scenarios, they are prone to severe computational delays and therefore cannot meet the real-time requirements of training scenarios. In contrast, the occlusion determination of large virtual training targets mainly focuses on the stable shielding effect caused by wide-area terrain structures and does not require excessive processing of pixel-level detailed occlusion. Therefore, how to design a lightweight occlusion-determination criterion suitable for wide-area scenarios and consistent with the principles of training terrain analysis, so as to greatly improve computational efficiency while ensuring judgment accuracy, is the key difficulty in achieving smooth real-time fusion of large virtual targets.

4. Implementation Method of Virtual–Real Occlusion Fusion Technology Based on Terrain Information

4.1 Terrain Data Preprocessing and Construction of a High-Precision Three-Dimensional Geographic Scene

In order to obtain high-quality terrain data supporting occlusion determination in wide-area scenarios, this study adopts terrain data collected collaboratively by UAV-borne LiDAR and oblique photography cameras as the basic data source, and constructs a three-dimensional terrain model with both geometric accuracy and texture information through a standardized preprocessing workflow, thereby providing complete and reliable basic data support for subsequent technical procedures.

First, for the acquired LiDAR point-cloud data and multi-view oblique-photography images, multi-source data registration and fusion are completed so as to realize accurate matching between terrain geometric structure and texture information [1]. Subsequently, gross-error correction is performed on the fused initial data to remove abnormal gross-error data and invalid information, thereby generating an initial high-precision three-dimensional terrain scene model and providing a unified data foundation for the subsequent steps of model lightweighting, spatial alignment, and occlusion determination.

4.2 Denoising and Lightweight Simplification of the Three-Dimensional Scene Model with Wide-Area Feature Preservation

To address the trade-off between wide-area terrain feature preservation and model lightweighting, this study completes model denoising and lightweight processing in two steps:

Model denoising. Statistical filtering, radius filtering, and other point-cloud filtering algorithms are adopted to automatically identify and remove point-cloud data of non-solidified objects such as leaves, weeds, and temporarily stacked materials. Statistical filtering removes outliers by calculating the statistical characteristics of neighborhood distances for each point cloud. The core decision formula is:

$$d_i = \frac{1}{k}\sum_{j=1}^{k} \| p_i - p_j \| \quad (1)$$

In the formula: p_i denotes the target point cloud, p_j represents the k-nearest neighbor point clouds of p_i and d_i indicates the average distance between p_i and its k-nearest neighbor point clouds. If d_i exceeds a predefined threshold, p_i is classified as an outlier and removed.

Through the above filtering process, non-solidified invalid point clouds can be effectively removed, and only rigid terrain data of core value for occlusion determination are retained.

Feature-preserving model simplification. For rigid terrain data such as ground surfaces and mountains retained after filtering, this study adopts an edge-collapse mesh simplification algorithm prioritizing wide-area terrain features [2]. By reducing the number of triangular facets and merging contiguous gentle-terrain regions, the volume of the model is compressed. The edge-collapse cost function is used to evaluate the influence of collapsing a certain edge on terrain features:

$$Cost(e_{ij}) = \alpha \cdot ||v_i - v_j||^2 + \beta \cdot |\kappa_i - \kappa_j| + \gamma \cdot |s_i - s_j| + \delta \cdot T_{mil}(e_{ij}) \quad (2)$$

Where:

e_{ij}: denotes the edge to be collapsed, connecting vertices v_i and v_j;

$||v_i - v_j||^2$: denotes the geometric distance term, used to control the overall simplification intensity of the model;

$|\kappa_i - \kappa_j|$: denotes the vertex-curvature difference term, used to preserve the undulating characteristics of the terrain;

$|s_i - s_j|$: denotes the slope-change difference term, used to preserve key occluding terrain such as steep cliffs and reverse slopes;

$T_{mil}(e_{ij})$: denotes the key terrain-feature weight, set for tactically relevant terrain such as ridgelines, commanding heights, and reverse slopes. The higher the weight value, the less likely the edge is to be collapsed;

α , β , γ , δ : denote adjustable weighting coefficients, which can be configured according to tactical requirements.

During the simplification process, the algorithm preferentially collapses edges with smaller values of the cost function, thereby ensuring that the geometric accuracy of key occluding terrain such as hills and high ground is not lost. It can simplify a terrain model with millions of facets to the order of hundreds of thousands of facets, greatly reducing model complexity and significantly improving the efficiency of subsequent terminal deployment and occlusion computation.

4.3 Model Deployment with Geographic Coordinate-System Binding and High-Precision Spatial Alignment

To address the difficulty of ensuring the spatial alignment accuracy and robustness between observation-and-aiming devices and terrain models in complex battlefield environments, this study mainly completes two core tasks so as to realize strict spatial matching between the field of view of the observation-and-aiming device and the terrain model and avoid global misalignment of occlusion determination caused by pose deviation.

The first task is binding the model to the geographic coordinate system. The preprocessed three-dimensional terrain model is deployed to the training observation-and-aiming device, and the model is accurately bound to the WGS84 standard geographic coordinate system of the target area. Longitude, latitude, and elevation information corresponding to each vertex of the model are assigned, thereby establishing a one-to-one mapping relationship between the terrain model and real geographic space.

The second task is real-time pose estimation based on multi-source data fusion**.** Through the BeiDou positioning unit and angle-encoder measurement unit integrated in the observation-and-aiming device, the system acquires in real time the device’s longitude and latitude coordinates as well as position and attitude information such as pitch angle, heading angle, and roll angle. Through a training-level coordinate-transformation algorithm, a mapping chain from the device observation image to the three-dimensional terrain model is established [3]. The homogeneous- coordinate transformation formula of the coordinate-transformation chain is:

$$P_{model} = T_{grid \leftarrow WGS84} \cdot T_{WGS84 \leftarrow ENU} \cdot T_{ENU \leftarrow device} \cdot P_{device} \quad (3)$$

Where:

P_{device} : denotes the homogeneous coordinates of the observation point in the device coordinate system;

$T_{ENU \leftarrow device}$: denotes the transformation matrix from the device coordinate system to the local east-north-up(ENU)horizontal coordinate system, calculated from the device’s pitch, roll, and yaw angles;

$T_{WGS84 \leftarrow ENU}$: denotes the transformation matrix from the local ENU coordinate system to the WGS84 geographic coordinate system, which needs to be calculated in combination with the coordinates of the reference point in the observation area;

$T_{grid \leftarrow WGS84}$: denotes the transformation matrix from the WGS84 coordinate system to the training grid coordinate system;

P_{model} : 3D denotes the corresponding homogeneous coordinates in the three-dimensional terrain-model coordinate system.

To suppress cumulative sensor errors and interference noise and improve the accuracy and robustness of pose estimation, this study adopts the Kalman filtering algorithm to fuse and correct multi-source sensor data such as BeiDou positioning and angle encoders [4]. The state equation and observation equation of the Kalman filter are respectively:

State equation:

$$x_k = Fx_{k-1} + Bu_k + w_{k-1} \quad (4)$$

Observation equation:

$$z_k = Hx_k + v_k \quad (5)$$

Where:

x_k: denotes the system state vector at time k, including the device's pitch angle ϕ, roll angle θ, yaw angle ψ, position coordinates p_x, p_y, p_z, angular velocity $\dot{\phi}, \dot{\theta}, \dot{\psi}$, and linear velocity v_x, v_y, v_z;

$$|x_k| = \begin{bmatrix} \phi, \theta, \psi, p_x, p_y, p_z, \\ \dot{\phi}, \dot{\theta}, \dot{\psi}, v_x, v_y, v_z \end{bmatrix}^T \quad (6)$$

F: denotes the state transition matrix, describing the change of the system state from time $k-1$ to time k;

B: denotes the control input matrix, where u_k denotes the control input vector;

w_{k-1}: denotes process noise, which follows a Gaussian distribution with mean 0 and covariance matrix Q.

z_k: denotes the sensor observation vector at time k, including observed values such as the position coordinates from BeiDou positioning and the attitude angles from the angle encoder;

H: denotes the observation matrix, associating the system state with the observation values;

v_k: observation noise, following a Gaussian distribution with mean 0 and covariance matrix R.

Through the two-step iteration of prediction and update in Kalman filtering, the globally optimal estimate of the device pose can be solved in real time. In this way, the pixel positions in the device's real-time observation image can be accurately mapped to the corresponding geographic coordinate positions in the three-dimensional terrain model, ultimately realizing strict spatial alignment between the device field of view and terrain data and providing an unbiased and precise spatial reference for the subsequent occlusion-determination stage.

4.4 Rapid Occlusion Determination Method Driven by Wide-Area Terrain Parameters

To address the difficulty of balancing the accuracy and real-time performance of occlusion determination for large virtual targets, and on the basis of high-precision spatial alignment, this study constructs a lightweight occlusion-determination method driven by wide-area terrain parameters for the occlusion-determination requirements of large virtual training targets. The core process is as follows. When a large virtual training target is deployed within the field of view of the observation-and-aiming device, the system first extracts the core spatial parameters of the virtual target, including the longitude-latitude boundary range of the target base and the elevation interval of the main structure. Subsequently, the spatial parameters of the virtual target are rapidly compared with the terrain-feature parameters of the corresponding region in the three-dimensional terrain model.

The core decision logic is as follows: For any point P_{obj} within the polygonal boundary of the virtual target's bottom surface, with coordinates (lon_{obj}, lat_{obj}) and bottom elevation h_{obj}, we search the 3D terrain model for the corresponding terrain point $P_{terrain}$ with elevation $h_{terrain}$. If $h_{terrain} > h_{obj} + \epsilon$ [5], the terrain area is determined to obstruct the virtual target's corresponding portion; otherwise, it is deemed non-obstructive. Here, ϵ represents a tolerance threshold used to mitigate minor spatial alignment errors in terrain data, typically set between 0.5m and 1m.

Then the corresponding part of the terrain region is judged to constitute occlusion to the corresponding part of the virtual target; otherwise, it is judged as non-occluding. In the formula, ϵ is a tolerance threshold used to eliminate the influence of slight errors in terrain data and spatial alignment, and it is usually set to 0.5 m to 1 m.

5. Experimental Verification and Analysis of Training Application Effects

To evaluate the effectiveness [6] of the proposed method in wide-area training scenarios, a hilly training area in Nanjing was selected as the experimental area. This area contains typical training terrain such as hills, gentle slopes, and gullies, and conforms to the characteristics of wide-area training scenarios. The experiment used a ruggedized training tablet as the test terminal [7]. Comparative tests were carried out using a traditional method [8] (control group) and the method proposed in this paper (experimental group). The core evaluation indicators included model-loading performance, real-time operational performance, occlusion-determination accuracy, and the visual effect of virtual–real fusion.

The experimental results show that, when the control group adopted the traditional high-precision model and pixel-by-pixel depth-detection algorithm, the system had a long initial loading time, the average time consumed for single-frame occlusion determination reached 33 ms, and the average operating frame rate of the terminal was only 11 fps. There was obvious rendering delay, memory occupation was high, and the device heated severely during long-term operation. Although its occlusion-determination accuracy reached 98.1%, its real-time performance could not meet the requirement for smooth fusion of large virtual targets.

After the experimental group adopted the method proposed in this paper, while completely preserving key occluding terrain features such as hills and high ground, the initial model-loading time was reduced by 74%, the average time consumed for single-frame occlusion determination was reduced to 4.3 ms, the average operating frame rate of the terminal remained stable above 38 fps, and memory occupation was reduced by 67%. After continuous operation, the device temperature rise was controlled within a reasonable range. Meanwhile, the occlusion-determination accuracy remained at 96.5%, showing only a slight and acceptable decrease and satisfying the accuracy requirements of applications.

References

[1] Liu Jun, Xiao Lei, Huang Xianfeng, et al. Rapid 3D reconstruction of scene environments combining UAV LiDAR and oblique photography. Journal of Surveying and Mapping, 2020, 49(10): 1287-1296.

[2] Garland M, Heckbert P S. Surface simplification using quadric error metrics//Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques. New York: ACM, 1997: 209-216.

[3] Zhang J, Zhang Y, Li J, et al. Real-time occlusion handling between virtual and real scenes in outdoor AR via GPU-accelerated depth fusion. IEEE Transactions on Visualization and Computer Graphics, 2021, 27 (11): 4221-4233.

[4] Kalman R E. A new approach to linear filtering and prediction problems. Journal of Basic Engineering, 1960, 82(1): 35-45.

[5] Wang Yong, Li Xiaopei, Wang Peng, et al. Real-time terrain occlusion processing method for augmented reality. Infrared and Laser Engineering, 2022, 51(5): 20210842.

[6] Jin Lute, Cheng Guo, Ye Zhuang, et al. Application of augmented reality technology in military affairs: A case study of map-guided training // China Simulation Society, Hefei Municipal Science and Technology Bureau. Proceedings of the 37th China Simulation Conference. National University of Defense Technology; 2025:97-103. DOI: 10.26914/c.cnkihy. 2025. 101895.

[7] Wang Mingyu, Wu Haiying, Deng Shujiang. Design and Implementation of a Mobile Augmented Reality Interactive Map System Based on Unity3D. Surveying, Mapping and Spatial Geographic Information, 2025,48(S2):107–109+113.

[8] Shi Hong. Research on Tracking and Registration Algorithms for Augmented Reality. Anhui University of Science and Technology, 2025. DOI: 10.26918/d.cnki. ghngc. 2025.001750.

[9] Monica S, Daniele D B, Francesco P. Grasping of Solid Industrial Objects Using 3D Registration. Machines, 2023, 11(3): 396-396. DOI: 10.3390/MACHINES11 030396.