Survey on Urban Warfare Augmented Reality

Xiong You [1], Weiwei Zhang [1,*], Meng Ma [2], Chen Deng [1] and Jian Yang [1]

[1] Zhengzhou Institute of Surveying and Mapping, Zhengzhou 50052, China; youarexiong@163.com (X.Y.); dclike@163.com (C.D.); jian.yang@tum.de (J.Y.)
[2] College of Computer, National University of Defense Technology, Changsha 410073, China; mameng437@gmail.com
*   Correspondence: jjxyzww@163.com; Tel.: +86-0371-8163-6047

Abstract: Urban warfare has become one of the main forms of modern combat in the twenty-first century. The main reason why urban warfare results in hundreds of casualties is that the situational information of the combatant is insufficient. Accessing information via an Augmented Reality system can elevate combatants' situational awareness to effectively improve the efficiency of decision-making and reduce the injuries. This paper begins with the concept of Urban Warfare Augmented Reality (UWAR) and illuminates the objectives of developing UWAR, i.e., transparent battlefield, intuitional perception and natural interaction. Real-time outdoor registration, information presentation and natural interaction are presented as key technologies of a practical UWAR system. Then, the history and current research state of these technologies are summarized and their future developments are highlighted from three perspectives, i.e., (1) Better integration with Geographic Information System and Virtual Geographic Environment; (2) More intelligent software; (3) More powerful hardware.

Keywords: augmented reality; urban warfare; situational awareness; registration; information representation; natural interaction

1. Introduction

Urban safety and security has always been of critical importance in winning a war, and hence, cities are often chosen as the targets of military operations. With the acceleration of global urbanization, more and more military operations are conducted in cities. In the Second World War, 40% of military engagements of European theatre took place in cities and large settlements and 90% of U.S. Marines' 250 overseas military interventions involved different cities. Therefore, urban area has become the main battlefield in the twenty-first century, and urban warfare has become one of the main forms of modern warfare.

Since urban area is fabricated with streets, buildings and underground facilities, urban terrain is considered a better place for defense than for offensive from the perspective of a military operation. It is easy for a defender to build well-protected fortresses and carry out fire attack, but relatively difficult for their opponents to exploit large scale attack strategies due to the blockage of the buildings. What's worse, the offensive side is vulnerable to heavy casualties without thorough intelligence concerning the urban context and the enemies. In the Chechnya war, 90% of Russian casualties were caused in the battle of Grozny, in which Russian's Army intended to siege and assault the Chechen capital. The case shows that the conditions for commanders to make correct decisions at all levels are accurately understanding of the complex urban terrain, timely grasping of the constantly changing states of enemies, friends and middle cube neutrals, and clearly maintaining the perception of the battlefield situation.

The commander usually improves the situational awareness of the battlefield via reading maps and marking situation symbols on maps (e.g., operational map). This involves complex cognitive

tasks that can be influenced by complex battlefield environment and subjective factors such as the commander's psychological state and operational experience. Therefore, different cognitive results can be yielded for the same battlefield situation. Furthermore, soldiers have to switch the line of sight between the map (heads-down) and the battlefield (heads-up), which imposes difficulties for them to maintain situational awareness.

Augmented Reality (AR), emerged in mid 1990s, is a technique that can enhance user's perception of surrounding environment through the use of helmet mounted displays, which overlays computer-created symbols of reality or virtual information on a real-world scene through registration in three-dimensional space. The technology has been supported by militaries around the world from the beginning. In recent years, more efforts are invested in this field in order to substantially improve the cognitive efficiency of combatants' situational awareness on the battlefield, which leads to a huge improvement of their capabilities of survival and collaborative operations. For example, ULTRA-Vis (Urban Leader Tactical Response, Awareness and Visualization) is such an AR-oriented research project launched by US's Defense Advanced Research Projects Agency (DARPA) in 2008, which intended to use AR to improve commander's situation awareness in urban warfare [1]. Through the perspective helmet mounted display, the system properly overlaid the geographical icons and information on soldiers' real vision, enabling them to perform tasks with the posture of "looking up at the top" and "trigger-on-the-trigger". This capability could improve soldiers' speed in warfare, strengthen command and control of the unit, and provide real-time situational awareness of friendly locations and tactical points [2]. It is necessary to point out that in such a system, real battlefield and battlefield information were connected for the soldier via geographic registration, making the warfighter in the course of action see the information registered with the battlefield accurately and enhancing their situation awareness. In other words, battlefield AR is not the enhancement of the battlefield itself, but to improve the perception of the battlefield. To achieve geographic registration, we need geographic coordinates and position and orientation of the soldier. Then, urban geographic information, urban virtual geographic environment and operational process model have to be employed to predict the operational situation.

There are substantial differences between urban warfare AR systems and common AR systems in terms of technical requirements. Urban warfare is conducted in a dynamic, indoor-outdoor integrated, serious electromagnetic-interference battlefield environment, and fast-changing mobility of soldiers requires equipment with high portability. However, current development is far behind the need of urban warfare. Thus, the capabilities of AR systems, especially stability, accuracy and usability, need to be improved. Towards this goal, it requires thorough review of key technologies, research problems and research trends. To begin with the concepts of Urban Warfare Augmented Reality (UWAR), this article discusses its basic characteristics, key technologies, core research problems, and predicts its future development, which can serve as a road map for researchers in the field.

2. Overview of UWAR

2.1. The Conception of UWAR

The correctness of a combat action and the speed of execution, will not only impact the success of a military confrontation, but also result in a significant difference in combatants' survival.

With the advancement of information, how to show the correct information at the right time in a proper way to the combatants has become the topic of in military science and technology innovation. After years of development, AR technology has gradually matured and will be adopted by a broad range of commercial applications. The popularity of the AR system will fundamentally change interactions between users and their surrounding environment and between users. The way users perceive environment and acquire information is also enhanced by the AR applications. By gaining access to information via AR technology, combatants can maintain a dominant position during the process of observing, locating, and destroying targets. AR systems can provide better situational

awareness, effectively reduce the friendly injury and indirect damage, and improve the efficiency of decision-making.

As a technology integrating a number of disciplines, AR registers between virtual world and real world via different sensors and then integrates computer-generated virtual elements (models, images, words, sounds, etc.) into the real-world space that can be naturally perceived by users, while allowing users to interact with the virtual elements in real time. AR can seamlessly mix the chosen information with users' world perceiving process, thereby enhancing users' cognitive capacity, reducing cognitive burden, and enhancing their ability to interact with information [3]. Military AR refers to the applications which enhance combat capability of troops or reduce military overhead in the military field using AR technology. Battlefield Augmented Reality (BAR) and Urban Warfare Augmented Reality (UWAR) are further refinements to AR applications in different military applications.

UWAR refers to the AR systems that enhance the combat capability of the combatants in the modern urban warfare environment via AR technology. The challenges of UWAR are:

(1) Keeping the system stable during intensive urban combat, in which soldiers need to move fast and often changes their head poses quickly.
(2) Developing a robust registration solution under poor operational conditions. For instance, vision-based registration would fail at night or in smoky environment because of poor imaging quality. Signal blockage and electromagnetic interference in urban battlefield will reduce the accuracy of GPS positioning.
(3) Improving computational efficiency of the UWAR system so as to retained portability of the overall computing hardware.
(4) Facilitating users, soldiers in urban battlefield, whose cognitive capabilities drop to a lower level due to their intense mental state.
(5) Enabling multi-tasking for soldiers during the operation.

2.2. The Effect of UWAR

The effectiveness of AR system in battle has been testified in quite a few experiments performed by researchers with different objectives over the years. Funded by the Canadian Ministry of defense, Colbert et al. evaluated the impact of AR in enhancing soldiers' situational awareness [4]. The experimental results and subjects' feedback showed that the system outperformed the traditional means of "map + compass" in terms of operational efficiency and cognitive workload in many tasks, such as aided navigation, way finding, target detection and judgment of azimuth and distance. Roux [5] performed a comparative study, in which commanders were equipped with different AR systems to evaluate how they benefitted from different command and control processes. He found that the AR system could provide soldiers with optimal path and team's locations to enhance their situational awareness. Moreover, the system could assist the commander in understanding and tracking battlefield information and enemy threats. Zysk et al. [6] stated that AR systems facilitated two capabilities for combat, namely enhanced situational awareness and precise navigation guidelines, since the highly-strained soldiers were prone to make mistakes when they used traditional maps to perceive the battlefield situation and AR provided an intuitive alternative in serving situational information to the combat personnel. Kenny [7] presented that the AR system could improve the efficiency of information processing pipeline including information collection, processing, exhibition and distribution, and help to distribute accurate information to the right fighters in the right time. In conclusions, the effectiveness of the UWAR system can be classified as:

(1) Enhancement of soldiers' ability to perceive battlefield information.
(2) Supporting commanders to do operational task planning.
(3) Empowering military training through a realistic battle scene simulation.

2.3. The Development of UWAR System

A number of countries such as the United States have begun the development of military augmented reality systems since 1990s. The two decades' efforts have resulted in many battlefield augmented reality systems, some of which related to urban operations are discussed as follows.

(1)   BARS

In 2000, the United States Naval Research Laboratory (NRL) developed a Battlefield Augmented Reality prototype System (BARS), which was mainly used in urban operations [8]. The system could provide situational awareness for infantry units in urban operations, and assist soldiers in carrying out operations when their sight was blocked, communication was insufficient, and it was difficult to distinguish enemy from friendly force. The system could also be used in vehicle division to improve the situational perception of driver. However, the system was only used in experiments and wasn't deployed in actual operations.

(2)   ARC4

The ARA company released a military Augmented Reality system, ARC4 (Augmented Reality Command Control Communicate Coordinate), in May 2014. This system provides commanders real world vision with accurately registered battlefield situational information, and showed the general COP to commanders and squads. ARC4 system also can provide the function of tracking, navigation, target delivery, image sharing, and tagging of features in the environment [9]. The system is also used in military training, in which virtual battlefield environment and virtual forces are superimposed on real battlefield. At present, the system is being tested and reformed.

(3)   TAR

Recently, a military AR conceptual system, Tactical Augmented Reality (TAR), is released by Communications-Electronics Research, Development and Engineering Centre (CERDEC) of U.S. Army. One of its core functions is to enhance commander's ability of battlefield situational awareness by using the Augmented Reality. Especially in urban warfare, the system can provide perception support of the battlefield situational information of integrated indoor- outdoor space [10,11]. The system can display real scene with indoor and outdoor targets, friendly forces and environment information in an overlapping way. Therefore, soldiers can obtain the battlefield situational information quickly and accurately.

Although a number of research efforts have been made, none of them can fully meet the requirements of urban warfare. In order to put augmented reality into practical use, further research work has to be done, such as to develop robust registration methods for extreme use cases, introduce intelligent information processing techniques into system building, and increase the computing power while retaining the portability of the system.

3. Key Technologies of UWAR

3.1. Architecture of UWAR

The UWAR system is mainly composed of real-time registration, information representation, human-computer interaction, wireless communication, control bus, urban battlefield environment database, urban operation augmented reality control cloud platform etc. (see Figure 1). The main task of the real-time registration module is to calculate the exact location of the battlefield information officers in the current scene. The information expression module is to display battlefield information in front of soldiers in humanoid form. The human-computer interaction module is employed to accurately understand users' command and respond quickly while minimizing disturbance to operations. The wireless communication module is responsible for uploading and downloading

information. The control bus module is mainly responsible for coordinating the transmission of information between modules, and assigning the computing and storage resources. The UWAR cloud platform on the one hand distributes commands, situational information, intelligence, battlefield environment information to every application terminal, and on the other hand processes reports, registration requests and collects environmental information from the terminals.
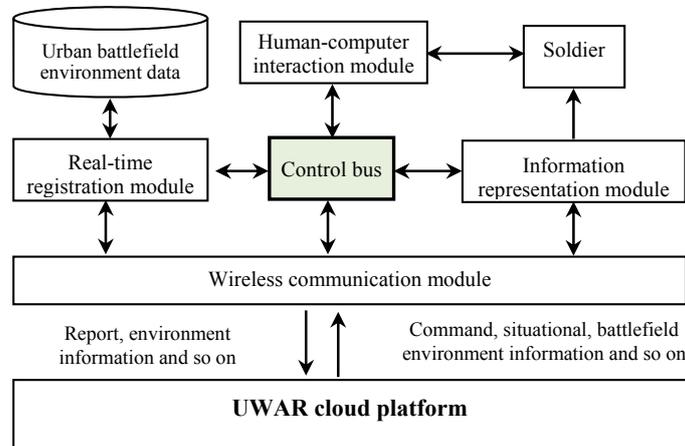

Figure 1. The architecture diagram of Urban Warfare Augmented Reality (UWAR).

The goal of UWAR is to accurately augment the battlefield situation information which soldiers need into the real world. To achieve this, it depends on three key technologies: real-time outdoor registration, information presentation, natural interaction. Figure 2 shows the composition of the three key technologies of UWAR and the relationship between them.
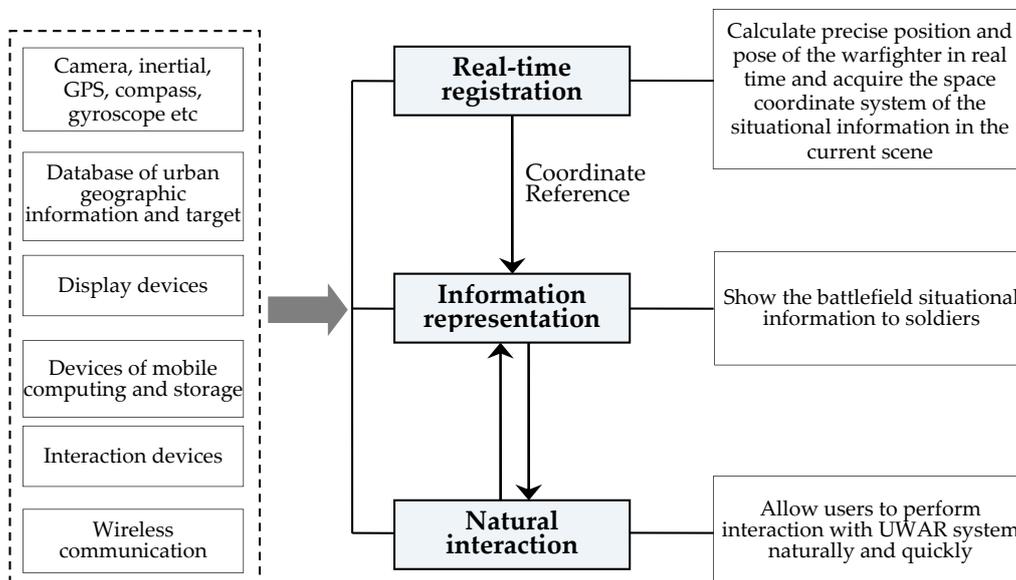

Figure 2. The composition of the three key technologies of UWAR.

3.2. Registration of UWAR

3.2.1. Main Issues of Registration Technology

In order to enhance the warfighters' situational awareness, the most important basis for UWAR is to provide real-time accurately geo-registered tactical information in the field of view of the soldiers.

The concept of geo-registration refers to rendering the tactical information in the correct position in the real world according to the geographic coordinate information of the target, such as longitude, latitude and elevation. Hence the target information and the corresponding object in the real world are accurately aligned in the view of the warfighter. Obviously, both the geographic information data and the surveying and mapping techniques can make a great contribution to achieve accurate geo-registration: Firstly, both the position of the warfighter and the location of the target need to be calculated in a unified geographical coordinate system. Secondly, the existing geographic information data can be directly used as the "augmentation" to highlight the main characteristics of the environment. Finally, the existing geographic information data can be used to support the estimation of the position of the warfighter; for example, there are several methods to estimate the current position of the warfighter, using the 2D maps, the DEMs or the 3D city models.

However, the distribution of targets in the urban warfare can be very dense, and the positioning signal is sensitive to urban terrain (e.g., signal blockage by high-rise building) and man-made interference. All of these would increase the difficulty of the registration. At present, the registration technologies are not good enough for the UWAR and should overcome impediments in three aspects: accuracy, robustness and real-time performance.

Besides of the acquisition of targets' coordinate and the scene 3D geometric model, the key challenge of registration technology is to obtain the user's position and orientation in real time. According to original data, registration methods are categorized into three classes: sensor-based, vision-based and hybrid methods. This section presents the registration methods and discusses their accuracy, efficiency and robustness from the perspective of a UWAR application.

3.2.2. Sensor-Based Registration

The sensor-based methods mainly contain magnetic sensor, inertial sensor (accelerometer and gyroscope) and GPS, etc. The inertial sensor can only get relative pose and need to be calibrated regularly as the error will accumulate over time. Earth magnetic sensor and GPS can obtain the orientation and three-dimensional position in the global coordinate system respectively, but the sensors on mobile terminal are usually less accurate and susceptible due to environmental disturbance (such as geomagnetic anomaly or blocked GPS signal).

In 1997, the first outdoor mobile AR used a head-worn differential GPS unit for global location and employed magnetometer and inclinometer to determine the orientation, and then directly displayed the registered geographic information in the user's field of view. The position accuracy was about 1 m. However, the three-axis orientation accuracy only measured by the geomagnetism and gravity was low, and the robustness was not strong enough [12].

In 2014, ARC4 team tried to develop an AR system in an outdoor unprepared environment to enhance the soldier's situational awareness. The system used GPS, accelerometer, gyroscope, magnetometer, barometer and other sensors to estimate the user's global pose in the outdoor field with less GPS interference. When the user moved at low speed, it could finish pose tracking in 2 ms to 40 ms and the orientation accuracy was 25 mard (about 1.45 degrees) [13]. Although the accuracy and robustness of pose tracking achieved by ARC4 was improved compared to the conventional INS-GPS framework, it still worked in a relatively ideal environment, because of the need for an accurate magnetic model and an open field with less GPS interference [14]. In the urban area with intensive building and geomagnetic interference, the above method is difficult to achieve the desired objectives.

In general, the advantages of sensor-based methods are characterized by high update frequency and low delay and suitable for operating in a wide range, but the accuracy is often not as high as that of the vision-based tracking since the vision-based methods can reach pixel-level accuracy when the quality of image is good enough. However, vision-based methods require more computing resource and are easily affected by quite a few environmental factors, such as light condition, so it is challenging to a realize low latency and strong robust vision-based tracking in outdoor [15].

3.2.3. Vision-Based Registration

According to the prepared dataset, the vision-based registration methods can be divided into image-based method, model-based method and combination of vision-based method (see Table 1).

Table 1. Vision-based registration methods.

Vision-Based Registration Methods	Algorithm	Basic Principle	Related Researches
Image-based method	Image retrieval-based method	The current image was located by using the most similar reference image's (GPS) position	Schindler 2007
			Zamir 2010 Zamir A.R. 2012
	Pre-reconstruction based on SFM	Pre-reconstruction the 3D point cloud based on SFM, and then restore the pose using (PnP) algorithm	Irschara 2009 Li 2012 Lim 2012 Lim 2015
			Wagner 2010 Arth 2009 Arth 2011 Ventura 2012
Model-based method	Restore the current image viewpoint by register to the existing 2D/3D model	Restoring the current pose through the sky silhouettes registration	Baatz 2012 Bansal 2014
		Based on the building outline edges to estimate the camera pose	Chu 2014 Cham 2010 David 2011
		Using the scene semantic segmentation information to improve the pose accuracy	Arth 2015 Arth 2015 Baatz 2012 Taneja 2012
Combination of Vision-based methods	Image-based localization + SLAM	Combine the 3D point cloud data with SLAM	Ventura 2014 Sattler 2014
	Model-based localization + SLAM	Integrate existing geographic information into SLAM	Arth 2015

(1)　Image-Based Method

In the early years, the image-based methods transformed the localization problem into the image retrieval problem [16,17]. The image database with GPS position marks was constructed in advance, and the most similar image was found by matching between the current image and the database. So, the found image's position was the solution of the localization problem [18–22]. This method could only roughly restore the current position (3DOF), but not the fully accurate 6-DoF pose.

In contrast, another image-based method [23,24] employed offline SFM to reconstruct the three-dimensional position of the feature points in the image sequence with exact global pose marks. In the stage of localization, the 6DOF pose of the camera was calculated via the standard three-point perspective (P3P) algorithm [25,26], by matching the feature points extracted from the current image to the 3D point cloud reconstructed by the SFM.

The current methods based on SFM pre-reconstruction can obtain a relatively high-precision global 6DOF pose [23,27–30] comparing to the image retrieval methods, but the difficulty is to perform large-scale real-time localization on mobile devices. Some researchers [31,32] proposed a real-time pose tracking method on the mobile devices, but it was only applied to small scenes, not good enough for a long time 6-DOF localization.

The image-based global localization method has made some progress in recent years, but there are still some problems while employed in the UWAR applications.

It heavily relies on pre-acquired images registered in the global coordinate system, and the procedure of generating the database consumes a lot of time and computing resource.

The real-time pose tracking does not work beyond the scope of off-line preparation, since the image database is created offline.

The image database would become obsolete when any of the geometry, appearance or lighting changed.

(2)　Model-Based Method

The model-based method is to use the existing geographic information for image registration. Currently, model-based location methods generally are classified into the following categories:

(a)　Restoring the pose via the registration of the sky silhouettes. Baatz et al. tried to align the contour lines extracted from the DEM (digital elevation model) and the silhouettes of the mountain and sky in the input image, to restore the current pose of view [33].

(b)　Based on the building outline to estimate the camera pose. Chu [34] calculated the intersection of two or more vertical contours of the building and the ground, and then calibrated the rough GPS position using an accurate 2D building map. Similar work was also presented by Cham [35] and David [36].

(c)　Using the scene semantic segmentation information to improve the accuracy of pose. Arth et al. [37] combined the building edges and the semantic segmentation information (such as building surface, roof, vegetation, ground, sky, etc.), and then aligned them with the 2.5D building model to restore the current global pose of the camera. Compared with the work of Chu [34], Arth registered the semantic segmentation of the input image with the existing map model to increase the accuracy in the pose estimation. The orientation error was less than 5° and the majority (87.5%) position error was less than 4 meters. Some similar research was also done by Baatz et al. [38] and Taneja et al. [39].

Compared with image-based localization methods, the model-based method takes advantage of the line and surface features to perform feature matching. On the one hand, the model-based approach is more robust to environmental light changing since it does not depend on the texture feature points; On the other hand, the model of the building sparse area does not contain enough information for localization.

Similar to the image-based method, the model-based method needs pre-prepared dataset, and the localization is limited to areas where the required information has been collected; However, the geographic information available for public access is more and more abundant and the coverage area is wider, and the model-based tracking would gain an advantage in the implementation of UWAR registration.

(3)　Combination of Vision-Based Methods

There are two representative combinations of visual localization methods.

Combination I: Image-based localization + SLAM. The basic idea of this method is to combine the 3D point cloud data reconstructed by the image sequence with SLAM to expand the range of global pose tracking. The current representative researches are finished by Middelberg [40] and Ventura [15].

Combination II: Model-based localization + SLAM. This solution was proposed by Arth [37] in 2015 and the existing geographic information was integrated into the SLAM to extend the SLAM's ability of tracking the position and orientation in the environment without prior preparation.

Combining SLAM with the prepared image data or geographic information is a trend in the study of the vision-based global localization methods, and it makes the vision-based methods more adaptable and robust in partially prepared areas or imperfectly pre-prepared environment. However, the disadvantage of the vision-based localization methods is that high quality images have to be acquired under good visual conditions but the objective conditions of the battlefield cannot guarantee this.

3.2.4. Hybrid Registration Technology

The hybrid registration method was proposed by Azuma in 1998 and has become an important research direction in recent years. In 2012, a registration method by combining inertial units, GPS and visual data was proposed by Oskiper using the extended Kalman filter, to construct a hybrid registration framework on a mobile phone [41]. Hartmann used the unscented Kalman filter to carry out the fusion of visual and inertial sensor data on a mobile platform in 2013 [42]. During the development of ARC4, Menozzi A et al. proposed a hybrid pose tracking method which combined visual methods based on landmarks, DEM terrain contours and the sun location with INS-GPS sensor components in 2014, and the method could calculate the global 6DOF pose with high precision in real time [43].

Kendall trained a convolutional neural network to regress the 6-DOF camera pose from a single RGB image, but the algorithm was not accuracy enough in outdoor for augmented reality system [44]. Rambach proposed a deep learning approach to visual-inertial camera pose estimation and the approach was able to integrate inertial measurements in the tracking framework, without complex modeling of the sensors noises, biases and non-linearity and the calibration between camera and sensor coordinate [45].

At present, the common hybrid registration models include vision-inertial registration, inertial-ultrasonic registration, inertial-compass registration, visual-GPS registration, compass-GPS registration etc. and the trend is to fuse more than three sensors for registration. In the aspect of the fusing method for hybrid registration, there are kalman filter, extended kalman filter, unscented kalman filter, partial filters etc.

3.2.5. Application of Registration Technology in UWAR

How to quickly and accurately calculate the head pose of commanders on the move is still a challenging problem. In urban warfare, the commanders' heads adjust frequently during a large scale operation, making it more difficult to obtain high-precision head pose in real time. Especially when the positioning signal is shielded or interfered in an unfamiliar environment, how to use a variety of methods to obtain the reliable pose information remains unsolved. In order to improve the accuracy and robustness of registration, the fusion of multi-modal registration method is a promising solution. Furthermore, fusion algorithm can become very complex as the number of sensors increases more than three, and deep learning technology could be employed to solve this problem in the future.

3.3. Information Representation of UWAR

3.3.1. Main Issues of Information Representation

In UWAR system, the information that needs to be displayed mainly includes: battlefield situation information, battlefield environment information, navigation information, combat information and other auxiliary information (see Table 2).

Table 2. The information needs to be displayed in UWAR system.

Information Type	Information Content
Battlefield situation information	Enemy's situation, the situation of friendly force etc.
Battlefield environment information	Geographic information, landmark information, electromagnetic information etc.
Navigation information	Navigation route, azimuth etc.
Combat information	Threaten area, command etc.
Other auxiliary information	Time, coordinate, system information etc.

Under conditions of equivalent information and scene, efficiency of commanders' understanding of situational information would vary a lot by using different information presentation methods.

Therefore, it is crucial to study how to represent battlefield situational information better and thus improve efficiency of commanders' understanding of battlefield situational information. Information representation of UWAR serves this purpose, which studies how to make soldiers to understand situational information quickly and accurately without interfering them observing environment [46,47].

Urban environment consists of dense buildings, complex spatial structure, dynamic lighting conditions, which raises lots of problems for battlefield situational information representation. Key problems are elaborated as follows: information overload display, information occlusion, view layout.

3.3.2. Information Overload

Information overload occurs when the number of information displayed in front of the user exceeds his ability to adapt, which often leads to negative effects such as mental strain and visual fatigue and thus affecting decision-making efficiency [48]. Especially in urban warfare, complex urban environment often results in huge amount of situational information which greatly affects the efficiency of commanders observing the environment and obtaining the information of targets. At present, there are two main types of ways to avoid information overload, namely information filtering and information clustering.

The representation method based on information filtering is to delete the information that is not relevant to users from display interface. An ideal method of information filtering is to reduce the amount of information to an acceptable level without losing the necessary information to users. Based on the characteristics of battlefield augmented reality, Livingston proposed an algorithm of information filtering based on user correlation [49].

The representation method based on information clustering is to aggregate attribute-related information and display them. Its purpose is to reduce the amount of information needs to be displayed. When users pay attention to a certain kind of information, the method would disaggregate the information and display them. Tatzgern et al. have done a lot of research in this field, and they proposed different representation methods based on information clustering [50–52]. Tatzgern et al. proposed a method of adaptive information density display for mobile augmented reality. This method could balance the amount of presented information against the potential clutter created by placing items on the screen, which used hierarchical clustering to create a level-of-detail structure. In another study, they evaluated effects of clustered annotation density on search and recall tasks [53].

3.3.3. Occlusion Handling

Occlusion handling method is to determine whether objects that need to be highlighted with information are blocked by other objects, which is a prevalent issue in urban warfare. For example, the problem occurs when the commanders need to perceive the situational information inside a building.

Furmanski et al. proposed a conception of information representation in augmented reality, Obscured Information Visualization (OIV). They also presented two design guidelines for OIV. The first guideline is that there is a way to represent the differences between normally perceptible and extra-sensory. The second is that visualization in a cluttered and complex environment should be algorithmically and perceptually easy to implement [54]. On the basis of previous research findings, Livingston et al. proposed a method that used metaphors to depict occluded objects [49].

3.3.4. View Management

In UWAR systems, disorderly information arrangement can affect the efficiency of the commanders' battlefield situational awareness. It is a problem that how to adjust the location of information adaptively and avoid the sight of the soldiers blocked by the information. Grasset et al. introduce an image-based approach, which combines a visual saliency algorithm with edge analysis to identify potentially important image regions and geometric constraints for placing labels [55]. Different information display layouts will have great impacts on the efficiency of user's environmental perception in urban environment. Tatzgern et al. proposed an information representation method

based on 3D geometric constraints. The method could effectively avoid the label overlapping gland display [56].

3.3.5. Application of Information Representation Technology in UWAR

Since soldiers' mental states become extremely intense in urban warfare, it's critical to display battlefield information adaptively according to soldiers' current tasks, state, and scenario, which is also the key problem needs to be broken through in the development of UWAR.

At present, information representation based on scene understanding is an important direction for future development. At present, information representation in augmented reality system is mainly based on geometric information of the scene, and does not take into account semantic information of the scene. In future development, information representation should be able to understand the semantic structure of scene and thus making UWAR systems more intelligent, which knows where to display information.

3.4. UWAR Interaction Technology

Collaboration and communication between the troop members are essential tasks in operations, which require UWAR system to explore the use of techniques such as human-computer interaction [46]. In urban warfare, soldiers are fully concentrated in combat operations in which they need to be well aware of the situation and make decision in time. For these extreme cases, interaction technologies should meet the requirements of intuitive interface, natural interaction process, real-time response and fault-tolerant.

Interaction methods employed by traditional AR applications include mouse and keyboard, touch screen, gesture, voice, gaze, etc. The traditional user interactions consist of the direct manipulation of graphic objects, such as icons and windows, using some type of pointing devices. Mice and keyboards are widely used and well-known for these applications, and recently, users can interact more intuitively using touch-screens, which is also very mature, especially for mobile devices. However, the main limitation for these techniques is that the user has to go and grab the device or touch it, which often cause interruptions to the on-going tasks on hand. This becomes a serve problem to the soldiers in urban warfare.

Gestures have long been considered as an interaction technique that can potentially deliver more natural, creative and intuitive methods for communicating with our computing devices [57]. Recognizing gestures for interaction can help achieve the ease and naturalness desired for human computer interaction. Users generally use different gestures for expressing their feelings, communicating and notification of their thoughts. The use of gestures provides an attractive and natural alternative to some traditional cumbersome devices for human computer interaction. The gesture can be the posture of a hand, an arm or a tangible device, which is hold by the user. In addition, the pose change and the moving track of fingers, hands, arms and tangible devices also belong to the category of gestures. Nowadays, the gesture interaction technology is mainly implemented by understanding soldiers' hand gesture, since hand gesture command is a natural interaction technique commonly used in both civil AR applications and traditional military operations. Data gloves and video sensors are the common devices employed in gesture detection. Argenta et al. designed an operational gesture interaction system, which could recognize more than 10 standard military gestures using a data glove [47]. Numerous methods for vision-based hand gesture recognition have been developed and evaluated. Since the release of commercial depth sensors, there have been numerous inspiring successes in finger tracking and hand gesture recognition for human computer interaction [58,59]. Zocco et al. [60] performed a user study, in which officers and soldiers can issue orders to enhance situational awareness via specific gestures using leap motion as a touchless natural user input device, and got a positive evaluation of the interactive method. Hand gesture is always a hot topic of natural interaction and a mature technology for general AR system. However, it suffers a lot of constraints in the UWAR systems since soldiers' hands are occupied in most tasks.

Voice commands would be one option, but noisy battlefield environment imposes a huge challenge for the off-the-shelf speech recognition technologies. Eye gaze is another natural input method and the soldier would need to deviate as little as possible from the normal routine. However, soldiers have to keep gaze moving to search the targets and hide from the threats, so it needs to work with other technologies, e.g., Brain-Computer interface.

In a vehicle, the soldiers are not directly exposed to the threat and do not always have to handheld weapons, and traditional hand-held input devices such as keys, joysticks or touch screens have certain feasibility. These technologies are mature, stable and low-cost, and users have been familiar with them. But for an unmounted soldier, it is a total different situation. The traditional interaction methods are not natural enough to react at a moment's notice. In order to enhance the user's immersion, AR applications tend to choose a more natural way of interaction, such as gesture and speech. In the existing UWAR systems, there are few instances of direct interactions with the virtual information and the interactions between users and virtual information are mostly passive reception. UWAR system can also take advantage of auditory and tactile methods to make up for the lack of visual perception. Colbert et al. [4] tried to feed GPS positioning information to soldiers in three different ways (visual, auditory and tactile methods) and the experimental results showed that the visual display is most popular. But in the case of visual overload, the tactile method is much easier to be accepted. Hence, for direct interaction with virtual information, the multi-modal interaction could be one solution as it integrates multiple interaction methods to mitigate limitations when hands are occupied. In conclusion, performing interaction with UWAR systems cannot disturb the soldier and should improve the soldier's ability to carry out their tasks without weighing heavily on their mind.

4. Future Development

According to VisonGain's forecast, the market size of military augmented reality will continue to grow in the next ten years (see Figure 3) [61]. Military augmented reality technology will get more and more attention in most countries. At present, the technology of urban warfare augmented reality is still in the stage of laboratory research and equipment testing and yet to be put into practical use. The development of UWAR system can also benefit from the fast advancing fields of GIS and virtual geographic environment (VGE). For GIS, the effort to collect high precision data of urban geography provides both references for registration in UWAR system and content-of-interest that needs to be represented in the system. For VGE, its research findings can be directly applied in UWAR system, such as the representation of urban building structural information, and the representation of obscured information in urban buildings.
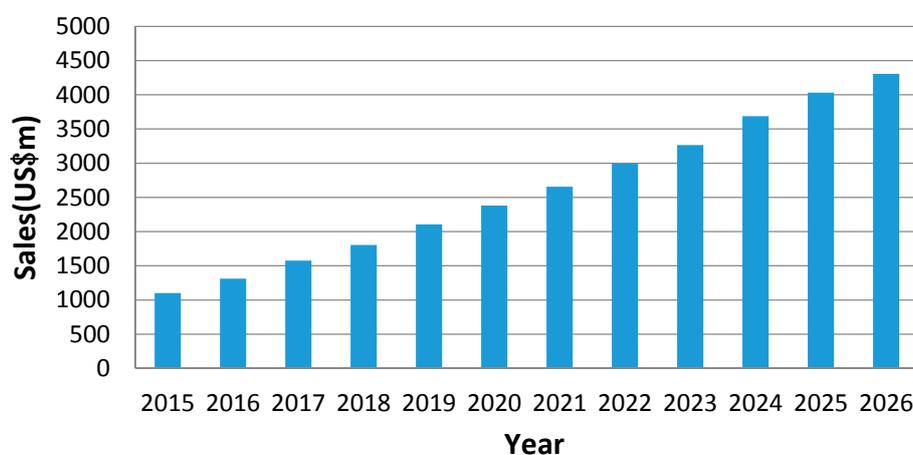

Figure 3. Global Military Augmented Reality Technologies Market Forecast 2016–2026.

In conclusion, the future development of the UWAR technology will be driven by following aspects:

(1) More powerful hardware. UWAR system will be more suitable for soldiers to use in battlefield with the development of the enabling technologies, such as more powerful portable computing device, see-through display with higher specification (e.g., brightness, resolution, and perspective), longer endurance, and more natural human-computer interaction. Therefore, the envisioned features of UWAR system can be realized and fully meet the needs of soldiers in urban warfare.

(2) More intelligent Software. With the development of artificial intelligence, it's able for the system to understand the intentions of soldiers and achieve high degree human-machine collaboration. What's more, the system can understand the geometric structure of the battlefield environment as well as its semantic structure, and thus to facilitate information filtering for adaptive display in terms of where (on the display), what (content), when. UWAR system will evolve into an indispensable combat assistant for soldiers in urban warfare.

(3) Better integration with GIS and VGE. GIS and VGE often serve as the spatial data infrastructure of UWAR, which play an important role in real-time registration and information representation of UWAR. However, seldom research work has been devoted to integration of real-world model developed in GIS and VGE into UWAR, which should be paid more attention to in order to put the UWAR system into practical use.

Author Contributions: Xiong You designed the framework of the paper. The introduction, Overview of UWAR and Future Development sections were developed jointly by Xiong You, Weiwei Zhang, Meng Ma and Jian Yang. Key Technologies section was developed jointly by Xiong You, Weiwei Zhang, Meng Ma and Chen Deng. Jian Yang, Meng Ma and Weiwei Zhang revised the manuscript draft.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Defense Advanced Research Projects Agency of USA (U.S. Department of Defense, State of Washington, USA). *Urban Leader Tactical Response, Awareness & Visualization (ULTRA-Vis)*; Defense Advanced Research Projects Agency of USA: Arlington, VA, USA, 2008.

2. Roberts, D.; Menozzi, A.; Clipp, B.; Menozzi, A.; Clipp, B.; Russler, P. Soldier-worn augmented reality system for tactical icon visualization. Head- and Helmet-Mounted Displays XVII; and Display Technologies and Applications for Defense. *SPIE* **2012**, *8383*. [CrossRef]

3. Livingston, M.A.; Swan, J.E.; Simon, J.J. Evaluating system capabilities and user performance in the battle eld augmented reality system. In Proceedings of the NIST/DARPA Workshop on Performance Metrics for Intelligent Systems, Gaithersburg, MD, USA, 24–26 August 2004.

4. Colbert, H.J.; Tack, D.W.; Bossi, L.C.L. *Augmented Reality for Battlefield Awareness*; DRDC Toronto CR-2005-053; HumanSystems® Incorporated: Guelph, ON, Canada, 2005.

5. Le Roux, W. The use of augmented reality in command and control situation awareness. *Sci. Mil. S. Afr. J. Mil. Stud.* **2010**, *38*, 115–133. [CrossRef]

6. Zysk, T.; Luce, J.; Cunningham, J. Augmented Reality for Precision Navigation-Enhancing performance in High-Stress Operations. *GPS World* **2012**, *23*, 47.

7. Kenny, R.J. *Augmented Reality at the Tactical and Operational Levels of War*; Naval War College Newport United States: Newport, RI, USA, 2015.

8. Julier, S.; Baillot, Y.; Lanzagorta, M. BARS: Battlefield Augmented Reality System. Nato Symposium on Information Processing Techniques for Military Systems. In Proceedings of the NATO Symposium on Information Processing Techniques for Military Systems, Istanbul, Turkey, 9–11 October 2000; pp. 9–11.

9. Gans, E.; Roberts, D.; Bennett, M. Augmented reality technology for day/night situational awareness for the dismounted Soldier. *Proc. SPIE* **2015**, *9470*, 947004:1–947004:11.

10. Heads Up: Augmented Reality Prepares for the Battlefield. Available online: https://arstechnica.com/information-technology/ (accessed on 1 January 2018).

11. Heads-Up Display to Give Soldiers Improved Situational Awareness. Available online: https://www.army.mil/ (accessed on 1 January 2018).

12. Feiner, S.; Macintyre, B.; Hollerer, T. A touring machine: Prototyping 3D mobile augmented reality systems for exploring the urban environment. *Pers. Technol.* **1997**, *1*, 208–217. [CrossRef]

13. Menozzi, A.; Clipp, B.; Wenger, E. Development of vision-aided navigation for a wearable outdoor augmented reality system. In Proceedings of the 2014 IEEE/ION Position, Location and Navigation Symposium—PLANS 2014, Monterey, CA, USA, 5–8 May 2014; pp. 460–472.

14. Maus, S. An ellipsoidal harmonic representation of Earth's lithospheric magnetic field to degree and order 720. *Geochem. Geophys. Geosyst.* **2010**, *11*, 1–12. [CrossRef]

15. Ventura, J.; Arth, C.; Reitmayr, G. Global localization from monocular slam on a mobile phone. *IEEE Trans. Visual. Comput. Graph.* **2014**, *20*, 531–539. [CrossRef] [PubMed]

16. Hays, J.; Efros, A.A. IM2GPS: Estimating geographic information from a single image. In Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, AK, USA, 23–28 June 2008; pp. 1–8.

17. Kalogerakis, E.; Vesselova, O.; Hays, J. Image sequence geolocation with human travel priors. In Proceedings of the IEEE 12th International Conference on Computer Vision, Tokyo, Japan, 29 September–2 October 2009; pp. 253–260.

18. Schindler, G.; Brown, M.; Szeliski, R. City-scale location recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Minneapolis, MN, USA, 17–22 June 2007; pp. 1–7.

19. Zhang, W.; Kosecka, J. Image based localization in urban environments. In Proceedings of the IEEE Third International Symposium on 3D Data Processing, Visualization, and Transmission, Chapel Hill, NC, USA, 4–16 June 2006; pp. 33–40.

20. Robertson, D.P.; Cipolla, R. An Image-Based System for Urban Navigation. In Proceedings of the 15th British Machine Vision Conference (BMVC'04), Kingston-upon-Thames, UK, 7–9 September 2004; pp. 819–828.

21. Zamir, A.R.; Shah, M. Accurate image localization based on google maps street view. In Proceedings of the 11th European Conference on Computer Vision, Crete, Greece, 5–11 September 2010; pp. 255–268.

22. Vaca, C.G.; Zamir, A.R.; Shah, M. City scale geo-spatial trajectory estimation of a moving camera. In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012; pp. 1186–1193.

23. Irschara, A.; Zach, C.; Frahm, J.M. From structure-from-motion point clouds to fast location recognition. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 2599–2606.

24. Klopschitz, M.; Irschara, A.; Reitmayr, G. Robust incremental structure from motion. In Proceedings of the Fifth International Symposium on 3D Data Processing, Visualization and Transmission (3DPVT), Paris, France, 17–20 May 2010; Volume 2, pp. 1–8.

25. Haralick, B.M.; Lee, C.N.; Ottenberg, K. Review and analysis of solutions of the three point perspective pose estimation problem. *Int. J. Comput. Vis.* **1994**, *13*, 331–356. [CrossRef]

26. Haralick, R.M.; Lee, D.; Ottenburg, K. Analysis and solutions of the three point perspective pose estimation problem. In Proceedings of the 1991 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Lahaina, Maui, HI, USA, 3–6 June 1991; pp. 592–598.

27. Li, Y.; Snavely, N.; Huttenlocher, D.P. Location recognition using prioritized feature matching. In Proceedings of the 11th European Conference on Computer Vision, Crete, Greece, 5–11 September 2010; pp. 791–804.

28. Li, Y.; Snavely, N.; Dan, H. Worldwide Pose Estimation Using 3D Point Clouds. In Proceedings of the 12th European Conference on Computer Vision, Florence, Italy, 7–13 October 2012; pp. 15–29.

29. Lim, H.; Sinha, S.N.; Cohen, M.F. Real-time image-based 6-dof localization in large-scale environments. In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012; pp. 1043–1050.

30. Lim, H.; Sinha, S.N.; Cohen, M.F. Real-time monocular image-based 6-DoF localization. *Int. J. Robot. Res.* **2015**, *34*, 476–492. [CrossRef]

31. Wagner, D.; Reitmayr, G.; Mulloni, A. Real-time detection and tracking for augmented reality on mobile phones. *IEEE Trans. Visual. Comput. Graph.* **2010**, *16*, 355–368. [CrossRef] [PubMed]

32. Arth, C.; Wagner, D.; Klopschitz, M. Wide area localization on mobile phones. In Proceedings of the 8th IEEE International Symposium on ISMAR 2009, Orlando, FL, USA, 19–22 October 2009; pp. 73–82.

33. Baatz, G.; Saurer, O.; Köser, K. Large Scale Visual Geo-Localization of Images in Mountainous Terrain. In Proceedings of the 12th European Conference on Computer Vision, Florence, Italy, 7–13 October 2012; pp. 517–530.

34. Chu, H.; Gallagher, A.; Chen, T. GPS refinement and camera orientation estimation from a single image and a 2D map. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 171–178.

35. Cham, T.J.; Ciptadi, A.; Tan, W.C. Estimating camera pose from a single urban ground-view omnidirectional image and a 2D building outline map. In Proceedings of the Twenty-Third IEEE Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 13–18 June 2010; pp. 366–373.

36. David, P.; Ho, S. Orientation descriptors for localization in urban environments. In Proceedings of the 2011 IEEE/RSJ International Conference on Intelligent Robots and Systems, San Francisco, CA, USA, 25–30 September 2011; pp. 494–501.

37. Arth, C.; Pirchheim, C.; Ventura, J. Global 6DOF Pose Estimation from Untextured 2D City Models. *Comput. Sci.* **2015**, *25*, 1–8.

38. Baatz, G.; Saurer, O.; Köser, K. Leveraging Topographic Maps for Image to Terrain Alignment. In Proceedings of the 2012 Second International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission (3DIMPVT), Zurich, Switzerland, 13–15 October 2012; pp. 487–492.

39. Taneja, A.; Ballan, L.; Pollefeys, M. Registration of spherical panoramic images with cadastral 3d models. In Proceedings of the 2012 Second International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission (3DIMPVT), Zurich, Switzerland, 13–15 October 2012; pp. 479–486.

40. Middelberg, S.; Sattler, T.; Untzelmann, O. Scalable 6-dof localization on mobile devices. In Proceedings of the 13th European Conference on Computer Vision, Zurich, Switzerland, 6–12 September 2014; pp. 268–283.

41. Oskiper, T.; Samarasekera, S.; Kumar, R. Multi-sensor navigation algorithm using monocular camera, IMU and GPS for large scale augmented reality. In Proceedings of the IEEE International Symposium on Mixed and Augmented Reality (ISMAR2012), Atlanta, GA, USA, 5–8 November 2012.

42. Hartmann, G.; Huang, F.; Klette, R. Landmark initialization for unscented kalman filter sensor fusion in monocular camera localization. *Int. J. Fuzzy Logic Intell. Syst.* **2013**, *13*, 1–11. [CrossRef]

43. Billinghurst, M.; Clark, A.; Lee, G. A survey of augmented reality. *Found. Trends Hum.–Comput. Interact.* **2015**, *8*, 73–272. [CrossRef]

44. Kendall, A.; Grimes, M.; Cipolla, R. Posenet: A convolutional network for real-time 6-dof camera relocalization. In Proceedings of the IEEE International Conference on Computer Vision, Shenzhen, China, 7–13 December 2015.

45. Rambach, J.; Tewari, A.; Pagani, A.; Stricker, A. Learning to Fuse: A Deep Learning Approach to Visual-Inertial Camera Pose Estimation. In Proceedings of the IEEE International Symposium on Mixed and Augmented Reality (ISMAR2016), Merida, Mexico, 19–21 September 2016.

46. Livingston, M.A.; Rosenblum, L.J.; Brown, D.G.; Schmidt, G.S.; Julier, S.J. Military applications of augmented reality. In *Handbook of Augmented Reality*; Springer: New York, NY, USA, 2011; pp. 671–706.

47. Argenta, C.; Murphy, A.; Hinton, J.; Cook, J.; Sherrill, T.; Snarski, S. Graphical User Interface Concepts for Tactical Augmented Reality. *Proc. SPIE* **2010**, *7688*. [CrossRef]

48. Tatzgern, M.; Kalkofen, D.; Grasset, R.; Schmalstieg, D. Hedgehog Labeling: View Management Techniques for External Labels in 3D Space. *Virtual Real.* **2014**. [CrossRef]

49. Grasset, R.; Langlotz, T.; Kalkofen, D. Image-driven view management for augmented reality browsers. In Proceedings of the 11th IEEE International Symposium on ISMAR, Atlanta, GA, USA, 5–8 November 2012; pp. 177–186.

50. Karlsson, M. *Challenges of Designing Augmented Reality for Military Use*; Karlstad University: Karlstad, Sweden, 2015.

51. Livingston, M.A.; Swan, J.; Gabbard, J.; Hollerer, T.H.; Hix, D.; Julier, S.J.; Baillot, Y.; Brown, D. Resolving multiple occluded layers in augmented reality. In Proceedings of the IEEE International Symposium on Mixed and Augmented Reality (ISMAR2003), Tokyo, Japan, 7–10 October 2003.

52. Livingston, M.A.; Ai, Z.M.; Karsch, K.; Gibson, G.O. User interface design for military AR applications. *Virtual Real.* **2011**, *15*, 175–184. [CrossRef]

53. Tatzgern, M.; Kalkofen, D.; Schmalstieg, D. Multi-perspective compact explosion diagrams. *Comput. Graph.* **2011**, *35*, 135–147. [CrossRef]

54. Tatzgern, M.; Kalkofen, D.; Schmalstieg, D. Dynamic compact visualizations for augmented reality. *IEEE Virtual Real.* **2013**, 3–6. [CrossRef]

55. Tatzgern, M.; Kalkofen, D.; Schmalstieg, D. Compact explosion diagrams. *Comput. Graph.* **2010**, *35*, 135–147. [CrossRef]

56. Tatzgern, M.; Orso, V.; Kalkofen, D. Adaptive information density for augmented reality displays. In Proceedings of the IEEE VR 2016 Conference, Greenville, SC, USA, 19–23 March 2016; pp. 83–92.

57. Furmanski, C.; Azuma, R.; Daily, M. Augmented-reality visualizations guided by cognition: Perceptual heuristics for combining visible and obscured information. In Proceedings of the IEEE International Symposium on Mixed and Augmented Reality (ISMAR2002), Darmstadt, Germany, 30 September–1 October 2002.

58. Zhou, R.; Jingjing, M.; Junsong, Y. Depth camera based hand gesture recognition and its applications in Human-Computer-Interaction. In Proceedings of the 2011 8th International Conference Information, Communication Signal Process, Singapore, 13–16 December 2011; Volume 5. [CrossRef]

59. Kulshreshth, A.; Zorn, C.; Laviola, J.J. Poster: Real-time markerless kinect based finger tracking and hand gesture recognition for HCI. In Proceedings of the 2013 IEEE Symposium on 3D User Interfaces (3DUI), Orlando, FL, USA, 16–17 March 2013; pp. 187–188.

60. Zocco, A.; Zocco, M.; Greco, A.; Livatino, S. Touchless Interaction for Command and Control in Military Operations. In Proceedings of the International Conference on Augmented and Virtual Reality AVR 2015: Augmented and Virtual Reality, Lecce, Italy, 31 August–3 September 2015; pp. 432–445.

61. Visiongain. *Military Augmented Reality (MAR) Technologies Market Report 2016–2026*; Business Intelligence Company: London, UK, 2016.