Taylor & Francis
Taylor & Francis Group

International Journal of Production Research

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tprs20

Artificial intelligence (AI) in augmented reality (AR)-assisted manufacturing applications: a review

Chandan K. Sahu , Crystal Young & Rahul Rai

To cite this article: Chandan K. Sahu , Crystal Young & Rahul Rai (2020): Artificial intelligence (AI) in augmented reality (AR)-assisted manufacturing applications: a review, International Journal of Production Research, DOI: 10.1080/00207543.2020.1859636

To link to this article: https://doi.org/10.1080/00207543.2020.1859636

Published online: 28 Dec 2020.

Submit your article to this journal

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=tprs20

INTERNATIONAL JOURNAL OF PRODUCTION RESEARCH
https://doi.org/10.1080/00207543.2020.1859636

Taylor & Francis
Taylor & Francis Group

Check for updates

Artificial intelligence (AI) in augmented reality (AR)-assisted manufacturing applications: a review

Chandan K. Sahu [a*], Crystal Young [b*] and Rahul Rai [a]

[a]Geometric Reasoning and Artificial Intelligence Lab (GRAIL), Department of Automotive Engineering, Clemson University, Greenville, SC, USA; [b]Manufacturing and Design Lab (MADLab), Department of Mechanical and Aerospace Engineering, University at Buffalo, Buffalo, NY, USA

ABSTRACT

Augmented reality (AR) has proven to be an invaluable interactive medium to reduce cognitive load by bridging the gap between the task-at-hand and relevant information by displaying information without disturbing the user's focus. AR is particularly useful in the manufacturing environment where a diverse set of tasks such as assembly and maintenance must be performed in the most cost-effective and efficient manner possible. While AR systems have seen immense research innovation in recent years, the current strategies utilised in AR for camera calibration, detection, tracking, camera position and orientation (pose) estimation, inverse rendering, procedure storage, virtual object creation, registration, and rendering are still mostly dominated by traditional non-AI approaches. This restricts their practicability to controlled environments with limited variations in the scene. Classical AR methods can be greatly improved through the incorporation of various AI strategies like deep learning, ontology, and expert systems for adapting to broader scene variations and user preferences. This research work provides a review of current AR strategies, critical appraisal for these strategies, and potential AI solutions for every component of the computational pipeline of AR systems. Given the review of current work in both fields, future research work directions are also outlined.

ARTICLE HISTORY
Received 16 June 2020
Accepted 14 November 2020

KEYWORDS
Augmented reality; artificial intelligence; deep learning; machine learning; tracking; manufacturing

1. Introduction

Although reality can never be altered, its perception can be improved. Augmented reality (AR) is an interface between reality and the perception of reality. AR is the medium that superimposes digital information on the physical world, thereby bridging the rift between the real and virtual worlds. AR enables a novel transformational paradigm of information delivery and assimilation to a human operator, and if properly implemented, leads to cognitive load[1] reduction.

AR can be defined as a human-computer interaction technology that interactively registers three-dimensional (3D) real and virtual objects in real time (Azuma et al. 2001). Although the term 'Augmented Reality' was coined by Thomas and David in 1992, the first application of AR technology was a head-mounted display (HMD) which dates back to 1968 by Sutherland. AR is in the process of revolutionising many industries including education (Radu 2012; Wu et al. 2013), construction and building management (Chi, Kang, and Wang 2013; Gheisari et al. 2014), e-commerce (Dacko 2017), robotics (Green et al. 2008; Makris et al. 2016), automotive (Regenbrecht, Baratoff, and Wilke 2005; Alhaija et al. 2017), defense (Livingston et al. 2011), healthcare (Fischer, Bartz, and Straßer 2004; Yamamoto et al. 2012), tourism (Wei, Ren, and O'Neill 2014; tom Dieck and Jung 2018), gaming (Piekarski and Thomas 2002), and product lifecycle management (PLM) (Doil et al. 2003; Ong, Yuan, and Nee 2008).

AR has the opportunity to be integrated into every stage of PLM. For instance, AR can be utilised in the facility development process by visualising the floorplan in an interactive environment for the developer (Wang et al. 2020). In product development, AR can simulate the various stages of manufacturing for a product or show the various design iterations of a product (Ong, Yuan, and Nee 2008). While the product is being manufactured, AR can be utilised for assembly (Wang et al. 2013), quality assurance (Schwerdtfeger et al. 2008), and for collaboration and safety (Makris et al. 2016). Once the product is ready to be marketed to consumers, AR can be used by the consumer to visualise the product as an integrated

CONTACT Rahul Rai rrai@clemson.edu, rahulrai@buffalo.edu Geometric Reasoning and Artificial Intelligence Lab (GRAIL), Department of Automotive Engineering, Clemson University, Greenville, SC 29607, USA
*This work has been contributed equally.

© 2020 Informa UK Limited, trading as Taylor & Francis Group

part of their daily life (Scholz and Smith 2016). When the user buys the product, AR can be used as a diagnostic tool in case of troubleshooting (Sanna et al. 2015). Finally, at the end of the product's lifecycle, AR can be utilised again in the manufacturing environment in order to disassemble the final product for material recovery (Osti et al. 2017; Chang, Nee, and Ong 2020).

AR can be an effective way to decrease costs and increase efficiency during the manufacturing stage of PLM. AR is a technology that can reduce cognitive load by supplementing human labour with the right information at the right moment and in the right mode without disturbing the user's focus (Abraham and Annunziata 2017), thereby enabling the industry to achieve the aforementioned goals. The focus of this paper is to present a review of the state-of-the-art literature at the intersection of AR, manufacturing, and artificial intelligence (AI).

Many manufacturing environments are noisy, and thus it is impractical to include audio or voice elements into AR-assisted manufacturing applications without consideration of alternative user interaction or rigorous noise reduction strategies (Syberfeldt, Danielsson, and Gustavsson 2017). Additionally, mandatory personal protective equipment limits the user from other modalities like haptics. For this reason, this review primarily focuses on the visual modality. To be specific, we focus on Red Green Blue (RGB) images or a live video stream (rather than RGB Depth (RGB-D), thermal, Light Detection and Ranging (LiDAR), audio or other sensory data) as the input data due to the fact that even miniature ergonomic devices like mobiles and wearable devices can equip both a camera to acquire the data and a display to disseminate the augmented information.

The computational pipeline for an AR system is shown in Figure 1. For a seamless and realistic integration of the real and virtual worlds in the AR device, the real and virtual environments must align with each other. The virtual camera must be *calibrated* with the real camera's intrinsic parameters to ensure that the real world is accurately depicted throughout the AR computational pipeline. Each frame from the real camera is then inputted into the detection, tracking, and camera position and orientation (pose) estimation modules in order to acquire the extrinsic parameters. The *detection* module detects the pertinent dynamic objects within the use environment. The *tracking* module then tracks these objects in real-time. Finally, the *camera pose estimation* module estimates the pose of the real camera in real-time. Additionally, each frame is *rendered inversely* for determining characteristics like illumination and depth to supplement the graphics processing unit to enable a realistic rendition of the virtual objects in the scene. Once

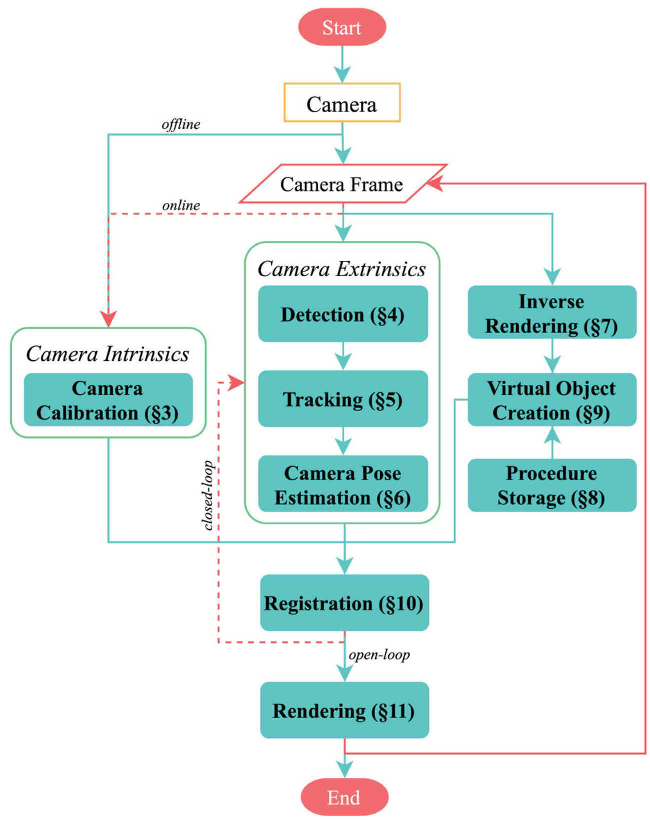


Figure 1. Computational pipeline for a procedure-based AR system.

the functionalities of the virtual camera imitating the real camera are achieved, any information within *procedure storage* is incorporated into the creation of virtual objects for AR in *virtual object creation*. These virtual objects are *registered* in the AR environment, which means that they are aligned with the detected and tracked objects in the operator's real environment. Finally, virtual objects are *rendered* on the main visual interface of the AR system. The rendering process places the virtual objects in an optimal way for the user and the current environment.

The pipeline underlying AR-based manufacturing applications suffer from the following drawbacks: (1) the computational strategies are computationally expensive and not robust in varying environments, (2) the current strategies utilised in AR systems are laborious to implement and restricted to known environments (*i.e.* camera calibration), (3) specialised and diverse feature construction is required for object detection and tracking, (4) there is a lack of robust pose estimation techniques, thereby affecting variable runtime and latency of the AR system, (5) storage methods are rigid and non-adaptable, (6) virtual objects are hard-coded within the AR system, (7) primarily open-loop registration approaches are utilised, and (8) current rendering methods implement simplified rules to promote context-awareness in the AR system and utilise destructive methods to optimise the visualisation of data. Clearly, improvements are needed in all aspects of procedure-based AR systems.

AI is a promising solution to increase the adaptability and versatility of AR systems. The potential of AI techniques in the AR space is clear from the unprecedented success of AI in image processing and its allied tasks. As AR starts with camera images and terminates with rendered images with augmentations projected onto the display, AI has the potential to bring revolutionary changes in AR applications similar to the way AR has in manufacturing applications. Already, AI has brought tremendous changes to the manufacturing field (Y. Liu et al. 2020) and there are many more applications which can be adopted into manufacturing applications like inspection (Zhang et al., "Convolutional Neural Network," 2019) and material information management (Furini et al. 2016). Deep learning (DL), a subset of AI techniques, has surpassed decades of performance of classical object detection techniques in just a decade. Similarly, AI techniques can introduce a paradigm shift in the performance of all the tasks related to AR. The AR space, which is mostly dominated by fiducial markers, can potentially adapt to natural features. AR, when powered by AI, can be adopted without modification to the environment. DL can improve the accuracy and robustness of calibration, detection, tracking, camera pose estimation, and registration tasks. Ontology, a subset of AI, can simplify storage and the dissemination of procedural information. DL can be used in the development of adaptable augmentations within virtual object creation. Finally, DL can be utilised when determining what data to display and the optimal way to display this data to the user in rendering. Recently, the untapped potential of AI in AR applications is being explored. However, so far, most researchers have limited the utility of DL techniques mostly to detection and to simpler tasks like rectangular bounding boxes to localise (Abdi and Meddeb 2017; Subakti and Jiang 2018) or annotate (Rao et al. 2017; Park et al. 2020) the identified objects.

The main contributions of this research work are as follows:

(1) Review the current strategies utilised within AR systems for each of the modules of the AR computational pipeline.
(2) Identify the shortcomings of the current approaches utilised in AR systems for each module.
(3) Justify the use of AI-based methods for each module.
(4) Review current AI-based methods utilised in AR for each module.
(5) Describe relevant AI-based research for each module as related to manufacturing.
(6) Introduce various future research directions relating to the AR pipeline.
(7) Provide conclusive remarks based on the review.

This review paper is structured as follows. Section 2 contains an overview of the current research and challenges at the intersection of AR, AI, and manufacturing. Sections 3 through 11 deliberate each module of the framework within Figure 1. Each section includes an *introduction*; *current approaches in* AR systems; a discussion about the *limitations of current approaches and the potential for AI*; current approaches *within AR that utilise AI*; *potential AI* methods that can be utilised in AR; and a *conclusion* in a sequential order. Section 12 outlines the key and immediate future research directions that are derived from and inspired by the insights gained from the preceding sections. Finally, Section 13 provides an overview and analysis of the work described in the previous sections. An overview of the topics reviewed and the conclusions drawn from all the modules of the procedure-based AR pipeline is provided in Table 1.

2. AR & AI in manufacturing

The manufacturing industry is currently undergoing an industrial revolution deemed Industry 4.0. The last three industrial revolutions included mechanisation in Industry 1.0, mass-production in Industry 2.0, and automation in Industry 3.0 (Ababsa 2020). Industry 3.0 automated all possible manufacturing activities and has seen a swarm of robots and sensors on the shop floor. Industry 4.0 or smart manufacturing is founded upon technologies whose aspirations include (1) accelerating the remaining segment of work that needs human expertise and skills; and (2) incorporating intelligence into the automated infrastructure to make robotic systems autonomous and able to make decisions from data (Rüßmann et al. 2015). Aspiration (1) made the provisions for the incorporation of AR into the scope of manufacturing. Aspiration (2) paved the way for integrating AI into the manufacturing industry. Significant progress in terms of theory and practice has been made in each of these directions. The manufacturing industry has been a huge beneficiary of these technologies.

2.1. AR-assisted manufacturing

As AR can facilitate uninterrupted information exchange without the need of altering the operator's attention, it can accelerate manufacturing activities. Thus, it has become one of the nine key disruptive technologies that are driving Industry 4.0 (Rüßmann et al. 2015). The majority of the applications of AR in the manufacturing industry are

Table 1. Overview of the concepts, limitations of classical approaches, and advantages of AI approaches for each module of the procedure-based AR system given in Figure 1.

Section	Classical methods	Limitations of classical methods	AI methods	Advantages of AI methods
(§3) Camera Calibration	• Linear Techniques (Offline, Intrinsics Unknown) • Non-Linear Techniques (Offline, Intrinsics Unknown) • Calibration-Free (Offline, Intrinsics Unknown) • Self-Calibration (Offline, Intrinsics Unknown) • Camera Pose Estimation for Extrinsics[a] (Offline, Intrinsics Known)	Classical methods: • Are computationally expensive or highly complex to implement (Donné et al. 2016); • Are cumbersome, time-consuming, and prone to errors (Tuceryan, Genc, and Navab 2002; Bogdan et al. 2018); • Require prior knowledge of the scene or strong assumptions to be imposed (for offline camera calibration) (Seo and Hong 2000; Bogdan et al. 2018); and • Are not robust or adaptive in general scenes (Seo, Ahn, and Hong 1998; Pang et al. 2006).	• Single Image Techniques (Online) • Non-Single Image Techniques (Online)	AI Methods: • Enable adaptive calibration without user intervention; • Require minimal inputs; • Are applicable to general scenes; and • Are not limited to use in predetermined environments.
(§4) Detection	• Fiducials (Artificial Features) • Simple Features (Natural Features) • Advanced Features (Natural Features)	Classical methods: • Cannot engineer all of the features that characterise an object; and • Cannot achieve generic object detection capability.	• Image Classification (Natural Features) • Object Detection (Natural Features) • Semantic Segmentation (Natural Features) • Instance Segmentation (Natural Features)	AI Methods: • Are able to work directly with raw images; • Are capable of learning all essential object features autonomously; and • Are capable of classifying and detecting a wide variety of generic objects.
(§5) Tracking	• Feature-based • Model-based	Classical methods: • Are prone to failure in cluttered environments; • Have limitations with complex scenes involving multiple dynamic objects, variations in illumination, motion blur, and occlusion; and • Exponentially explode as the number of tracked objects increases, making them ineffective for real-time tracking (for MOT[b] algorithms).	• Detection and Tracking • Tracking by Detection	AI Methods: • Eliminate the need for hand-crafted features to design robust trackers; and • Do not explode exponentially as the number of tracked objects increases.

(*continued*).

Table 1. Continued.

Section	Classical methods	Limitations of classical methods	AI methods	Advantages of AI methods
(§6) Camera Pose Estimation	• Pose Estimation from *n* points (P*n*P)	Classical methods: • Lack robustness with large changes in viewpoint, variation in illumination, textureless scenes, motion blur, and other situations; • Are not able to find accurate matching points in all scenarios; • Grow in computational complexity and the proportion of outliers as the size of the model grows (for SfM[c]-based approaches); and • Do not have a constant runtime at inference (for RANSAC[d]).	• Without Auxiliary Learning • With Auxiliary Learning	AI Methods: • Do not require engineered features; • Can be trained to be robust against changes in illumination, perspective, and others; • Have relatively constant complexity; • Have consistent performance at run-time; and • Are able to directly infer the pose from RGB[e] images (thus skipping the Detection & Tracking steps) (Kendall, Grimes, and Cipolla 2015).
(§7) Inverse Rendering - Illumination	• With Auxiliary Information • Without Auxiliary Information	Classical methods: • Have a complex capturing process (for light probes and methods using auxiliary information); • Are computationally expensive and not scalable; • Disturb real-time performance (with light probes); and • Have low accuracy (for methods not using auxiliary information).	• Supervised • Unsupervised	AI Methods: • Do not require definitive models of the geometric or photometric attributes; and • Exploit the attributes of the images within the FOV[f] rather than those beyond the FOV[f].

(§7) Inverse Rendering - Depth Inference	• Perceptual Depth Cues
	• Image Attributes
		Classical methods:
		• Are ineffective in recovering depth information despite the use of perceptual cues due to:
		  ◦ The heavily under-constrained nature of the problem;
		  ◦ Camera characteristics like LDR[g]; and
		  ◦ The use of few cues.
			• Supervised
			• Semi-supervised
			• Unsupervised
				AI Methods:
				• Are capable of exploiting almost all cues at once, thus offering better depth inference; and
				• Are able to estimate depth information from RGB[e] images simultaneously with other tasks (detection, tracking, and pose estimation).

(§8) Procedure Storage	• RDBMS[h] (Databases) • PLM[i] (Databases) • PIM[j] (Databases) • CMMS[k] (Databases) • XML[l] (Databases)	Classical methods: • Are static in nature and do not have the capability to derive further knowledge and relationships.	• Ontology	AI Methods: • Are able to dynamically manipulate data over time; and • Add flexibility into AR systems.

(*continued*).

Table 1. Continued.

Section	Classical methods	Limitations of classical methods	AI methods	Advantages of AI methods
(§9) Virtual Object Creation	• Manual Creation	Classical methods: • Require extensive time and energy from the AR system developer to make augmentations; and • Do not enable adaptive or in-situ augmentations.	• Caption Generation (1D[m]) • Sketch or Contour Generation (2D[n]) • 3D[o] Reconstruction (3D[o])	AI Methods: • Enable augmentations to be created and adapted in-situ; and • Require minimal input for augmentation creation.
(§10) Registration	• Open-Loop • VVS[p] (Closed-Loop)	Classical methods: • Are not fully developed; • Are not in widespread use in AR systems; and • Require predefined features in the scene (for VVS[p] methods).	• Incorporate NNs[q] in VVS[p] • MPC[r] for VVS[p] • Reinforcement Learning for VVS[p]	AI Methods: • Can be used in unknown environments (Saxena et al. 2017); and • Are more accurate, efficient, stable, and robust than IBVS[s] strategies (for AI-based VVS[p] methods) (Comport et al. 2006).
(§11) Rendering	• Rule-Based Methods (Data) • Spatial (Data Visualisation, Filtering) • Knowledge-Based or Semantic (Data Visualisation, Filtering) • Hybrid (Data Visualisation, Filtering) • Geometric-Based (Data Visualisation, Visualization Layouts) • Image-Based (Data Visualisation, Visualization Layouts) • Data Clustering (Data Visualisation)	Classical methods: • Are rigid and based on hard-coded rules (for Data); • Can be destructive and include data loss (for Data Visualisation) (Tatzgern et al. 2016); and • Can fail when the amount of data grows too large (for Data Visualisation) (Tatzgern et al. 2016).	• Layout Generation (Data Visualisation) • Data Clustering (Data Visualisation)	AI Methods: • Are able to incorporate context-awareness, user customisation, expert knowledge, and flexibility into AR systems; • Are more efficient and effective for rendering (with AI-based layout creation); and • Are able to transform data into clustering-friendly nonlinear representations (with DL-based data clustering).

[a] Refer to 'Camera Pose Estimation' Section (§6).
[b] Multiple Object Tracking (MOT).
[c] Structure from Motion (SfM).
[d] RANdom SAmple Consensus (RANSAC).
[e] Red Green Blue (RGB).
[f] Field Of View (FOV).
[g] Low Dynamic Range (LDR).
[h] Relational DataBase Management System (RDBMS).
[i] Product Lifecycle Management (PLM).
[j] Product Information Management (PIM).
[k] Computerised Maintenance Management Systems (CMMS).
[l] Extensible Markup Language (XML).
[m] One-Dimensional (1D).
[n] Two-Dimensional (2D).
[o] Three-Dimensional (3D).
[p] Virtual Visual Servoing (VVS).
[q] Neural Network (NN).
[r] Model-Predictive Control (MPC).
[s] Image-Based Visual Servoing (IBVS).

in three activities: assembly, maintenance, and training. Within assembly, maintenance, and training tasks, technicians must follow specific tasks and procedures in order to perform assembly sequences and procedures. Seamlessly displaying this procedural information to the user is essential to minimise their cognitive load. This directly translates into minimal operating costs, efficient processes, and increased productivity. In AR-assisted assembly and maintenance activities, the operators can perform their work without altering their focus. Operators who use AR for assembly and maintenance can also handle parts that they are working with for the first time. AR eliminates the need of memorising operational instructions. Hence, customisation can be added to products effortlessly. AR-assisted assembly operations have seen a reduction of error of up to 82% (Tang et al. 2003). While information supply remains the primary motivation, several secondary advantages offered by AR are creditworthy in improving productivity. Reduction of the distraction that was imminent in using computers or instruction manuals as the source of information is one of them. Others include collaboration capabilities with the AR system, interconnectivity of the AR system with the manufacturing environment, and the capability for AR systems to adapt to the manufacturing facility and to users' current status.

Many projects have been established to incorporate AR in assembly operations. An AR-based interactive manual assembly design system was developed to predict the user's assembly intents and to manipulate virtual objects (Wang, Ong, and Nee 2013). Chen, Hong, and Wang (2015) used AR to assist the assembly of a gearbox. AR has also been used to remotely support assembly operations to cater to customer's customisation requests (Mourtzis, Zogopoulos, and Xanthi 2019). The AR system offered the part list and assembly instructions at the workstation by retrieving its production schedule.

AR has been utilised for maintenance tasks. Maintenance tasks encompass a wide variety of activities such as inspection, measurement, and service in addition to assembly. A robotic arm was maintained using AR, where an animated ratchet was superimposed (Zubizarreta, Aguinaga, and Amundarain 2019). The AR system estimated the 6-DOF of the rigid parts associated with the robotic arm and showed the motion of the ratchet. The relevant product structure was displayed in the maintenance of a ubiquitous car (Lee and Rhee 2008a). The pertinent technical information was projected on the engine of a motorbike during its maintenance (Uva et al. 2018). The electronic components of a wind turbine were inspected with the help of AR (Quandt et al. 2018). AR has successfully assisted maintenance technicians in improving overall productivity by allowing bi-directional context-aware authoring (Zhu, Ong, and Nee 2013).

AR has been recognised as an effective tool to disseminate knowledge by improving training methods. An HMD was utilised to train employees assembling engines at the vehicle manufacturer BMW (Werrlich, Nitsche, and Notni 2017). Users have been trained for assembling a milling machine (Peniche et al. 2012) as well as for welding (Quandt et al. 2018). AR-assisted training for complex manufacturing activities has been evaluated to be as effective as face-to-face training (Gonzalez-Franco et al. 2016). Gavish et al. (2015) concluded from a usability study that using AR for training in industrial assembly and maintenance tasks shall be encouraged, and AR is more effective than virtual reality.

Additionally, AR has assisted in product design by annotating design variations (Bruno et al. 2019); in assembly and disassembly sequence planning (Wang et al. 2013; Chang, Nee, and Ong 2020); in packaging by overlaying the virtual correction template on the die itself (Álvarez et al. 2019); in planning the orientation of robots (Fang, Ong, and Nee 2013); in casting to study users' hand movement (Motoyama et al. 2020); to highlight spot-weld locations (Doshi et al. 2017); and in shop floor management (Wang et al. 2020). Furthermore, AR can assist in simulating and developing manufacturing processes and products before they are actually carried out. That will provide support in getting the activities done correctly in their first attempt. Moreover, collaborative AR applications can open up the opportunity to simultaneously leverage the expertise of multiple people from geographically diverse locations.

2.2. AI-driven manufacturing

The advent of automation (Industry 3.0) in manufacturing brought about the widespread use of sensors and robots. The use of sensors has thus resulted in a large amount of data available from the manufacturing process, or big data. Deriving decisions from big data and making robots autonomous to enable them to make their own decisions have been two of the key goals of Industry 4.0 (Rüßmann et al. 2015). Autonomous robots and big data analytics have resulted in the widespread use of AI. AI technologies provide a way to manage and utilise this big data (Aggour et al. 2019). AI strategies have been utilised in all aspects of the manufacturing process and supply chain from design, operations management, and production to maintenance and assembly (Baryannis et al. 2019; Zhang et al., "A Reference Framework," 2019). Real-time data from sensors can inform decision-making in the manufacturing process (Deac et al. 2017). This data can allow manufacturing systems to respond in real-time to changing demands and conditions throughout the PLM with respect to the industry, supply network, and customer needs (Zhang and Kwok 2018; Yao et al., "From Intelligent Manufacturing," 2017). Smart manufacturing aims to be able to utilise big data collected from the entire product lifecycle into the manufacturing system in order to improve every aspect of the process (Tao et al. 2018). The main goal behind using big data is to achieve fault-free or defect-free processes (Escobar and Morales-Menendez 2018).

AI strategies can be used to aid in the design of manufacturing systems. For instance, Stocker, Schmid, and Reinhart (2019) utilise reinforcement learning to automate the design process for vibratory bowl feeders (VBFs). The reinforcement learning mimics the manual trial-and-error approach currently used to find the trap configuration with the highest efficiency. Sanderson, Chaplin, and Ratchev (2019) developed an innovative Function-Behaviour-Structure model for Evolvable Assembly Systems (EAS) consisting of an ontology and design process. The Function-Behaviour-Structure model is used as input to a modelling design process for assembly systems. In this methodology, the design process is integrated into the manufacturing control system.

AI methods can be utilised for handling operations management. Baryannis et al. (2019) highlight the usage of various AI methods for supply chain risk management, including stochastic programming, robust optimisation, fuzzy programming, hybrid approaches with mathematical programming, network-based approaches, agent-based approaches, reasoning, and machine learning (ML) with big data. Li et al., "Machine Learning" (2020) propose a rescheduling framework that balances rescheduling frequency and the accumulation of delays. The framework combines ML with optimisation in order to solve a flexible job-shop scheduling problem (FJSP).

Various AI strategies can be applied to production decision-making and support. For instance, Palmer et al. (2018) developed a manufacturing reference ontology to support flexible manufacturing decision-making. This ontology is not only useful to elucidate understanding across domains but for information sharing and support for interoperable manufacturing systems as well. Zhou et al. (2020) present a framework for knowledge-driven digital twin manufacturing cell (KDTMC) to support autonomous manufacturing. The framework utilises digital twins, dynamic knowledge bases, and knowledge-based intelligent skills. Possible applications for the framework include process planning, production scheduling, and production process analysis and regulation.

When considering maintenance, service, repair, and inspection, AI techniques can be used for quality inspections and predictive maintenance. Yacob, Semere, and Nordgren (2019) propose Skin Model Shapes, which are able to generate digital twins of manufactured parts. The Skin Model Shapes are used in conjunction with ML methods to detect anomalies and unfamiliar changes with parts to ensure part quality. Kang et al. (2018) compared both linear regression and artificial neural networks (ANNs) for predictive models for part failure. Part failure predictive models are used to predict product quality indicators based on both product quality and manufacturing parameters. AI has also been widely used in inspection (Yang et al. 2020), diagnosis (Matei et al. 2020), and prognostics and health monitoring of machines (Yang and Rai 2019).

Finally, AI can be used for scheduling and production in assembly lines. Huo, Zhang, and Chan (2020) use a fuzzy control system to adjust the assembly line depending on real-time machine health information. The control system is designed to smooth the workloads of workstations in the assembly line. Azadeh et al. (2012) describe a multi-layer feed-forward back-propagation ANN developed for a stochastic Two-Stage Assembly Flow-Shop Scheduling Problem (TSAFSP). The TSAFSP consists of (1) manufacturing of a product's parts by various machines in parallel and (2) assembly of all of the product's parts. The TSAFSP addressed by Azadeh et al. (2012) accounts for stochastic machine breakdown and processing times.

2.3. AI-driven AR-assisted manufacturing

Both AR-assisted manufacturing and AI-driven manufacturing have evolved independently of each other and have accelerated manufacturing separately. Most past work has used AR as a tool to supplement manufacturing activities. Those attempts, although numerous, suffer from a variety of shortcomings, as discussed in Section 1. These shortcomings can be eliminated by using AI as a tool in the computational framework of AR. There is a domain of research that uses AI as a tool in AR applications that has been overlooked so far. The succeeding sections (3 to 11) detail the mode of incorporation of AI in AR-assisted manufacturing applications. While AI-driven AR-assisted manufacturing applications will be instrumental in the future, the adoption of AI into AR-assisted manufacturing faces challenges from both technical and managerial fronts.

There are many challenges that are presented from the technical front while incorporating AI strategies within AR-assisted manufacturing applications. First, manufacturing environments offer limited variation in objects to support detection and tracking. Discerning between various objects such as nuts or bolts which differ by size, or objects that have undergone the same terminal machining process, is very difficult. That makes creating generic AI challenging. Training AI algorithms to recognise anomalies is an expensive affair because creating data that capture the anomalies is usually time and resource-intensive. Further, collecting and sharing big data from the manufacturing processes to the AR system is another challenge. Transmission of relevant data to AR systems is essential in order for the AR system to be up-to-date and context-aware. Privacy and security issues can arise when industrial data are stored in cloud platforms (Bajic et al. 2018). Finally, manufacturing processes frequently include time-varying uncertainties (Qin and Chiang 2019). These uncertainties must be accounted for in order for data-driven or AI methods to perform accurately within AR-assisted manufacturing systems.

There are many challenges that emerge from the managerial side when incorporating AI technologies into the manufacturing environment. As described by Arinez et al. (2020), implementing AI on a large scale such as a shop-floor is a challenge in itself because most implementations to date are designed and tested in a laboratory. Integrating AI with current manufacturing systems is another issue; finding solutions to issues with parallelism in computing (Jordan and Mitchell 2015) and AI compatibility with first principles models and processes (Qin and Chiang 2019) are some major challenges with implementing AI in manufacturing. When creating training datasets for AI algorithms, it may be necessary to collect data from actual machine processes. In cases of predictive maintenance, diagnostics, or prognostics, data would need to be collected from machine failures or malfunctions. This data collection process may be dangerous for the developer and costly for the manufacturer. Supervised learning, which has been the most successful segment of AI, needs labels of the data. Labeling is expensive in terms of time and can require skilled labour in manufacturing applications. Next, utilising the results from AI-driven manufacturing requires special interpretation (Wuest et al. 2016; Qin and Chiang 2019). Interpretation of results includes an analysis of the output, characteristics of the chosen algorithm, parameter settings, the expected output, and the input data along with any preprocessing done (Wuest et al. 2016). The correct interpretation is a challenge that manufacturers must face when incorporating AI into their processes. Next, manufacturing processes are often dangerous and require special safety considerations. Dagnaw (2020) describes challenges with incorporating safety into AI-driven manufacturing processes, including a need to add labels relating to safety in AI algorithms, incorporating knowledge-based reasoning into AI algorithms, and utilising AI algorithms in the manufacturing space for dynamic risk assessment. Integrating these subsystems in their current forms together in the manufacturing environment requires programming expertise, thus necessitating organisational restructuring (Martinetti et al. 2019; Masood and Egger 2020). Next, AI methods tend to be black-box in nature; they are not explainable. That questions their credibility. Lack of reasoning diminishes users' trust in AI-driven AR-systems. Once AI is included in AR-assisted manufacturing, there may be a prolonged adjustment period for employees because of a lack of trust in AI (Qin and Chiang 2019; Arinez et al. 2020). Moreover, a lack of explainability can hinder the organisation's compliance with safety and legal requirements. On the organisational level, access to computing and human resources in order to implement AI in manufacturing along with legal and regulatory issues are other major challenges (Dagnaw 2020). Dagnaw (2020) concisely summarises this challenge as the need for the development of AI applications that are ethical, explainable, and acceptable.

While the integration of AI into AR-assisted manufacturing will face the above-mentioned technical and managerial challenges, the advantages of AI-driven AR far outweigh these shortcomings. With AI, AR systems will become intelligent like autonomous robots and sensors. They will be able to detect and track generic objects without mistake. They will be ubiquitous like sensors in Industry 3.0. AI-driven AR systems will be able to work autonomously in the manufacturing environment with minimal effort and input from humans. AI provides AR systems with the ability to work efficiently and effectively as a unified whole with the manufacturing environment and with human operators, thus enabling Industry 4.0. While the ubiquitous use of AI methods within AR systems is perceived as a logical step towards the development of smart industry, only a few AR systems in the research literature utilise AI. As a way to bridge the gap between the industry of today and Industry 4.0, next, we highlight the shortcomings of classical approaches in each area of the AR pipeline and introduce better AI alternatives, starting with camera calibration.

3. Camera calibration

The first step in the creation of an AR application is camera calibration. Camera calibration is a process that captures the internal parameters of the real camera so that the virtual camera can mimic the real camera. Camera calibration readies the virtual environment for detection, tracking, and camera pose estimation so that the virtual objects can be correctly aligned with real objects. A hierarchical description of the topics covered within camera calibration can be found in Figure 2.

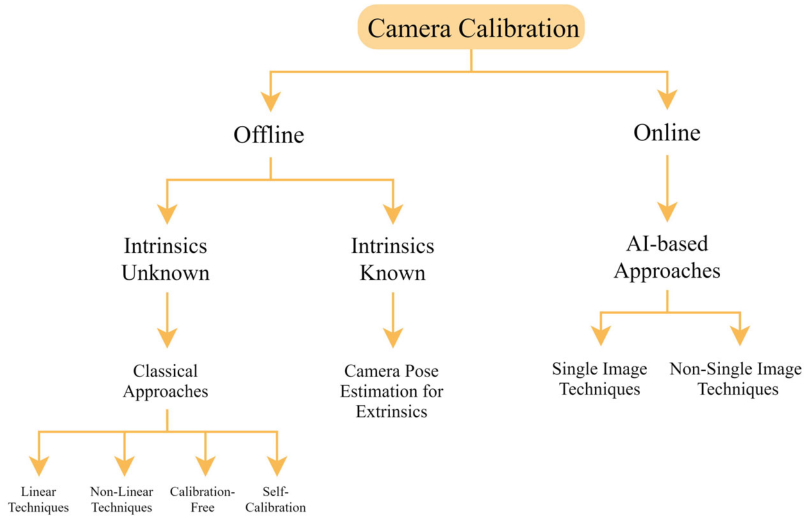


Figure 2. Visual depiction of the topics reviewed in camera calibration.

Many AR systems are developed using Software Development Kits (SDKs) in order to streamline the tasks required to develop AR systems, including camera calibration. These SDKs are either available as a supplement to proprietary devices such as Microsoft HoloLens, Schenker META, Magic Leap, Android devices, Apple devices, and Lightform projectors, or as separate libraries available for use in development. If the SDK is associated with a proprietary device, a calibration profile is available for use. For SDKs that are independently developed, manual or direct calibration techniques are used with calibration patterns as in the Wikitude (Wikitude GmbH 2018), Vuforia (PTC 2020), and Kudan (Kudan Inc. 2020) SDKs.

3.1. Current approaches for camera calibration in AR

The camera calibration techniques currently in use can be classified into offline and online methods. While camera calibration is traditionally an offline process, many researchers have found ways to perform the necessary camera calibration steps online by utilising user intervention or by using computer vision (CV) techniques. Offline camera calibration methods can further be split into two categories: (1) methods where the camera calibration parameters do not need to be known in advance, and (2) methods where the intrinsic camera parameters are pre-calibrated (Pang et al. 2006). The majority of AR systems that perform offline camera calibration belong to the second category. In this case, the camera calibration process solely involves the determination of the extrinsic parameters.

Within the first category of methods where the calibration parameters are not known *a priori*, there are two main categories of strategies implemented in AR systems to find the intrinsic camera parameters: (1) methods that utilise linear techniques and (2) methods that utilise non-linear techniques (Malek et al. 2008). Additionally, some strategies find the intrinsic and extrinsic parameters separately or find the camera matrix as a whole. If the latter strategy is implemented, it is possible to decompose the resultant camera matrix into its intrinsic and extrinsic matrix parts (Axholt 2011) using the Kruppa equations (Kruppa 1913). Having both the intrinsic and extrinsic parameters defined separately is required to generate shading, shadow, and illumination effects on virtual objects (Seo and Hong 2000; Pang et al. 2006).

Linear calibration techniques involve the use of the linear least-squares method to obtain the projection matrix (Malek et al. 2008). The most common method utilised within linear calibration techniques is Direct Linear Transform (DLT), as in Malek et al. (2008) and Teichrieb et al. (2007). Numerical or iterative techniques are utilised to perform linear optimisation, as in Malek et al. (2008), where the Gauss-Jacobi numerical technique was used to solve the linear system of equations. Linear techniques are fast but do not take into account any non-linear radial or tangential lens distortion (Malek et al. 2008). Omitting this non-linear information relating to lens distortion leads to inaccurate calibration results (Gomez, Simon, and Berger 2005).

Non-linear optimisation calibration techniques estimate the camera parameters through iteration based on a predefined function (Malek et al. 2008). Some common functions used in non-linear methods are the reprojection error (Dash et al. 2018) and the residual errors in calibration (Yuan, Ong, and Nee 2006). Numerical or iterative techniques such as the Levenberg-Marquardt

(LM) algorithm (Dash et al. 2018) and Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm (Xie et al. 2018) are utilised to perform non-linear optimisation in many cases (Comport et al. 2006). Many of the non-linear optimisation strategies in AR systems derive the camera parameters based on the motion of the camera or Structure from Motion (SfM) as in Seo and Hong (2000), Teichrieb et al. (2007), Simões et al. (2013), Seo, Ahn, and Hong (1998), Gordon and Lowe (2006), and Skrypnyk and Lowe (2004). Although non-linear techniques have good accuracy, they are slow and have a high computational cost (Malek et al. 2008). Additionally, non-linear techniques tend to converge to local minima (Comport et al. 2006).

For augmentation of video streams or in fully offline AR applications, methods such as the bundle adjustment technique can be used for camera calibration, as in Gordon and Lowe (2006), Heyden and Astrom (1997), and Heyden and Åkström (1998).

A subset of camera calibration focuses upon self-calibration (i.e. auto-calibration) and calibration-free methods. Self-calibration and calibration-free methods are useful in cases where the intrinsic parameters change over time (e.g. zooming of the camera) and fiducial points are not available (Seo and Hong 2000) or when attempting to calibrate general video images (Seo, Ahn, and Hong 1998). Self-calibration methods provide the Euclidean motion of the real camera in addition to the calibration parameters (Seo and Hong 2000). The method by Faugeras, Luong, and Maybank (1992) utilises the epipolar transformation found between point matches of image sequences and the Kruppa equations to link the epipolar transformations to the absolute conic in order to find the intrinsic camera parameters. Pollefeys, Koch, and Van Gool (1999) work off of the absence of skew in the image plane in order to allow for metric self-calibration of zooming lens cameras. Taketomi et al. (2014) dynamically calculated the intrinsic and extrinsic parameters of a zooming lens camera using an energy minimisation formulation. Chendeb, Fawaz, and Guitteny (2013) apply self-calibration to AR by combining interferometry and pattern recognition. The methods by Seo, Ahn, and Hong (1998) and Seo and Hong (2000) use reference and video images to recover structure and motion information for the object of interest. The parameterised cuboid structure (PCS) method (Chen, Yu, and Hung 1999) is used to insert animated virtual objects into an AR scene by using a cuboid structure as the reference object. Finally, in Kutulakos and Vallino (1998), an affine camera model with an orthographic camera is used in order to produce virtual augmentations. Kutulakos and Vallino (1998)'s method is not generally applicable to cameras with perspective projection and is unable to take advantage of scene illumination effects because it does not utilise Euclidean geometry (Seo and Hong 2000).

3.2. Limitations of current approaches for camera calibration in AR & potential for AI

The camera calibration methods currently utilised in AR systems are computationally expensive or highly complex to implement (Donné et al. 2016). Further, any methods that require the user to align elements in the space or to collect data for the system are cumbersome, time-consuming, and prone to errors (Tuceryan, Genc, and Navab 2002; Bogdan et al. 2018). Any camera calibration methods that are performed offline require some knowledge of the scene or strong assumptions to be imposed on the scene prior to the use of the system, and this is limiting for applications where the use environment changes frequently (Seo and Hong 2000; Bogdan et al. 2018). Even in cases of self-calibration methods, none of these strategies have been found to be robust in general scenes (Pang et al. 2006).

In order to overcome the limitations of existing classical camera calibration methods, AI-based camera calibration strategies are recommended. AI-based camera calibration methods: (1) enable adaptive calibration without the need for user interaction, (2) are applicable to general scenes, and (3) are not limited to use in predetermined environments.

3.3. AI approaches for camera calibration in AR

To the best of our knowledge, there have been no AI-based camera calibration methods utilised within AR-assisted manufacturing applications.

3.4. Potential AI approaches for camera calibration in AR

The AI-based camera calibration methods for use in AR systems can be categorised into single image and non-single image methods. Single image methods are solutions that take as input a single image. Non-single image methods are those which take as input multiple consecutive frames or a video segment.

3.4.1. Single image methods

Within single image strategies, Bogdan et al. (2018) present DeepCalib, a Convolutional Neural Network (CNN) with Inception-V3 architecture (Szegedy et al. 2016) that estimates the intrinsic parameters of the camera. The CNN is trained using omnidirectional images from the Internet. Their method does not require motion estimation, a calibration target, consecutive frame inputs, or information about the structure of the scene. The AI-based self-calibration method by Zhuang et al. (2019) finds camera intrinsics and radial distortion for Simultaneous Localization and Mapping (SLAM) systems using two-view epipolar constraints. In this work, the CNN is trained on a set of images with varying intrinsics and radial distortion. The solution by Lopez et al. (2019) finds both intrinsic and extrinsic camera parameters from single images with radial distortion. They utilise a CNN that is trained to perform regression on varying extrinsic and intrinsic parameters. Donné et al. (2016) present ML for adaptive calibration template detection or MATE. A CNN is trained using standard checkerboard calibration patterns in order to detect these calibration patterns in single images. After the calibration pattern is detected, other methods can be used to derive the camera parameters. He, Wang, and Hu (2018) estimate depth and focal length from a single image. While this method cannot be used to find all of the intrinsic parameters of the camera, it effectively finds the focal length of an image. Hold-Geoffroy et al. (2018) utilise a deep CNN (DCNN) to infer the focal length and camera orientation parameters of a camera from a single image. The DCNN is trained using the SUN360 database (Xiao et al. 2012), a large-scale panorama dataset. A simple pinhole camera model is utilised, so distortion parameters are omitted.

3.4.2. Non-single image methods

Within AI-based non-single image camera calibration methods, Brunken and Gühmann (2020) utilise multiple consecutive image frames as input to a Neural Network (NN) that automatically determines intrinsic, extrinsic, and distortion parameters from a camera. In this method, two views are transformed to match a third reference view through plane homography mappings. Gordon et al. (2019) utilise frames from videos to estimate the camera calibration parameters. Depth, egomotion, object motion, and camera intrinsics are learned from video frames in an unsupervised manner. Their method utilises two NNs: one for depth estimation and one used to predict camera motion and camera intrinsics from two consecutive video frames.

3.5. Conclusion

The classical camera calibration techniques currently used in AR systems are not adaptive to general scenes (Seo, Ahn, and Hong 1998), which are commonly encountered in manufacturing. When offline calibration parameter estimations are used, any changes that occur to the internal parameters of the camera while in use, such as zooming, render the originally-estimated camera parameters as no longer useful unless known two-dimensional (2D)-3D correspondences are still present in the environment (Seo and Hong 2000). In order to make the camera calibration process more robust, AI-based approaches are recommended. AI-based methods are able to dynamically predict the calibration parameters for a camera while in use, given a single image or multiple image frames. This then saves time for developers and operators when utilising AR systems in the manufacturing environment. While AI-based NN methods have their own challenges, such as extensive effort required for training, these challenges are outweighed by the benefits gained from being able to dynamically estimate camera intrinsics in unknown environments (Zhang et al., "Line-Based Geometric," 2019).

4. Detection

In a real-time setting, object detection initialises the process of determination of camera extrinsics. The real objects in the environments are detected and localised to instantly or recurrently (with tracking) estimate the pose of the camera. The techniques discussed in this section are summarised in Figure 3.

4.1. Current approaches for detection in AR

The classical methods of detection can be classified into marker-based and markerless. Marker-based methods place fiducials, which have characteristic properties (shape, patterns, etc.) for easier identification in the scene. Circular 2D barcodes (Naimark and Foxlin 2002), color-coded 3D markers (Mohring, Lessig, and Bimber 2004), 3D-cones (Steinbis, Hoff, and Vincent 2008), and libraries like ARToolKit (Kato and Billinghurst 1999), ARTag (Fiala 2005), and ARToolKitPlus (Wagner and Schmalstieg 2007) are a few fiducial markers that are widely used for AR applications. Although they are easier to track, these fiducials need to be placed in the environment, which may not be possible (in industrial scenes) or desirable (aesthetically unpleasant or a distraction).

The markerless methods detect natural features available in the scene. They identify specific key points that are unique in their neighbourhood and create a feature. Hence, they are local in nature, and this makes them invariant to scaling and rotation. Scale-invariant feature transform (SIFT) (Lowe 1999, 2004) and speeded-up robust features (SURF) (Bay, Tuytelaars, and Van Gool 2006) are two of the most widely used markerless methods for matching images of an object from different viewpoints. Natural features like edge and texture information (Vacchetti, Lepetit, and Fua 2004), edges and

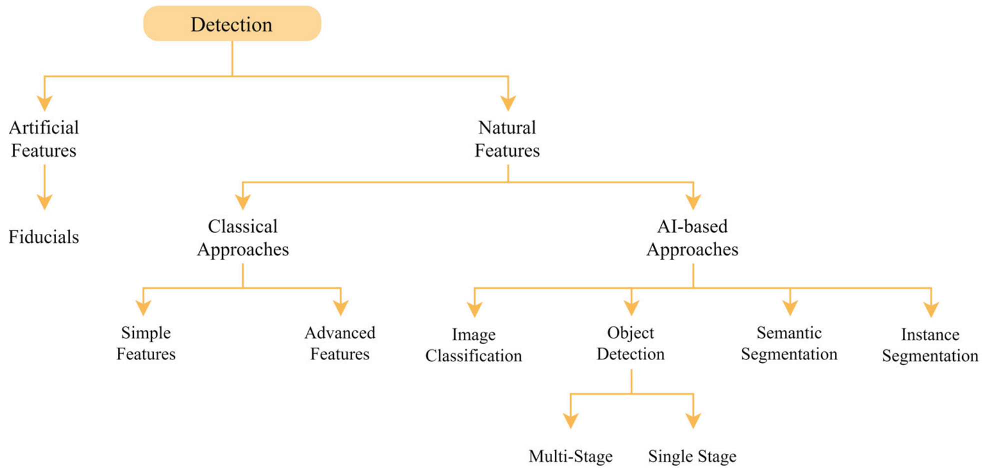


Figure 3. Visual depiction of the topics reviewed in detection.

corners (Park and Park 2010), lines with a 3D model (Wuest, Vial, and Strieker 2005; Ababsa et al. 2008), SIFT (Lee and Hollerer 2008; Wagner et al. 2008, 2009; Ha et al. 2011; Teng and Wu 2012), Ferns (Ozuysal, Fua, and Lepetit 2007; Wagner et al. 2008, 2009), SURF (Takacs et al. 2008) etc. have already been used for object detection in AR applications.

Usually, such point-based natural features like SIFT and SURF are susceptible to failure in cases of variations in lighting, intrinsic parameters of the camera, and motion blur. So, they were further developed as Principal Components Analysis (PCA)-SIFT (Ke and Sukthankar 2004) to make SIFT more robust to deformations, Affine (A)SIFT (Morel and Yu 2009) to make SIFT fully *affine* invariant, Global (G)SIFT (Mortensen, Deng, and Shapiro 2005) for adding *global* characteristics, Colored (C)SIFT (Abdel-Hakim and Farag 2006) for incorporating *color* invariant attributes, and fully affine invariant (Fair) SURF (Pang et al. 2012) to make SIFT both fully affine invariant and computationally efficient.

4.2. Limitations of current approaches for detection in AR & potential for AI

Even after several improvements on the point-based natural features, their performance at image processing did not portray significant improvement. They still failed in the presence of occlusion, noise, clutter, distortions, and variations in illumination characteristics. That can be attributed to the following factors:

- Identification, modelling, and inclusion of all the features is a formidable task. Every additional improvement created a new feature descriptor, which also lacked some other characteristics, e.g. CSIFT lacked global contextual information, whereas GSIFT lacked color attributes.
- Hand-crafted features have limited flexibility. They have to be engineered for each class of images. That amplifies the effort in feature engineering as the number of classes increases. Thus, it limits the design of a generic object detection framework, thereby preventing their proliferation in AR applications.

ML-based approaches, although developed alongside feature-based approaches, could not perform comparably until the development of huge datasets and computational capacities. The new wave of ML techniques, under the umbrella of DL, discarded the manual feature extraction step, and work directly with the raw images. The CNNs, a class of DL, eliminated the need for feature engineering in object detection. CNNs are capable of learning all the essential features of the objects by themselves. Moreover, state-of-the-art DL techniques are capable of classifying and detecting a wide variety of generic objects.

4.3. AI approaches for detection in AR

Mask Region-based CNN (R-CNN) (He et al. 2017) was used for object detection and instance segmentation and subsequent 3D annotation (Park et al. 2020). Its practical utility was studied in the case of the maintenance and inspection of a 3D printer. You Only Look Once (YOLO)9000 (Redmon and Farhadi 2017) was used for detecting objects, and the exemplar-based Criminisi's method (Criminisi, Perez, and Toyama 2003) was used for filling the void created by removing the object in case of an inpainting application (Kim et al. 2018). The use of ML eliminated the selection of the area to be removed and replaced it with label selection. A CNN powered by AlexNet (Krizhevsky, Sutskever, and Hinton 2012) was used in DeepAR (Akgul, Penekli, and Genc 2016) for detecting 2D images of the target as part of the camera-based tracking mechanism. SqueezeNet (Iandola et al. 2016), trained on the pattern analysis, statistical modelling, and computational learning (PASCAL) Visual Object Classes (VOC) (Everingham et al. 2010), was implemented on a mobile device for constructing a fast markerless indoor mobile AR system (Subakti and Jiang 2018). Another fast markerless indoor mobile AR system (Subakti and Jiang 2018) was powered by MobileNet (Howard et al. 2017; Sandler et al. 2018) for recognising different machines and retrieving machine status in an indoor industrial setting. The MobileNet classified the image and detected the boundary for labeling. The AR application offered the user broad characteristics of the machine when the user was far and detailed machine information when the user was close.

4.4. Potential AI approaches for detection in AR

Most of the current work utilising DL techniques in AR applications limits the utility of DL at detecting objects using rectangular bounding boxes. However, DL offers precise pixel-level details. The complete understanding of a scene primarily consists of four tasks: image classification, object detection, semantic segmentation, and instance segmentation, in the order of increasing fine-grained inference (Garcia-Garcia et al. 2017; Wu, Sahoo, and Hoi 2020). They are briefly discussed below.

4.4.1. Image classification

Image classification attributes the whole scene into a single class or labels the scene using a single label. The NN acts as an operator to attribute the whole scene to a single class with a significantly higher probability than other classes. Although image classification by itself does not have significant importance in AR applications, it is a predecessor to subsequent tasks of object detection and others. The successful architectures in image classification are adopted for all other tasks.

LeNet-5 (LeCun et al. 1998) was among the first CNNs to taste success on a comparatively simple Modified National Institute of Standards and Technology (MNIST) dataset consisting of handwritten digits. AlexNet's (Krizhevsky, Sutskever, and Hinton 2012) success in classifying the relatively complex and diverse ImageNet database (Deng et al. 2009) triggered the widespread adoption of CNNs for image classification. It revolutionised the task of image classification by including the rectified linear unit (ReLU) (Nair and Hinton 2010) for improving training times, repeated blocks of convolutional layers and max-pooling layers, data augmentation, dropout regularisation (Hinton et al. 2012) to combat overfitting, softmax classifier, and parallel computing. It was succeeded by VGG-Net (Simonyan and Zisserman 2014), which endorsed the use of deeper networks, stacked convolutional layers, and smaller filters with stride and transfer learning. GoogLeNet (Szegedy et al. 2015) used inception modules consisting of parallel convolutional layers with filters of different sizes to allow multi-scale processing. ResNet (He et al. 2016) advocated the use of extremely deeper networks to learn better features and the use of additional residual connections for mitigating the exploded computational expense.

Recurrent NNs (RNNs) have also been explored for image classification (Visin et al. 2015) as an alternative to CNNs. SqueezeNet (Iandola et al. 2016) based on AlexNet, MobileNet (Howard et al. 2017) based on depth-wise separable convolutions, and ShuffleNet (Zhang et al. 2018) based on channel shuffling and point-wise group convolutions are a few more CNNs developed for edge devices that have performance comparable to the state-of-the-art architectures.

4.4.2. Object detection

In addition to classifying the image (or video frame) for (an) object(s), the object detection task also has to localise it. Usually, the DL techniques detect approximate boundaries and bound the detected objects using rectangular bounding boxes. Most object detection algorithms can be broadly categorised into two classes: multi-stage and single-stage.

4.4.2.1. Multi-stage approach. This is a traditional approach where a set of candidate bounding boxes (region proposals) are generated first. The features are extracted from those boxes using a CNN. And then, the proposed regions are classified into different categories for labeling them.

R-CNN (Girshick et al. 2014) was the first successful attempt at object detection. It integrated selective search (Uijlings et al. 2013) for region proposal with AlexNet (Krizhevsky, Sutskever, and Hinton 2012) for feature extraction, and a Support Vector Machine (SVM) for classification. Its multi-stage training pipeline needed separate training for each stage. Thus, it was slow, and the overall optimisation was difficult. Hence, it was developed into Fast R-CNN (Girshick 2015) to jointly train for classification and regression of the bounding boxes. It also supported parameter sharing. However, the external region proposal by the selective search was still a bottleneck. It was replaced by a region proposal network (RPN) based on a Fully Convolutional Network (FCN). That gave birth to Faster R-CNN (Ren et al. 2015), a completely end-to-end trainable CNN-based framework with near real-time performance.

Spatial Pyramid Pooling (SPP)-Net (He et al. 2015), Feature Pyramid Network (FPN) (Lin et al. 2017), Region-based FCN (R-FCN) (Dai et al. 2016), Mask R-CNN (He et al. 2017) etc. are a few more architectures for object detection that follow a multi-stage approach. Although accurate, all these multi-stage approaches are expensive in space and time. That limits their efficacy in AR applications which call for real-time performance.

4.4.2.2. Single-stage approach. In a single-stage approach, both the location of the objects and the class probabilities are predicted from the full image using a single CNN.

Overfeat (Sermanet et al. 2013), based on FCNs, was the first single-stage object detector that received significant recognition. It had sub-par accuracy compared to R-CNN (Girshick et al. 2014). YOLO (Redmon et al. 2016) divides the image into a $G \times G$ grid. Each cell has to predict the object in the cell. Each of those grid cells predicts b bounding box locations and c class probabilities. YOLO directly predicted the bounding boxes and associated probabilities from the pixels. YOLO showcased outstanding real-time performance of 45 Frames Per Second (*FPS*), and a miniature version referred to as 'Fast YOLO' (Redmon et al. 2016) offered 155 *FPS*. YOLO was further developed as YOLO9000 (Redmon and Farhadi 2017) for detecting more than 9000 object categories, and as YOLOv3 (Redmon and Farhadi 2018) for improving speed and accuracy. However, because of the coarse grid structure, and as each of the cells can only be attributed to one class, YOLO was unable to properly detect small objects which may be present in groups. To overcome these shortcomings, Single-Shot Detector (SSD) (Liu et al. 2016) was developed. SSD consolidated concepts from YOLO, multi-scale convolutional features (Hariharan et al. 2016) and RPN from R-CNN (Girshick et al. 2014) for faster detection without compromising the accuracy of detection.

AttentionNet (Yoo et al. 2015), G-CNN (Najibi, Rastegari, and Davis 2016), Multibox (Erhan et al. 2014), Deconvolutional SSD (DSSD) (Fu et al. 2017), Deeply Supervised Object Detector (DSOD) (Shen et al. 2017) etc. are few other single-stage object detectors.

4.4.3. Semantic segmentation

Unlike classification and detection, in which a group of pixels is attributed to a single class, *semantic segmentation* attributes each pixel to a particular class. The DL learns a pixel-wise classifier. Semantic segmentation identifies almost precise pixel-level boundaries of the objects.

FCN (Long, Shelhamer, and Darrell 2015), built upon state-of-the-art image classification models like AlexNet (Krizhevsky, Sutskever, and Hinton 2012), VGG16 (Simonyan and Zisserman 2014), and GoogLeNet (Szegedy et al. 2015), replaced the final fully-connected layer with fully-convolutional layers. Although it was quite successful, it lacked global contextual information, instance awareness, and real-time performance. Encoder-decoder variants referred to as semantic pixel-wise segmentation network (SegNet) (Badrinarayanan, Kendall, and Cipolla 2017) and U-Net (Ronneberger, Fischer, and Brox 2015) were also developed, where a set of upsampling and deconvolutional layers preceded the final softmax classifier which is used for predicting pixel-wise class probabilities. Dilated convolutions were used to aggregate multi-scale contextual information (Yu and Koltun 2015; Chen et al., "DeepLab," 2017; Paszke et al. 2016; Wang et al., "Understanding Convolution," 2018). As the convolutional filters operate at a particular scale, and hence the extracted features are limited to that scale, multi-scale networks were developed. The multi-scale networks use multiple networks targeting different scales and fuse the individual outputs to produce the final output (Eigen and Fergus 2015; Bian, Lim, and Zhou 2016; Roy and Todorovic 2016; Zhao et al. 2017). ParseNet (Liu, Rabinovich, and Berg 2015) concatenated global features from a preceding layer to the local features of a succeeding layer. RNNs have also been utilised to capture the short- and long-spatial relationships in ReSeg (Visin et al. 2016) and other works (Pinheiro and Collobert 2014; Byeon et al. 2015). Adversarial networks have also been explored for semantic segmentation (Luc et al. 2016; Souly, Spampinato, and Shah 2017; Hung et al. 2018). The Efficient NN (ENet) (Paszke et al. 2016) developed using dilated convolutions for semantic segmentation offers good real-time performance. It is a promising solution for AR.

4.4.4. Instance segmentation

Instance segmentation is the most advanced state of scene understanding, which requires *differentiating different instances of the same object* in addition to semantic segmentation. Lack of knowledge of the number of instances of a particular object in the scene makes it more difficult than other tasks. It is a partially solved problem and requires significant further research. Successful instance segmentation can offer information about occlusion that can help in AR-assisted precise robotic pick-and-place applications among several other applications.

Simultaneous Detection and Segmentation (SDS) (Hariharan et al. 2014), built upon R-CNN (Girshick et al. 2014), is one of the first significant attempts at instance segmentation. Mask R-CNN (He et al. 2017) added a third output to Faster R-CNN (Ren et al. 2015) to compute the binary mask for segmentation in addition to two other outputs: bounding box offset and class

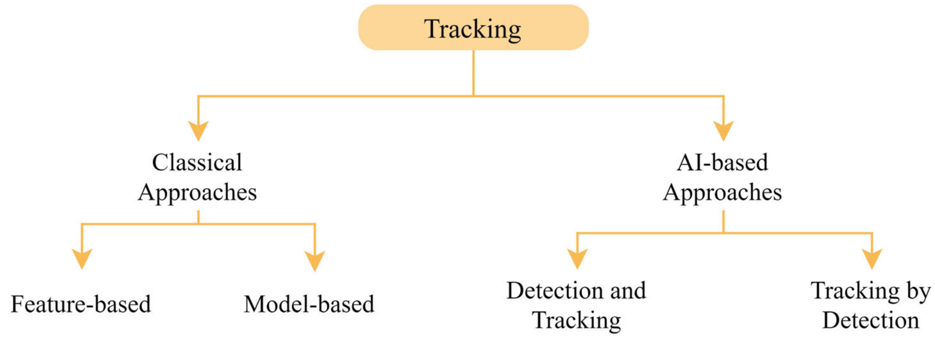


Figure 4. Visual depiction of the topics reviewed in tracking.

label. Mask Labeling (MaskLab) (Chen et al. 2018), based on Faster R-CNN (Ren et al. 2015), produces three outputs: box detection, direction prediction, and semantic segmentation. The direction prediction estimates each pixel's direction towards its center and hence assists in separating instances of the same class. Path Aggregation Network (PANet) (Liu et al. 2018), based on FPN (Lin et al. 2017) and path augmentation, was also developed. Mask R-CNN (He et al. 2017), boundary-aware instance segmentation (Hayder, He, and Salzmann 2017), MaskLab (Chen et al. 2018), and PANet (Liu et al. 2018) produce instances with bounding boxes, not pixel-wise segmentations.

A multi-task network cascade (Dai, He, and Sun 2016) consisting of three networks was developed for instance-aware semantic segmentation. The three networks were dedicated to differentiating instances, categorising obje cts, and estimating masks. In addition to multi-task network cascade, TensorMask (Chen et al. 2019), DeepMask (Pinheiro, Collobert, and Dollár 2015), and SharpMask (Pinheiro et al. 2016) predict pixel-wise instance segmentation.

4.5. Conclusion

Object detection and its allied tasks have seen unprecedented success with the advent of DL. Their performance has surpassed that of classical feature-based approaches. Currently, object detection is mostly considered to be a solved problem because of DL. However, the object detection task in AR-assisted manufacturing applications is still mostly dominated by fiducial markers. Moreover, most of the existing work using DL techniques in AR-assisted manufacturing limit DL for object detection for simpler augmentations like rectangular bounding boxes to localise (Abdi and Meddeb 2017; Subakti and Jiang 2018) or annotate (Rao et al. 2017; Park et al. 2020) the identified objects. Advanced pixel level and instance-aware techniques are yet to be adopted into AR-assisted manufacturing applications.

5. Tracking

Detection can lead to an estimation of the pose of the camera in a static scene. In reality, no interactive environment is static. Even minute movements of the camera alter the viewpoint and disturb the alignment between the virtual and real content. Objects in the scene can also be under motion. So, tracking is essential to estimate the pose of the camera in real-time. An effective tracker needs to be accurate within fractions of a millimetre in distance and degrees in orientation. It has to be quick enough to maintain temporal coherency. The combined latency of the tracker and the renderer has to be low enough to maintain sensible registration. Figure 4 depicts the organisation of the discussed tracking techniques.

5.1. Current approaches for tracking in AR

The classical approaches of tracking can be broadly be split into *feature-based* and *model-based*. The *feature-based* approaches use feature descriptors of the current images of the object to match it with the reference image. Techniques like mean shift (Comaniciu, Ramesh, and Meer 2000; Comaniciu and Meer 2002; Zhou, Yuan, and Shi 2009), keypoint tracking by matching the point neighbourhood (Lowe 2004), homography (Benhimane and Malis 2007), differential tracking like the Lucas-Kanade (LK) algorithm (Lucas and Kanade 1981) etc. are used to match the feature descriptors. However, such feature-based methods are not effective under occlusions and changes in illumination. Sum of conditional covariance (Rogério et al. 2011), M-estimation (Comport, Marchand, and Chaumette 2003), local normalised correlation (Irani and Anandan 1998), and maximisation of mutual information (Viola and Wells 1997; Panin and Knoll 2008; Dame and Marchand 2010) are a few techniques to increase the robustness of feature-based methods.

Virtual visual servoing (VVS) with M-estimator was used for real-time tracking of circles, lines, and spheres in a monocular video see-through (VST) system (Comport, Marchand, and Chaumette 2003). Visual servoing with Lie algebra of affine transformations (Drummond and Cipolla 1999) was used in a model-based tracker that used edge features (Reitmayr and Drummond 2006). Multi-stage optical flow estimation (Neumann and You 1998) was used to track both fiducials (colored circles and triangles (Neumann and Cho 1996)) and natural features (e.g. textures and corners) (Park, You, and Neumann 1999). Multiple tracking hypotheses based on the moving edges algorithm (Bouthemy 1989) were retained for robust estimation in Vacchetti, Lepetit, and Fua (2004) and Wuest, Vial, and Strieker (2005), instead of one like those in Drummond and Cipolla (2002) and Marchand, Bouthemy, and Chaumette (2001). The tracking algorithm discussed by Vacchetti, Lepetit, and Fua (2004) can track both textured and non-textured objects as they combine both edges and feature points. Point features extracted using Features from Accelerated Segment Test (FAST) corner detector were tracked by tracking-by-synthesis approach (Simon 2011), where the points were matched using normalised cross-correlation. Real-time Attitude and Position Determination (RAPiD) (Harris 1993) was a popular markerless 3D tracking system.

The *model-based* approaches use a 3D model of the object (Lowe 1991; Drummond and Cipolla 2002; Comport et al. 2006). From a bird's eye view, the model-based approaches minimise the error between the features extracted from the current image and the projection of the 3D model. Ordinarily, they minimise the distance between the points of the image and the projection of the 3D lines (Uchiyama and Marchand 2012). Edges are used widely due to their computational efficiency and robustness against changes in lighting. Both non-linear optimisation methods (e.g. VVS (Comport et al. 2006), iterative reweighted least squares (Drummond and Cipolla 2002)) and Bayesian methods (e.g. particle filter (Klein and Murray 2006; Pupilli and Calway 2006)) were used for solving the minimisation problem.

Multiple objects were tracked using image retrieval and online SfM (Kim, Lepetit, and Woo 2010). SLAM (Bailey and Durrant-Whyte 2006; Durrant-Whyte and Bailey 2006) has been the most significant breakthrough in tracking applications. It can track the location of the agent (e.g. robot, camera) in unknown environments, in addition to creating a map of the environment using natural features. Its widespread adoption can be attributed to the elimination of the requirement of prior information like maps or Computer-Aided Design (CAD) models. SLAM was used for annotating unknown environments (Reitmayr, Eade, and Drummond 2007; Reitmayr et al. 2010) in addition to tracking. It was also used in wearable and handheld cameras (Castle et al. 2007). It was further developed as parallel tracking and mapping (PTAM) (Klein and Murray 2007) and Oriented FAST and Rotated Binary Robust Independent Elementary Features (BRIEF) (ORB)-SLAM (Mur-Artal, Montiel, and Tardos 2015) for various applications.

5.2. Limitations of current approaches for tracking in AR & potential for AI

The optimisation methods necessary for tracking are prone to failure in cluttered environments, which hinders the matching process. In Bayesian approaches, ensuring a fair representation of all possible candidate poses is a challenging task due to the huge number of possible poses.

As a whole, the classical approaches of tracking have the following shortcomings:

- Complex scenes involving multiple dynamic objects, variations in illumination, and motion blur are difficult to model. Moreover, the tracked objects may undergo partial or full occlusion. Capturing the variations and impacts of occlusion still remains a hurdle in traditional approaches of tracking.
- The combinatorial complexity of traditional Multiple Object Tracking (MOT) algorithms (e.g. multiple hypothesis tracking (MHT) (Reid 1979) and joint probabilistic data association filters (Bar-Shalom, Fortmann, and Cable 1990; Hamid Rezatofighi et al. 2015)) explode exponentially as the number of tracked objects increases. So, their complexity renders them ineffective for real-time tracking of multiple objects.

The elimination of hand-crafted features to design trackers that are robust against occlusion and variations in the object attributes makes DL-based techniques better than the conventional ones. Moreover, the complexity of the DL-based trackers is not proportional to the number of tracked objects in the scene. Together, these advantages make DL-based trackers a wise choice for AR applications.

5.3. AI approaches for tracking in AR

To the best of our knowledge, there is not a single research work that uses AI in the tracking component of AR-assisted manufacturing. The only instance of AR that uses AI-based tracking is a DCNN. The DCNN was used for temporal tracking of the 6-Degrees of Freedom (DoF) pose of an object for superimposing the virtual model of the object on the real object using two frames of inputs (Lalonde 2018). It was more resistant to occlusions. DL-based tracking techniques have not received the deserved attention. This opens up the possibility for radical applications of DL-based trackers for adoption into AR-assisted manufacturing applications.

5.4. Potential AI approaches for tracking in AR

From a functional viewpoint, we split the tracking methods into two parts: single object tracking (SOT) and MOT, depending on the number of objects being tracked. Although tracking a single object is sufficient to estimate the camera pose, SOTs are prone to failure in case of occlusion or insignificant relative motion between the tracked object and the camera. In the case of occlusion, the camera pose can be estimated by tracking other objects that are not occluded. That makes MOTs a better choice. From a technical perspective, the tracking approaches can be categorised into *detection and tracking* (Wagner et al. 2008; Pilet and Saito 2010) and *tracking by detection* (Lepetit and Fua 2006; Ozuysal et al. 2009). In *detection and tracking*, the object in the current frame is detected. Its instances in the future frames are tracked using a tracking algorithm. *Tracking by detection* detects the object in every frame and tracks it using association techniques. It does not utilise the location information from the previous frames. As the *detection and tracking* methods utilise the information from the past frames, they are computationally inexpensive and are quicker than *tracking by detection*-based methods. However, *detection and tracking* methods are prone to failure in the presence of multiple similar objects, or in cases which affect the visibility of the tracked object (e.g. under occlusion or instances where the tracked object moves beyond the Field Of View (FOV) of the camera). Due to the presence of numerous unknown objects, which calls for extraction of unknown features, MOT tasks are heavily dominated by *tracking by detection* approaches. Most of the SOT tasks use *detection and tracking* approaches as the initial bounding box for the target object can be prescribed or estimated.

5.4.1. Single object tracking

Discriminative correlation filter (DCF) based approaches, which exploit properties of circular correlations, used to dominate most of the tracking benchmarks in terms of accuracy and robustness. Further improvements of DCF-based approaches for increasing the tracking speed reduced the accuracy significantly (Bolme et al. 2010). So, pre-trained CNNs were appended to increase their performance. DCF-based methods needed large models and had a huge number of parameters. Thus, they were prone to overfitting. Efficient Convolution Operators (ECO) (Danelljan et al. 2017) introduced factorised convolution operators to reduce the number of parameters, a conservative update strategy to increase robustness, and a compact generative model to reduce complexity. Multi-Domain Network (MDNet) (Nam and Han 2016) and structured output deep learning tracker (SO-DLT) (Wang et al., "Transferring Rich Feature," 2015) are a few early DL methods to utilise CNNs for tracking. They were limited by real-time performance. Successive networks like DeepSRDCF (Spatially Regularised Discriminative Correlation Filters) (Danelljan et al. 2015) and fully convolutional network based tracker (FCNT) (Wang et al., "Visual Tracking," 2015) used shallow networks to process pre-trained CNNs as features to improve their performance. The first fully convolutional Siamese network for tracking, SiamFC, was powered by AlexNet (Krizhevsky, Sutskever, and Hinton 2012) and had a cosine window for temporal constraints (Bertinetto et al. 2016). SiamFC had poor generalisation capability and its performance degraded if the target experienced significant changes to its appearance. Semantic appearance Siamese network (SA-Siam) (He et al. 2018b), inspired from ensemble trackers, branched into an *appearance* branch and a *semantic* branch. Each of the branches was a Siamese network. The networks were trained separately and were fused after computing the similarity scores. It offered comparatively better real-time performance.

Most of the prior Siamese networks like SiamFC (Bertinetto et al. 2016), Siamese INstance search Tracker (SINT) (Tao, Gavves, and Smeulders 2016), SA-Siam (He et al. 2018b), and Siamese RPN (SiamRPN) (Li et al. 2018) used shallow networks like AlexNet (Krizhevsky, Sutske ver, and Hinton 2012). Despite replacing the shallower AlexNet foundation with deeper networks like ResNet (He et al. 2016) or GoogLeNet (Szegedy et al. 2015), the performance of the network did not improve significantly. The authors of SiamFC+, SiamRPN+ (Zhang and Peng 2019), and SiamRPN++ (Li et al., "Siamrpn++," 2019) identified two reasons for the limited performance: (1) deeper networks restricted the translation invariance; and (2) the network padding used in deeper networks induced a positional bias in the learning. It hindered the object relocalization. SiamFC+ and SiamRPN+ (Zhang and Peng 2019) were equipped with a set of cropping-inside residual (CIR) units, and those CIR units were stacked to create deeper and wider networks. SiamRPN++ (Li et al., "Siamrpn++," 2019) was trained using a spatially aware sampling strategy to facilitate spatial invariance. It incorporated a multi-layer aggregation module and a depth-wise correlation layer on top of SiamRPN (Li et al. 2018). SiamRPN++ (Li et al., "Siamrpn++," 2019) could achieve 35 *FPS* with a ResNet-50 (He et al. 2016) backbone and 70 *FPS* with a

MobileNet (Howard et al. 2017) backbone. The Siamese networks were further developed as a Distractor-aware Siamese RPN (DaSiamRPN) (Zhu et al. 2018) for improving robustness against occlusion and clutter, memory networks (Yang and Chan 2018) using Long Short-Term Memory networks (LSTMs) to capture the variation of the target's appearance, the Siamese better match network (Siam-BM) (He et al. 2018a) to handle rotations, Dynamic Siamese network (DSiam) (Guo et al. 2017) to learn background suppression and variation in the target appearance, and Residual Attentional Siamese Network (RASNet) (Wang et al., "Learning Attentions," 2018) to enhance discriminative power and reduce overfitting.

The networks like SiamFC (Bertinetto et al. 2016), SA-Siam (He et al. 2018b), and SiamRPN (Li et al. 2018) track the target object using axis-aligned rectangular bounding boxes. Although the rectangular bounding box is efficient, it is far from an accurate representation of the object. Moreover, for AR applications where the pose of the object is necessary, a rectangular bounding box does not meet the necessity. SiamMask (Wang et al. 2019) produced both rotating bounding boxes and object segmentation masks at 55 *FPS* by including two more tasks of bounding box regression using an RPN (Ren et al. 2015; Li et al. 2018) and class agnostic binary segmentation (Pinheiro, Collobert, and Dollár 2015) in addition to the usual similarity metric of the Siamese networks. More importantly, SiamMask can estimate the binary segmentation masks from a boundary box initialisation. SiamMask_E (Chen and Tsotsos 2019) improved the bounding box fitting process by fitting an ellipse to estimate the size and angle of rotation of the bounding box.

Pure DL-based algorithms still have a long way to go to meet the expected performance in tracking problems. They are not as successful as they are in object detection. Although Siamese networks are quite promising and are successful, they could not fully own the stage of visual object tracking challenge (VOTC) 2018 (Kristan et al. 2018) and 2019 (Kristan et al. 2019). Siamese networks shared the stage of VOTC 2019 with algorithms based on DCFs for *short-term*,[2] *long-term*[3] as well as *real-time*[4] tracking. Most of the successful trackers in VOTC 2019 used CNNs for object localisation. Instead of classical DCFs, CNN-based formulation of DCFs was widely used. The successful Siamese trackers used Siamese correlation filters. Many networks combined characteristics of both Siamese networks and DCFs. Unlike VOTC 2018, the best VOTC 2019 trackers (except FuCoLot in *long-term* tracking) did not use hand-crafted features. The top trackers relied on a deeper ResNet backbone. The future trackers seem to be deeper hybrid networks utilising concepts from both Siamese networks and DCFs.

5.4.2. Multiple object tracking

MOT algorithms have to detect and track multiple objects without prior information about the number of target objects and their appearance. The detected objects in MOT have to be tagged for proper identification. In addition to occlusion and interaction among the objects, inter-class similarity and intra-class variation make the MOT task quite challenging. Most DL-based MOT algorithms use the *tracking by detection* approach because of the number of unknown objects. In a *tracking by detection* framework, there are two tasks: detecting the moving object in every frame and tracking them by associating the instances of the same object in each frame. So, holistically, tracking is a spatio-temporal task.

Before the advent of DL techniques in object detection, the detection task was the bottleneck of tracking applications. After the success of DL techniques in object detection, many trackers used state-of-the-art DL-based detectors to detect multiple objects. So, simple online and real-time tracking (SORT) (Bewley et al. 2016) included faster R-CNN (Ren et al. 2015) for detecting multiple objects. It employed Kalman filters (Kalman 1960) for motion prediction and the Hungarian method (Kuhn 1955) for data association. To capture the temporal information of the videos, the spatial features extracted by SSD (Liu et al. 2016) were fed to LSTMs for regressing and classifying locations of the objects in addition to associating the features of the objects across the frames (Lu, Lu, and Tang 2017). Kieritz, Hubner, and Arens (2018) jointly detected and tracked multiple objects using an integrated network of SSD and RNNs. An end-to-end trainable enhanced Siamese network (Kim, Alletto, and Rigazio 2016) was used to extract appearance and temporal geometric information for supporting tracking.

In addition to detection, *data association* is an important task for tracking applications. Occlusion, drastic variation in appearance or motion, and interactions of the objects impede data association. Failure in data association leads to fragmentation of the tracks, misidentification of the objects, etc. Traditionally, clique graphs (Zamir, Dehghan, and Shah 2012), global optimisation (Zhang, Li, and Nevatia 2008; Berclaz et al. 2011), discrete-continuous optimisation (Andriyenko, Schindler, and Roth 2012), object re-identification (Kuo and Nevatia 2011) etc. were used for data association. These traditional approaches are of discrete combinatorial complexity. So, to improve the data association in the case of MHT, MHT was combined with bi-linear LSTMs (Kim, Li, and Rehg 2018) to incorporate motion and appearance models and LSTM-CNNs (Chen, Peng, and Ren 2018) to handle long temporal variations in the object. Further, only RNNs were utilised in three blocks for motion prediction, state update, and track management (initiation and termination) for online MOT (Milan et al. 2017). Although it was quicker ($\approx$165 *FPS*), its accuracy was sub-par. Multiple LSTM models were used for learning position, velocity, and appearance features in MOT (Liang and Zhou 2018). However, LSTMs are known to fail at fully capturing the temporal information across the frames, as tracking is a spatio-temporal task. So, leveraging both CNNs and LSTMs is quite intuitive. Thus, an RNN-based appearance model built upon a CNN was used for MOT (Sadeghian, Alahi, and Savarese 2017). The appearance model was supplemented with motion and object interaction models based on LSTMs. State-of-the-art Siamese networks were combined with LSTMs to capture the appearance, motion, and velocity cues for MOT (Wan, Wang, and Zhou 2018). The tracks were initialised using traditional methods (Kalman filters (Kalman 1960) with Hungarian methods (Kuhn 1955) and LK optical flow (Lucas and Kanade 1981) with Intersection-over-Union (IoU) distance computation) for feeding into the LSTMs.

The data association task can also be solved by estimating a similarity score among different frames by using CNNs only. Siamese networks, which are highly successful in SOT, have become the de facto choice to compute the similarity score. A Siamese CNN and an appearance-based tracklet affinity model were jointly learned for increasing the robustness against track fragmentation and identity switching (Wang et al. 2016). Spatio-temporal descriptors learned from a Siamese CNN were combined with gradient boosting to associate the image patches (Leal-Taixé, Canton-Ferrer, and Schindler 2016). It was validated using linear programming. Spatio-temporal motion features were added to a Feature Pyramid Siamese Network (FPSN) to supplement the similarity in appearance information of the Siamese networks with motion information (Lee and Kim 2018). Additionally, reinforcement learning (Ren et al. 2018; Rosello and Kochenderfer 2018) and autoencoders (AEs) (Wang et al. 2014) have been explored for MOT applications.

5.5. Conclusion

Most of the MOT tasks focus on tracking vehicles (in autonomous driving applications e.g. KITTI (Geiger, Lenz, and Urtasun 2012; Geiger et al. 2013), traffic scenes (Wen et al. 2015)), or people[5] (pedestrians in crowded places or on the street). AR applications, in general, and in manufacturing applications, focus on generic MOT. So, generic MOT still needs considerable focus as the premise of the two problems are quite different. Trackers for tracking pedestrians or vehicles have to mostly capture the intra-class variations, which need not be the case in generic MOT. Generic MOT lays almost equal emphasis on both intra- and inter-class variations. The SOT datasets include tracking of generic objects (e.g. VOTC 2020[6], object tracking benchmark (Wu, Lim, and Yang 2013)). The most important reason for the lack of attention in generic MOT is the unavailability of germane datasets; the first generic MOT dataset was developed in 2020 (Dave et al. 2020) with labeled bounding boxes. We expect generic MOT to gain significant momentum in the future. That will expedite the development of AR tracking technologies in general and in manufacturing applications. Additionally, state-of-the-art SOTs like SiamMask (Wang et al. 2019) and SiamMask_E (Chen and Tsotsos 2019) can track segmentation masks, whereas MOT methods are still exclusively using axis-aligned rectangular bounding boxes to track objects.

6. Camera pose estimation

The ultimate goal of both detection and tracking is to estimate the camera pose. Both detection and tracking produce information about the pose of the object, from which the pose of the camera can be estimated. The transformation from the pose of the object to the camera is termed *pose estimation*. In general, *pose estimation* is the process of estimation of the position and orientation of the camera from the scene. Pose estimation is different from visual place recognition in the sense that visual place recognition offers the topological location without the exact position, whereas pose estimation offers both global position and orientation of the camera. A precise 6-DoF pose of the camera is essential for a realistic blend of virtual content. The pose estimation techniques can be broadly categorised as shown in Figure 5.

6.1. Current approaches for camera pose estimation in AR

As the camera pose consists of six independent parameters, ideally, three points (P3P - Pose estimation from 3 points) should suffice. However, the solution involves a quartic polynomial. That calls for an additional point to disambiguate the results. However, multiple point correspondence (P*n*P - pose estimation from n points) is preferred, as the accuracy of the estimated pose increases with the number of point correspondences. The over-constrained pose estimation problem gets transformed into an optimisation problem. Techniques like DLT and pose from orthography and scaling with iterations (POSIT) (Dementhon and Davis 1995) can solve the optimisation problem. DLT can find the complete projection matrix, which consists of both camera intrinsics

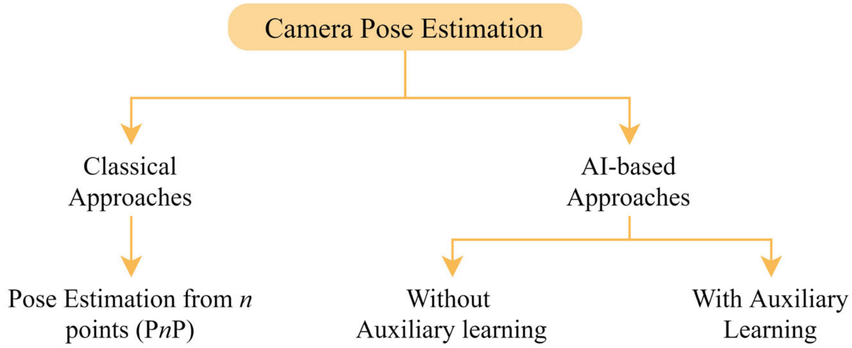


Figure 5. Visual depiction of the topics reviewed in camera pose estimation.

and extrinsics, by solving the linear system of equations using Singular Value Decomposition (SVD). However, estimating only the extrinsics yields better results in 3D tracking. If the internal parameters are known, the formulation becomes over-parametrised; this reduces stability. POSIT is an iterative process, which has become a standard method to solve the PnP problem. Although it does not need any initialisation, it is sensitive to noise and is unsuitable for coplanar points. Homography (Benhimane and Malis 2007) can be used for estimating the pose from coplanar points. In the case of noisy data, the camera pose is estimated by minimising the sum of reprojection errors. That requires an iterative optimisation scheme, which again needs an initialisation. If the total error is linear, linear least squares or iterative re-weighted least squares can be used. If the total reprojection error is non-linear, Gauss-Newton or LM algorithm can be employed. The pose increment in LM includes an additional stability factor. If the stability factor is small, LM imitates the Gauss-Newton algorithm. If the stability factor is large, the LM behaves like gradient descent (Lepetit and Fua 2005). As pose estimators face a large number of outliers from the images, Random Sample Consensus (RANSAC) (Fischler and Bolles 1981), Progressive Sample Consensus (PROSAC) (Chum and Matas 2005), or M-estimators (like Huber estimator (Huber 1992), Tukey estimator (Rousseeuw and Leroy 2005)) can be used for improving the robustness. M-estimators need an initial pose estimate, whereas RANSAC lacks precision. Additionally, Bayesian techniques like Kalman filters and particle filters can also be utilised for pose estimation.

Numerous works used classical methods to estimate the pose of the camera in AR applications. A particle filtering framework was used for estimating the pose of a handheld camera for real-time AR applications (Ababsa and Mallem 2011). It tracked both point and line features. Extended Kalman filter (EKF) was employed to estimate the camera pose from square-shaped fiducials (Maidi et al. 2010). Four points were matched to estimate the initial pose. Orthogonal iteration algorithm (Lu, Hager, and Mjolsness 2000) was utilised to estimate the camera pose from square fiducial markers (Ababsa and Mallem 2004). POSIT (Dementhon and Davis 1995) was used to calculate the pose of a camera from binocular images and retro-reflective markers in a stereo-vision system (Dorfmüller 1999). POSIT for coplanar points (Oberkampf, DeMenthon, and Davis 1996) was used along with RANSAC for determining the pose of the camera in a markerless mobile AR system for geovisualisation (Schall et al. 2008). POSIT was also used by Bleser, Pastarmov, and Stricker (2005) with RANSAC, and with Tukey estimator by Lepetit et al. (2003) for AR applications. A non-iterative PnP algorithm (Moreno-Noguer, Lepetit, and Fua 2007) with RANSAC was utilised for approximating the camera pose from keyframes (Park, Lepetit, and Woo 2008). P3P with RANSAC was used to find the 3D camera pose from natural features in an HMD. P5P was used with RANSAC to estimate the camera pose relative to the user's hand in a tabletop AR setup (Lee and Hollerer 2008).

6.2. *Limitations of current approaches for camera pose estimation in AR & potential for AI*

The classical approaches are plagued by a few limitations:

- Despite the popularity of classical approaches, they lack robustness. Large changes in viewpoint (e.g. at corners of buildings/machines), variation in illumination, textureless scenes, motion blur, etc. deteriorates their performance. They are unable to find accurate matching points in all scenarios.
- The computational complexity of the SfM based approaches grows with the size of the model. With larger models, the proportion of outliers increases considerably. Although RANSAC-like algorithms can be used to increase robustness, computational expense shoots high in the presence of a large number of outliers. Moreover, RANSAC does not have a constant runtime at inference.

On the other hand, as DL-based techniques do not require engineered features, and they can be trained to be robust against changes in illumination, perspective, and others. The complexity of DL techniques remains relatively constant. Unlike RANSAC, the performance of DL remains consistent at runtime. Additionally, Kendall, Grimes, and Cipolla (2015) showed that monocular images have a one-to-one (injective) relationship with the pose of the camera. So, while the classical approaches have to follow a sequential approach of detection, tracking, and pose estimation, DL can directly infer the pose from the RGB images by skipping the detection and tracking tasks.

6.3. AI approaches for camera pose estimation in AR

To the best of our knowledge, there is not a single instance of research work that uses AI to estimate the pose of the camera in AR-assisted manufacturing applications. The only work that used AI trained a CNN using RGB images for fine-tuning the coarse pose acquired from a consumer-grade Global Positioning System (GPS)/Inertial Measurement Unit (IMU) (Wang et al., "DeLS-3D," 2018). They used RNNs to learn the temporal correlation of the pose. To the best of our knowledge, neither their work nor any other work in AR used RGB images for directly estimating the pose from images.

6.4. Potential AI approaches for camera pose estimation in AR

PoseNet (Kendall, Grimes, and Cipolla 2015) conceived the idea of directly estimating the 6-DoF pose including the depth and out-of-plane rotation of a monocular camera in real-time directly from a single RGB image by using a pre-trained GoogLeNet (Szegedy et al. 2015). For training, it used SfM to label the camera pose for each frame in the video. It was robust to variations in lighting, motion blur, and camera intrinsics as it localised the camera pose from deep features. Moreover, it was fast (5 ms/frame) and had a low memory footprint. However, its accuracy in both outdoor (2 m and 6°) and indoor (0.5m and 10°) settings was poorer than feature-based state-of-the-art methods. Efforts at increasing the accuracy of PoseNet included incorporating a geometric loss function (Kendall and Cipolla 2017) and a Bayesian CNN (MacKay 1992) to model the uncertainty (Kendall and Cipolla 2016). PoseNet was prone to overfitting. Although it used dropout regularisation to overcome overfitting, its accuracy did not improve. So, the final 2048-dimensional fully connected layer of the CNN was fed to an LSTM to reduce the dimensionality of the feature vector (Walch et al. 2017). Additionally, alternative architectures like VGG-Net (Naseer and Burgard 2017) and BranchNet (Wu, Ma, and Hu 2017) that split the GoogLeNet into two branches for learning the high-level features of translation and orientation have also attempted to estimate the camera pose.

Even after several architectural modifications, the DL-methods were not able to achieve performance comparable to the classical methods. The reason was identified as the limited capacity of the model to learn the 3D structural connections and constraints from monocular images. So, auxiliary learning methods were employed by Valada, Radwan, and Burgard (2018) to improve camera localisation. The primary task of camera localisation was supplemented with the auxiliary task of visual odometry in visual localisation and odometry network (VLocNet) (Valada, Radwan, and Burgard 2018). It used both a global pose regression network and a Siamese (based on ResNet (He et al. 2016)) relative pose estimation network. VLocNet++ (Radwan, Valada, and Burgard 2018) learned the odometry and semantic segmentation of the scene as auxiliary tasks.

The absolute pose of the camera is dependent on the global coordinate frame. Constructing a generic DL model to learn the camera pose from all scenes is challenging, as every scene has a different coordinate system. A model trained on a particular scene learns the mapping from that particular scene to the camera pose. Altering the scene alters the training data. Hence, the trained model is likely to fail in other scenes. That necessitates re-training of the models with new scenes. So, there has been a considerable thrust to learn the relative pose of the camera to decouple the coordinate frame of the scene from the task of pose estimation. Efforts at estimating the relative pose of sample images with respect to reference images include Siamese CNNs (Laskar et al. 2017; Balntas, Li, and Prisacariu 2018) and hourglass networks (Melekhov et al. 2017; a symmetric encoder-decoder structure based on ResNet (He et al. 2016)).

Additionally, in structure-based learning techniques like Hierarchical Feature Network (HF-Net) (Sarlin et al. 2019), which supplements RGB images with a 3D model of the scene for camera localisation, the best hypothesis was selected using RANSAC (Fischler and Bolles 1981). As RANSAC (Fischler and Bolles 1981) is not differentiable, it prohibited end-to-end learning. DSAC (differentiable RANSAC, Brachmann et al. 2017) substituted the hard hypothesis selection in RANSAC with a probabilistic one and enabled end-to-end learning of the camera pose. DSAC used two CNNs, one for predicting the scene coordinates and creating a pool of hypotheses, and another *scoring CNN* for scoring

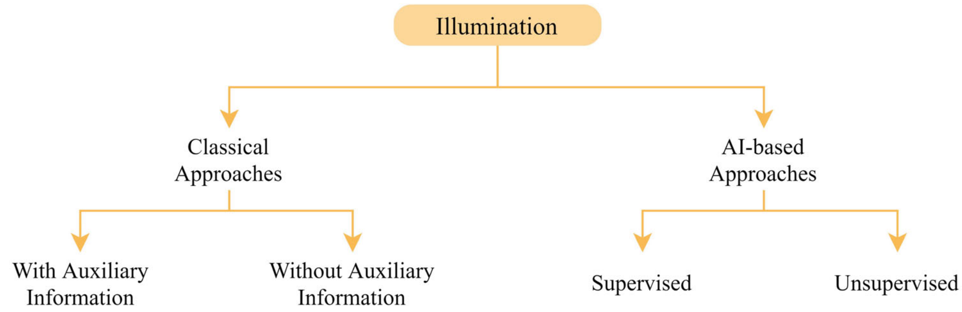


Figure 6. Schematic diagram of the topics reviewed in illumination.

the hypotheses for global consensus. Brachmann and Rother (2018) replaced the scoring CNN with a soft inlier count to reduce overfitting. Moreover, Brachmann and Rother (2018) also showed that RGB-images with ground truth pose are enough for regressing the scene coordinates. The 3D model of the scene is optional.

6.5. Conclusion

DL-based techniques cannot be used without appropriately labeled data. Acquiring the ground truth pose information is not as easy as it is in detection and tracking. Currently, classical methods are used to acquire the ground truth pose. That makes DL-techniques dependent on classical methods. Moreover, as the data collection process is expensive, the pose labels in the training data are usually sparse. Techniques based on odometry accumulate drift, thereby deteriorating the accuracy of the pose estimates over time. Although most DL-based techniques are pre-trained on ImageNet-like datasets, it is important to keep in mind that pose estimation and image classification are largely unrelated tasks. That limits the effectiveness of transfer learning. So far, DL-based techniques could not achieve accuracy at par with classical structure-based methods. Sattler et al. (2019) argued that the reason behind the limited accuracy is that the CNN-based absolute pose regression models are close to image retrieval which are ordinarily used for place recognition, not pose estimation.

7. Inverse rendering

An image is formed by integrating illumination, scene geometry, camera properties, and surface reflectance. Inverse rendering involves recovering the constituent factors from a rendered image. Thus it is an underconstrained problem. We focus on inferring illumination (photometry) and depth (geometry) characteristics of a scene.

7.1. Illumination

For accurate visual registration and coherence, the illumination characteristics of the scene must be determined. Determining the illumination of a scene from a single image, captured using a low dynamic range (LDR) camera with a limited FOV, is a challenging task because of the fact that the scene is illuminated by light received from the full sphere of directions around the object, most of which are beyond the FOV of the camera. Its solution needs modelling of the geometric and photometric information contained in the image. The structure of the reviewed illumination techniques is pictorially depicted in Figure 6.

7.1.1. Current approaches for illumination in AR

Most of the methods for estimating the illumination characteristics of a scene can be broadly classified into those that use auxiliary information and those that do not use auxiliary information. Methods utilising auxiliary information leverage RGB-D data (Jiddi, Robert, and Marchand 2016) or information acquired from light probes or that from other sources. The light probes are quite popular. They can be active (like the fisheye camera used by Kán and Kaufmann 2013) or passive (like reflective spheres used by Debevec 2008). The auxiliary information can also be assumptions of some attributes or simpler models like Lambertian illumination (Lu et al. 2010).

The radiance transfer function was used by Knorr and Kurz (2014) to estimate the lighting on a face, and it was expressed using spherical harmonics. A hemispherical photon emission model was used to estimate the global illumination with the presumption of uniformly distributed multiple light sources (Zhang, Zhao, and Wang 2019).

Methods that do not use any auxiliary information attempt to directly estimate the lighting of the scene from various features contained in the image itself. Shadows (Panagopoulos et al. 2011; Jiddi, Robert, and Marchand 2017) and faces (Shim 2012; Yi et al. 2018) are widely used. Kán, Unterguggenberger, and Kaufmann

(2015) acquired the reflection maps of the virtual objects by convolving the image using bidirectional reflectance distribution functions (BRDFs). Sato, Sato, and Ikeuchi (2003) estimated the distribution of light by exploiting the correlation between the brightness of the image with that of shadows. Gruber, Richter-Trummer, and Schmalstieg (2012) estimated the lighting conditions of the scene using arbitrary geometry for photometric registration.

7.1.2. Limitations of current approaches for illumination in AR & potential for AI

Although the light probes and other methods using auxiliary information offer better visual fidelity, the capturing process is complex, and the whole process is computationally expensive and not scalable. Using light probes to infer the illumination characteristics disturbs the real-time performance. The accuracy of the methods that do not use auxiliary information is low. Additional preprocessing steps like depth estimation, reflectance estimation, and scene geometry reconstruction are employed for improvement. On the other hand, methods based on AI do not need the definitive models of the geometric or photometric attributes. Moreover, they exploit the attributes of the images within the FOV rather than those beyond the FOV, which indirectly affect the scene.

7.1.3. AI approaches for illumination in AR

None of the existing works that use AI for estimating the photometric characteristics are applied to manufacturing applications. There are several research works that use AI to estimate illumination characteristics in AR applications. An AE, with the encoder based on MobileNetV2 (Sandler et al. 2018), was used for estimating the omni-directional illumination in indoor and outdoor environments from an unconstrained LDR image acquired from the limited FOV camera of a mobile AR device (LeGendre et al. 2019). A Generative Adversarial Network (GAN)-based network (Goodfellow et al. 2014) was employed for learning the perspective independent illumination and shadow characteristics of virtual objects for blending them realistically in an AR application (Wang, Wang, and Lian 2019). They adjusted the illumination and shadow of virtual objects using pix2pixHD (Wang et al., "High-resolution Image Synthesis," 2018). A ResNet was designed for determining the illumination of the real scene in terms of spherical harmonics for an AR game (Marques, Clua, and Vasconcelos 2018). Marques et al. (2018) used ResNet (He et al. 2016) for estimating the position of the single point light source from a single color image. Lalonde (2018) used a CNN to estimate the high dynamic range (HDR) illumination from a single RGB LDR image of an indoor scene. Deep NNs (DNNs) have also been used for estimating the illumination characteristics in real-time for AR applications (Xu, Li, and Zhang 2020).

7.1.4. Potential AI approaches for illumination in AR

Gardner et al. (2017) initiated the inference of HDR illumination from single limited FOV LDR images using DCNNs in an end-to-end manner. They used LDR and a sphere to approximate the scene geometry, which are weak approximations to the ground truth. A cascade of AEs was used by Li et al., "Inverse Rendering" (2020) for inferring spatially varying illumination, material, and geometry of the scene. Cheng et al. (2018) trained a DCNN to estimate the low-frequency scene illumination, expressed in terms of spherical harmonics, from a pair of photos.

In AR applications, the illumination of the rendered objects has to be adapted for its location in the real scene. So, the illumination of a specific location (Garon et al. 2019) and a pixel (Song and Funkhouser 2019) were estimated using AEs. As the material of the objects affects illumination, Meka et al. (2018) estimated the material of generic objects appearing in a single image using an AE. ARShadowGAN (D. Liu et al. 2020) used a GAN that leveraged an attention mechanism to create realistic shadows of virtual objects. The 3D CNNs used by Srinivasan et al. (2020) showed a way for learning the 3D illumination of a scene without the need for ground truth scene geometry. It used a stereo pair and a spherical panorama for training.

7.1.5. Conclusion

Most of the manufacturing environments are illuminated by multiple sources. So, adopting state-of-the-art AI techniques for determining the illumination characteristics of the scene would be advantageous. Unavailability of large scale data is impeding the full-fledged adoption of AI into AR-assisted manufacturing applications to estimate the illumination characteristics.

7.2. Depth inference

The modus operandi of AR systems is to overlay virtual objects directly on top of the real video feed, thereby occluding the real objects. For spatially sensible augmentations, the virtual objects should appear to be at the same depth as the co-located real objects. The co-location is infeasible without acquiring proper depth information from the image. Depth inference also aids with mutual occlusion. The reviewed techniques for depth inference can be classified, as shown in Figure 7.

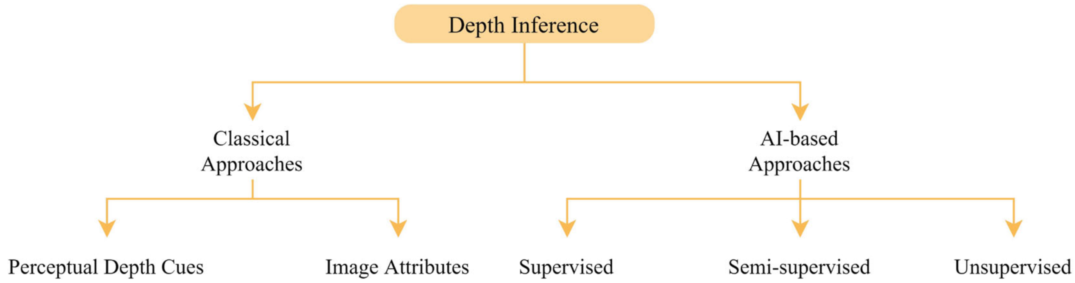


Figure 7. Visual depiction of the topics reviewed in depth inference.

7.2.1. Current approaches for depth inference in AR

The lion's share of AR applications infer depth directly from the scene using the ten perceptual depth cues recognised by Swan et al. (2007). The cues are occlusion, shading, texture gradient, motion parallax, haze, linear perspective and shortening, accommodative focus, height in the visual field, binocular convergence, and binocular disparity. Although they are popular, the perception of every user is different. That leads to inconsistent interpretation. Classical image-based approaches also try to exploit these cues. They are comparatively more consistent as the depth is directly inferred from the video feed. Stereo (Kanade and Okutomi 1994; Zhu and Pan 2008), motion (Weng, Huang, and Ahuja 1989), shadow casting, object shading (Horn and Brooks 1989), and texture (Blostein and Ahuja 1989) have been used for depth inference. Monocular cameras can also be used for depth inference using multi-focus images or acquiring images from different perspectives using a moving camera (Newcombe and Davison 2010). Breen, Rose, and Whitaker (1995) used Z-buffer and Sanches et al. (2012) used frame buffer to acquire depth maps. Hahne and Alexa (2009) combined time-of-flight range data with stereo images to compute depth images at interactive rates. Stereo matching was used to estimate the depth of the real world by projecting the bounding box of the virtual object using the model view matrix (Kanbara et al. 2000). Zhu et al. (2010) combined the depth information acquired from stereo images with color and neighbourhood information to reduce noise in handling occlusion in a video-based AR system. Dense depth maps were acquired using geo-registered images and sparse 3D models acquired from geographic information system (GIS) data using back-projection (Zollmann and Reitmayr 2012). The depth information was calculated by stereo matching after background subtraction and clustering the objects based on their silhouette in video-based AR (Zhu and Pan 2008). Optical flow and sparse SLAM reconstruction were used for inferring pixel-level depth information (Holynski and Kopf 2018). Census correlation (Zabih and Woodfill 1994) was used as the stereo matching algorithm to compute the per-pixel depth map without any prior information (Gordon et al. 2002).

7.2.2. Limitations of current approaches for depth inference in AR & potential for AI

The main reason for the lack of effectiveness of classical methods in recovering depth information from an image is that it is a heavily under-constrained problem. While perceptual cues can aid in recovering relative depth information from images, their efficacy gets deterred by camera characteristics like LDR. Most of the classical methods utilise only a few cues. AI methods, on the other hand, can exploit almost all the cues at once, thereby offering better depth inference.

7.2.3. AI approaches for depth inference in AR

Owing to the limited effectiveness of classical approaches at depth inference, AI-based techniques for inferring depth have been explored in AR applications. A Siamese network and a self-supervised AE based on depth estimation network (Garg et al. 2016) and DeConvNet (Noh, Hong, and Han 2015) was trained for estimating the per-pixel inverse depth (disparity maps) from stereo image pairs without ground truth data in a surgical application (Ye et al. 2017). This is the only instance of AI in AR for inferring depth information. It is not applied within manufacturing. We briefly discuss other state-of-the-art DL techniques that are not being used in AR but can be useful for depth inference in AR-assisted manufacturing applications.

7.2.4. Potential AI approaches for depth inference in AR

Eigen, Puhrsch, and Fergus (2014) were the first to estimate the depth from monocular single images using two CNNs, one for coarse global estimation and the other for local refinement, in an end-to-end manner. Frameworks based on AlexNet and VGG-Net were used for predicting the depth in addition to semantic labeling and estimating the surface normals (Eigen and Fergus 2015). A ResNet backbone with reversed Huber loss (Zwald and

Lambert-Lacroix 2012) has also been used for predicting depth from single RGB images (Laina et al. 2016). An encoder-decoder framework, inspired from SegNet (Badrinarayanan, Kendall, and Cipolla 2017) and FlowNet (Dosovitskiy et al. 2015) was designed for detecting obstacles from images and corresponding optical flow (Mancini et al. 2016). These works use absolute depth information, which is ordinarily difficult to acquire. So, a multi-scale deep network modified from GoogLeNet (Szegedy et al. 2015) was developed for inferring the pixel-wise depth information from single images using relative depth information (Chen et al. 2016).

Additionally, Geometry and Context Network (GC-Net) (Kendall et al. 2017) used 3D convolutions on stereo images. Camera-Aware Multi-scale Convolutions (CAM-Convs) (Facil et al. 2019) incorporated camera models into a U-Net (Ronneberger, Fischer, and Brox 2015) based AE for estimating the depth. Li et al. (2015) and Wang et al., "Towards Unified Depth" (2015) explored Conditional Random Fields (CRFs) for refining the results by exploiting the continuous nature of depth among the pixels. Adversarial networks have also forayed into depth estimation tasks (Gwn Lore et al. 2018; Feng and Gu 2019).

Due to the expensive process of acquisition of ground truth depth information, the geometric correlation between the frames has been exploited to train the model under self-supervision (Casser et al. 2019; Godard et al. 2019). However, due to the lack of ground truth information, the performance of unsupervised learning techniques remains quite low. So, visual odometry (Wang et al., "Learning Depth," 2018) and auxiliary learning tasks (Yin and Shi 2018; Zou, Luo, and Huang 2018) were used to assist depth estimation. Even after several innovations, the performance of unsupervised learning remained low. So, semi-supervised methods using stereo knowledge (Tosi et al. 2019) and sparse ground truth data (Kuznietsov, Stuckler, and Leibe 2017) have also been explored to balance accurate depth estimation with the reliance on ground truth data.

7.2.5. Conclusion

While DL-based techniques can estimate the depth information from RGB images simultaneously with other tasks such as detection, tracking, and pose estimation, the major hurdle that still remains is the acquisition of ground truth data for training. The existing datasets are limited in their capabilities: Cityscapes (Cordts et al. 2016) lack ground truth data; KITTI (Geiger, Lenz, and Urtasun 2012) has sparse ground truth data; and New York University (NYU) Depth dataset (Silberman et al. 2012) is restricted to indoor settings. Although promising, unsupervised and semi-supervised learning still have a long way to go to achieve the desirable performance in order to be applied within the industrial setting.

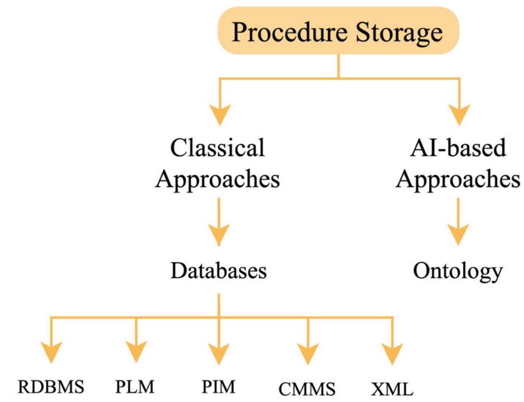


Figure 8. Visual depiction of the topics reviewed in procedure storage.

8. Procedure storage

Procedural information is required for both assembly and maintenance tasks within the manufacturing environment. Having accurate and effective procedures available for use in the AR system is important so that the AR system can be utilised to its fullest capacity. A hierarchical description of the topics covered within procedure storage can be found in Figure 8.

8.1. Current approaches for procedure storage in AR

The most common procedural storage mechanisms in AR systems are databases. Databases store and organise data so that the data can be easily accessed and managed (Tonnis 2003). The most popular types of databases used in AR systems are relational database management systems (DBMS) or RDBMS, PLM systems, product information management (PIM) systems, computerised maintenance management systems (CMMS), and extensible markup language (XML) databases.

An RDBMS facilitates the organisation of relational databases. Relational databases provide structure to data by storing it in tables that comply with database schema (Tonnis 2003). MySQL databases store relational databases on servers (Lytridis, Tsinakos, and Kazanidis 2018). The smart maintenance solution given by Terenzi and Basile (2013) includes two main parts: an administration web client and a mobile user client. The administration web client utilises a server-based

RDBMS to store maintenance and user information and to connect users to maintenance information depending on the environmental context. The AR solution for distance learning given by Lytridis, Tsinakos, and Kazanidis (2018) stores book and asset information in a server-based MySQL database. Their ARTutor platform is able to access the database in order to download images and files. While RDBMS is optimised to perform queries quickly, this optimisation results in varying data structures that are difficult to distribute between systems (Tonnis 2003).

PLM systems store enterprise data (Rentzos et al. 2013). The AR solution given by Setti, Bosetti, and Ragni (2016) for computer numerical control (CNC) machine setup and maintenance utilises a PLM to store system description and diagnostic information for the CNC machine. The PLM database is composed of YAML Ain't Markup Language (YAML) files. The AR solution for manual assembly processes given in Rentzos et al. (2013) stores virtual instructions in a repository through Visual C# .NET files and can be incorporated with an external enterprise database such as PLM in order to gather geometrical information about parts. This geometrical information is utilised to match parts with corresponding equipment or tools that relate to the virtual instructions for assembly.

PIM systems support the management of product and part structure and development (Lee and Rhee 2008b). The solution for AR as a part of ubiquitous computing in cars by Lee and Rhee (2008b) utilises a PIM system to store information about a car for the purposes of maintenance.

CMMS, also referred to as enterprise asset management or EAM systems, store maintenance data (Neges, Wolf, and Abramovici 2015). The solution for mobile maintenance support through AR by Neges, Wolf, and Abramovici (2015) utilises CMMS to manage maintenance data, employees, and tasks.

XML databases store XML documents that are either text-based or model-based (Tonnis 2003). Schmalstieg and Wagner (2007) utilise a server called Muddleware that has three components: (1) a memory-mapped database, (2) a networking component, and (3) a controller component in the form of a finite state machine (FSM). The memory-mapped database is structured using a model-based XML database. The AR navigation system presented by Reitmayr and Schmalstieg (2003) uses an XML database to store domain-specific preprocessing operations. While the XML format is convenient because of its uniform structure, XML databases are not suitable for AR systems because they are resource-intensive (Tonnis 2003).

8.2. Limitations of current approaches for procedure storage in AR & potential for AI

While all of the database structures described are effective in their storage and organisation of data for manufacturing, they are static in nature and do not have the capability to derive further knowledge and relationships based on the data they hold. One possible solution that is able to dynamically manipulate data over time is AI-based storage methods such as ontology.

8.3. AI approaches for procedure storage in AR

An ontology is an AI-based storage method that is able to not only store data in a logical and organised manner but is able to derive relationships between data instances. Ontologies are data storage mechanisms that are a part of the Semantic Web (Crowder et al. 2009) and can be defined as a specification of a conceptualisation (Gruber 1993). One of the benefits of using an ontology to describe domain knowledge as opposed to a database is that semantic queries can be completed in ontologies using SPARQL Protocol and Resource Description Framework (RDF) Query Language (SPARQL) (Zhao and Qian 2017). Ontologies are composed of concepts or classes and instances or individuals where concepts describe categories within a domain, and instances are specifications of these categories or concepts. Ontologies that have instances or individuals defined are deemed knowledge bases (Noy and McGuinness 2000). Many file formats have been created to store ontologies, including RDF, web ontology language (OWL) and its variants, and United States Defense Advanced Research Project Agency (DARPA) agent markup language and the ontology inference layer (DAML + OIL) (Munir and Anjum 2018). The main advantage of using an ontology to store data is the ability to use an ontological or semantic reasoner such as Jena, HermiT, ELK, or Pellet. These reasoners are able to infer logical consequences between ontology instances based on asserted facts or axioms (Abburu 2012). An additional advantage of using an ontology is the ability to integrate various sources of data throughout the PLM (Ali et al. 2019).

Ontologies have been utilised in AR systems in order to store procedural information. Flatt et al. (2015) uses an ontology and knowledge base to store information for their AR system for maintenance applications. Jo et al. (2014) stores Component-Task-Subtask-Instruction pairs in a knowledge-based system (KBS). The AR system is used for maintenance tasks and the stored information is imported from maintenance manuals such as Aircraft Maintenance Manual (AMM), Illustrated Tool and Equipment Manual (TEM), Component Maintenance

Manual (CMM), and Illustrated Parts Catalog (IPC) for aircraft maintenance. Wang, Ong, and Nee (2014) present an integrated AR-assisted assembly environment (IARAE) that includes an ontology-based assembly information model (OAIM) to assist users with AR assembly tasks. During the assembly process, the OAIM is queried for information. Toro, Vaquero, and Posada (2009) used an ontology to store knowledge about industrial maintenance by extending the standard ontology for ubiquitous and pervasive applications (SOUPA) set of ontologies. This maintenance information is combined with AR virtual elements to create a knowledge-based industrial maintenance system structure called UDKE (user, device, knowledge, and experience). Chang, Ong, and Nee (2017) utilise an ontology to store product information in order to form a disassembly sequence table as a part of the automatic content generation module in the AR-guided product disassembly (ARDIS) system. Rentzos et al. (2013) use a semantic-based AR system to integrate product and process in an ontology model that is used to generate work instructions for the user.

8.4. Potential AI approaches for procedure storage in AR

To the best of our knowledge, there are no additional AI strategies other than ontology that can be utilised for procedure storage in AR-assisted manufacturing applications.

8.5. Conclusion

Ontologies or knowledge bases introduce added flexibility into AR systems. If knowledge bases are formed with additional domain knowledge that is available from enterprise management systems and geometric data, such as in Rentzos et al. (2013) and Järvenpää et al. (2019), additional capabilities and automation can be introduced into AR systems that allow for greater integration of multiple sources of data. This utilisation of process information increases the adaptability of AR systems within the manufacturing space.

9. Virtual object creation

Virtual objects are a major component in any AR system. For AR-assisted manufacturing applications, in particular, virtual objects are essential to display data and procedures to users. A hierarchical description of the topics covered within virtual object creation can be found in Figure 9.

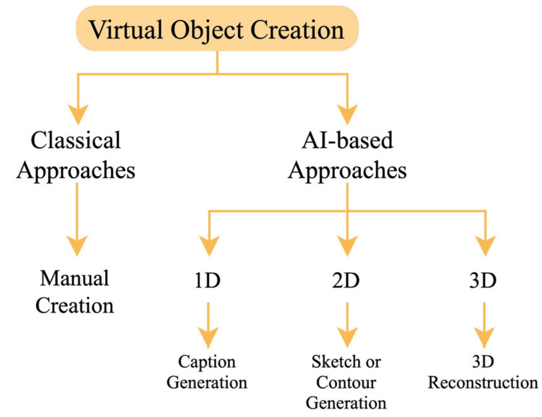


Figure 9. Visual depiction of the topics reviewed in virtual object creation.

9.1. Current approaches for virtual object creation in AR

Virtual objects or augmentations for AR can take the form of one-dimensional (1D) text, 2D graphics or images, and 3D models. The most complex but intuitive of these options are 3D models. An additional layer of interactivity can be added to 3D models with animation. Animation can be incorporated into the system for any type of augmentation (Chang, Ong, and Nee 2017).

In traditional AR systems, virtual objects or augmentations are created manually and stored with respect to tracked objects in the user's environment (Chang, Ong, and Nee 2017). These augmentations are developed offline by experts and loaded into the application before use. The manual virtual object creation process is costly in terms of both time and capital. Even in adaptive AR systems such as the AR Maintenance System (ARMS) by Siew, Ong, and Nee (2019) where augmentations are chosen based on the user's actions, augmentations are predetermined and stored within the system before use.

9.2. Limitations of current approaches for virtual object creation in AR & potential for AI

AI-based methods provide an alternative approach to virtual object creation: *in situ* augmentation creation. In this way, minimal input, such as a single image, can be utilised to automatically create 1D, 2D, and 3D augmentations. This saves the developer of the AR system time and energy in manually creating and predefining these augmentations for use in the system.

9.3. AI approaches for virtual object creation in AR

To the best of our knowledge, there have been no AI-based virtual object creation strategies utilised within AR-assisted manufacturing applications.

9.4. Potential AI approaches for virtual object creation in AR

For AI-based augmentations, there are strategies that can be used to automatically create 1D, 2D, and 3D augmentations for AR systems given minimal input information.

9.4.1. 1D

In order to automatically generate 1D text-based augmentations based on objects seen within a camera frame, AI-based caption generation methods can be utilised. DNN-based image captioning methods can be categorised into (Bai and An 2018):

- DNN-based retrieval and template-based methods;
- Multimodal learning methods;
- Encoder-decoder methods;
- Attention-guided image captioning;
- Compositional architectures; and
- Methods for images with novelties.

The outputs of all DNN-based image captioning methods besides those belonging to the last group of methods are limited to word dictionaries and their training data that comes in the form of image-sentence pairs (Bai and An 2018). Thus, image captioning strategies that deal with novel data (i.e. 'novel captioning') are the most flexible for systems that aim for minimal training time and effort. Mao et al. (2015), Yao et al., "Incorporating Copying Mechanism" (2017), and Li et al., "Pointing Novel Objects" (2019) utilise LSTM for novel captioning. Specifically, Yao et al., "Incorporating Copying Mechanism" (2017) utilise LSTM with a copying mechanism (LSTM-C) as a new architecture to incorporate copying into the CNN plus RNN captioning framework. Li et al., "Pointing Novel Objects" (2019) use LSTM with pointing (LSTM-P) to facilitate vocabulary expansion and produe novel objects via pointing. Hendricks et al. (2016) train a deep multimodal caption model to integrate a lexical classifier and a language model in their Deep Compositional Captioner (DCC) method. Venugopalan et al. (2017) present a Novel Object Captioner (NOC), a semantic captioning model that can produce captions for object categories not present in the original training dataset. Wu et al., "Decoupled Novel" (2018) and Baig et al. (2018) utilise Zero-Shot Learning concepts for the novel captioning task. Finally, Feng et al. (2020) present Cascaded Revision Network (CRN), an approach that integrates out-of-domain knowledge in the novel captioning process.

9.4.2. 2D

For 2D graphics or image-based augmentations, AI-based strategies can be utilised to produce 2D sketches based on inputted images. These sketches can then be utilised on the screen of the AR system as added visual cues for the user. Multiple examples of producing sketches from images can be found in AI literature. Ha and Eck (2017) present Sketch-RNN, an RNN that uses a sequence-to-sequence Variational AE (VAE) to construct stroke-based drawings of common objects. Chen et al., "Sketch-pix2seq," (2017) provide improvements to the Sketch-RNN architecture by using a CNN instead of a bidirectional RNN encoder and by removing the Kullback-Leibler (KL) divergence from the objective function of the VAE. These improvements to the original Sketch-RNN model were found to outperform the state-of-the-art methods with RNN encoders. The method by Song et al. (2018) takes as input one photo of a single object and outputs a sketch of the object. The image is encoded via a CNN and fed into a neural sketcher that composes a generative sequence model with an RNN. The neural sketcher is trained to learn the sketch stroke-by-stroke. Riaz Muhammad et al. (2018) use reinforcement learning to produce a sketch given an input image. Wu et al., "SketchsegNet" (2018) present SketchSegNet, a sequence-to-sequence VAE-based model that implements stroke-level sketch segmentation. Li et al., "Photo-sketching" (2019) propose a contour generation algorithm or *image-conditioned contour generator* that detects salient boundaries in input images and outputs hand-drawn contour drawings based on objects in the input image.

9.4.3. 3D

Some AR systems have already found ways to derive 3D models while the system is in use. All of these methods derive a 3D structure from multiple sequential images. Reitinger, Zach, and Schmalstieg (2007) describe an interactive 3D reconstruction system that is used to model urban scenes. Their 'Reconstruction Engine' takes as input a sequence of 2D images, extracts corners from the image using Harris Corner Detector, finds point correspondences between the images using RANSAC, and finally utilises the plane sweeping principle for depth estimation to create a depth map. Yang et al. (2013) derive reconstructed 3D models from a pair of images by combining multiple techniques: SfM, Clustering Views for Multi-View Stereo (CMVS), Patch-based Multi-View Stereo (PMVS), and Poisson surface reconstruction.

AI-based 3D reconstruction methods can produce 3D models, point clouds, or meshes using a single image. The method by Wu et al. (2017) takes as input a single image, converts the objects within the image to two-and-a-half dimensional (2.5D) sketches, and outputs a 3D model. By deriving 2.5D sketches from the image, this method avoids modelling object appearance variations within the real image. The method by Fan, Su, and Guibas (2017) produces point cloud coordinates from a single image using PointOutNet, a network with encoder and predictor stages. Pontes et al. (2018) present Image2Mesh which takes as input a single image and produces a mesh of the objects found in the image. 3D meshes are used because 3D models and point clouds have higher computational complexity, are less ordered, and lack finer geometry when compared to 3D meshes (Pontes et al. 2018).

9.5. Conclusion

The use of descriptive and expressive virtual objects or augmentations within AR systems is essential in order for the user to derive the most benefit from the system. While augmentations are typically predetermined and saved in the AR application before use, methods are available to create 1D, 2D, and 3D augmentations while the application is being used. Caption generation is a well-researched method within DL that can produce 1D text descriptions of objects within a single image. AI-based sketch or contour generation methods are able to produce a simple 2D sketch based on objects within an image. Current strategies within AR systems are able to use sequential images with SfM formulations to determine the 3D structure of objects within a scene. AI-based virtual object creation methods are even more powerful in that they can determine 3D structure from single images. Further, animation can be combined with these augmentations to increase the interactivity of the system. Automatic virtual object creation methods not only save the system developer time during the AR development process but are able to adapt in real time to the manufacturing environment.

10. Registration

Registration refers to the alignment of the virtual objects or augmentations with the user's environment. Registration combines both the intrinsic and extrinsic parameters found during the camera calibration and detection and tracking steps in order to determine the correct placement and orientation of the virtual objects or augmentations on the screen of the visual interface for the user. One major issue in AR is misregistration, which is when the pose of the virtual objects rendered on the screen do not align perfectly with the real world. There are various types of registration errors that contribute to misregistration, including (Holloway 1997):

- Linear registration error;
- Lateral registration error;
- Depth registration error; and
- Angular registration error.

Linear registration error describes the amount of separation between the real and virtual objects. It is measured by the length of the distance vector between the pose of these objects. Lateral registration error describes the offset between the real and virtual objects if the virtual object were placed at the same distance as the real object with respect to the user's eye. It is measured by the length of the vector that describes the distance between the real object and the virtual object at the same depth as the real object. Depth registration error describes how close or far the virtual object is relative to the real object from the user's perspective. It is measured by subtracting the length of the line of sight for the real object from the line of sight for the virtual object. Finally, angular registration error describes the pitch offset between the objects from the user's perspective. It is measured by the angle that subtends the lateral error line segment (Holloway 1997). A visual depiction of these various types of error can be found in Figure 10.

Registration error has two main sources: static or dynamic (Axholt 2011; Zheng 2015). A static registration error is spatial and is visible to the user when they are stationary (Axholt 2011). Static or spatial errors

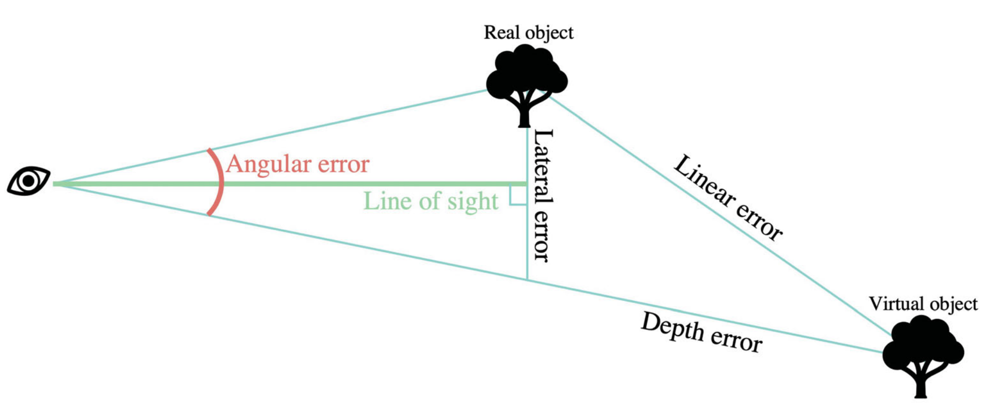


Figure 10. Visual depiction of the registration errors present in any AR system: angular error, lateral error, linear error, and depth error.

result from inaccurate measurements of geometrical relationships in the system (Henderson and Feiner 2007). A dynamic registration error is temporal and is apparent when the user moves the AR system (MacIntyre, Coelho, and Julier 2002; Axholt 2011). Dynamic registration errors tend to have a higher impact on AR systems than static errors (Bauer 2007). System latency or delay, which is the result of delays in every process of the system (Holloway 1997), is the main source of dynamic or temporal registration error (Holloway 1997; Zheng 2015). An additional type of registration error that relates to inverse rendering is photometric error. Zheng (2015) considers five components that contribute to system latency or synchronisation delay: tracking delay, application delay, rendering delay, scanout delay, and display delay. The four approaches currently utilised to reduce system latency and thus dynamic registration errors are latency minimisation, just-in-time image generation, predictive tracking, and video feedback. For a full review of these latency-reducing methods, see Section 2.3.2 of Zheng (2015). For the current review, we are focussing upon closed-loop registration error minimisation strategies, not strategies used within open-loop systems to reduce registration error. This is because the only way to reduce registration error in an open-loop system is to improve the error in each individual source of error throughout the computational pipeline and, as discussed in Section 10.1, this is not a sustainable solution for correcting registration error.

A hierarchical description of the topics covered within registration can be found in Figure 11.

10.1. Current approaches for registration in AR

The solutions developed to minimise registration error in AR systems are diverse. It is difficult to estimate registration error because of the wide array of sources of the error, but a starting point to estimate this error can be the error estimate from the manufacturer for the tracking system that is used (MacIntyre and Coelho 2000).

The majority of AR systems in use and development utilise an open-loop registration approach. In an open-loop system, the registration error propagates through the system and is seen by the user, but not by the system (Zheng, Schubert, and Weich 2013). The only way to improve the registration error in open-loop AR systems is to make each component of the system more accurate (Bajura and Neumann 1995b). In a closed-loop system, the visual feedback from registration is utilised during the detection and tracking steps in order to minimise registration error (Zheng, Schubert, and Weich 2013). Closed-loop AR systems are tolerant of transformation errors (Bajura and Neumann 1995b).

Within open- or closed-loop AR systems, one solution to account for registration error is to modify the final graphics rendered in order to account for the error. Forward warping for AR, or FW-AR, corrects for registration error by distorting augmentations to match the real camera image. Backward warping for Augmented Virtuality (AV), or BW-AV, distorts the real camera image in order to match the augmentations on the screen. FW-AR is preferred over BW-AV because FW-AR does not modify the real camera image. Because backward warping modifies the real camera image, it converts the system into AV, not AR (Zheng, Schmalstieg, and Welch 2014). Additionally, BW-AV only works for video-based interfaces where the real environment is not visible to the user (Bajura and Neumann 1995b). According to Holloway (1997), the only possible way to decrease registration error in Optical See-Through (OST) HMDs is to use predictive head tracking methods.

Many researchers have attempted to better estimate or measure registration error. Holloway (1997) separates the registration error into the following categories:

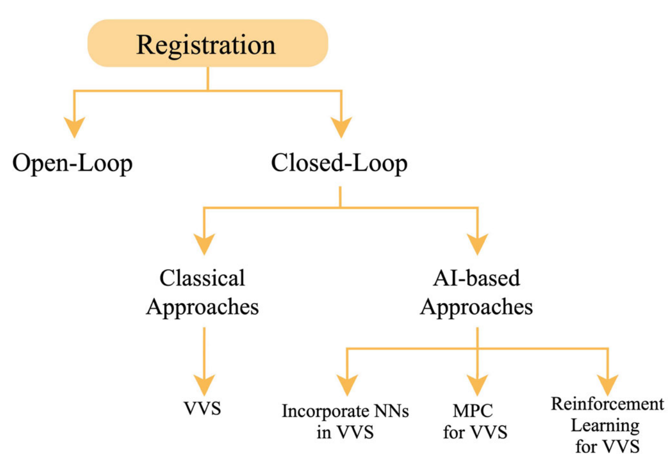


Figure 11. Visual depiction of the topics reviewed in registration.

acquisition or alignment, tracking, display, and viewing (for HMDs). MacIntyre and Coelho (2000) and MacIntyre, Coelho, and Julier (2002) determine registration error based on the expansion and contraction of objects' 2D convex hulls within images. MacIntyre and Coelho (2000) and MacIntyre, Coelho, and Julier (2002) utilise the measured registration error in a Level of Error (LOE) filtering approach where different representations or versions of augmentations are automatically rendered based on the changes in the registration error. Baillot et al. (2003) measure registration error using a calibration method to estimate transformations between the tracking system and the world.

Bajura and Neumann (1995a, 1995b) were some of the earliest investigators into closed-loop AR systems. In their works, 3D coordinate system registration error is dynamically estimated and corrected using the registration error estimate from 2D objects in images. These works use a backward warping or BW-AV approach to distort the real camera image in order to match the parameters of the virtual objects in the scene; thus, this method can only be used for video-based systems. Zheng, Schubert, and Weich (2013) utilise the registration error within 2D images within the model-based tracking procedure in order to create a closed registration loop. Zheng, Schmalstieg, and Welch (2014) handle registration errors on both the global scale and local scale by correcting the camera pose in 3D and by correcting pixel-wise registration errors in 2D. The method utilises a reference model of the real scene in a registration-enforcing model-based tracking (RE-MBT) method to correct the 3D camera pose. FW-AR is utilised to distort the augmentations to match the real camera image. Naik et al. (2015) take inspiration from Bajura and Neumann (1995b) in implementing a closed-loop registration approach that dynamically corrects augmentations for projector-based AR by detecting projected features in each camera image and modifying the augmentations in an FW-AR approach to correct for registration error.

VVS control theory has been utilised by various researchers for closed-loop registration error minimisation. In VVS control theory, features are extracted from images and the error in reference 2D image features is minimised in a closed-loop fashion (Behringer, Park, and Sundareswaran 2002; Liu and Li 2019). Sundareswaran and Behringer's motion estimation-based tracking algorithm is based on the theory of VVS (1999). Behringer, Park, and Sundareswaran (2002) utilise VVS in an outdoor AR system based on 3D model-based tracking. Marchand and Chaumette (2002, 2005) present a pose estimation algorithm that iteratively uses the VVS control law in order to extract the desired features from an image. Comport et al. (2006) utilise VVS control theory with an M-estimator for pose estimation. Pressigout and Marchand (2004) utilise VVS control theory for markerless tracking with limited knowledge of the real scene. Zheng (2015) utilise VVS theory in the development of their closed-loop registration method published in Zheng, Schmalstieg, and Welch (2014), which utilises model-based tracking for global pose estimation and optical flow methods for pixel-based registration adjustment.

10.2. Limitations of current approaches for registration in AR & potential for AI

Closed-loop registration provides AR systems with the ability to minimise registration error without explicitly determining the causes and potential solutions for the registration error present within the system. While closed-loop registration for AR systems is an improvement over open-loop registration, the current strategies for closed-loop registration are not fully developed and are not in widespread use in AR systems. Additionally, VVS closed-loop registration methods require features to be defined in the scene before use.

One possible way to further the effectiveness of closed-loop registration within AR systems is to incorporate DL strategies in the VVS-inspired closed-loop registration approaches. AI-based VVS strategies are designed to work in unknown environments in many cases (Saxena et al. 2017). Additionally, AI-based VVS methods are more accurate, efficient, stable, and robust than traditional image-based visual servoing (IBVS) strategies (Comport et al. 2006).

10.3. AI approaches for registration in AR

To the best of our knowledge, there have been no AI-based registration minimisation strategies utilised within AR-assisted manufacturing applications.

10.4. Potential AI approaches for registration in AR

One AI-based non-VVS registration error minimisation method was found in the literature. Jiang et al. (2020) use two deep networks: (1) initial registration network and (2) registration error network. The initial registration network is a feed-forward network based on VGG-11 (Simonyan and Zisserman 2014) that is used to warp the feature template to match the input image. The registration error DNN then estimates the error of the resultant warping and optimises the image transformation parameters.

The AI-based VVS or IBVS methods either incorporate NNs into the current framework for VVS, utilise predictive models (model-predictive control or MPC) to enable VVS, or use reinforcement learning to enable VVS. Reinforcement learning-based methods for image-based VVS are termed *RL-IBVS* methods.

Within AI-based VVS methods that use NNs in the current VVS framework, Bateux et al. (2017, 2018) incorporate an AlexNet-based CNN with direct visual servoing (DVS) in order to estimate the relative pose between the current and reference image. DVS is a method that enables VVS using a full image with no feature extraction (Collewet and Marchand 2011) and no prior information of the scene or object (Cui 2016). Once the relative pose is determined, the resulting registration error is incorporated back into a classical VVS control law. Saxena et al. (2017) use a similar approach in using a CNN for visual feedback to find the relative camera pose between a pair of images. Liu and Li (2019) use two CNNs, one that takes as input the reference image and one that takes as input the current image, and calculates the error between the reference and current pose of a robot arm end effector.

MPC is a method that enables real-time prediction and correction for image-to-image registration (Ebert et al. 2018). Finn and Levine (2017) perform feedback control at the pixel level by utilising an LSTM RNN structure to predict the pose of objects in images. Ebert et al. (2017) introduce a video prediction model called Skip Connection Neural Advection Model (SNA), based on the dynamic neural advection (DNA) model (Finn, Goodfellow, and Levine 2016) that keeps track of images over multiple frames. Ebert et al. (2018) measure the distance to a goal within a calibrated image as the cost function for their visual MPC VVS approach. Lampe and Riedmiller (2013) use reinforcement learning for use in a visual feedback loop for control of the reaching and grasping mechanism of a robotic arm. Finn et al. (2016) utilise reinforcement learning to relate feature points with a specific task. An automated approach is used for deriving the state-space representation from inputted camera images and a deep spatial AE is used to determine these features. Lee, Levine, and Abbeel (2017) utilise a hybrid reinforcement learning approach between MPC and conventional IBVS. Predictive models learn with pre-trained visual features to track objects within images over time. A Q-iteration procedure is used to choose the most relevant visual features. Sampedro et al. (2018) present deep deterministic policy gradients (DDPG), a deep reinforcement learning algorithm for VVS-based controlling of multirotor aerial robots. The DDPG algorithm controls the aerial robots based on the determined errors in the image plane compared to a reference image.

10.5. Conclusion

While most AR systems in use today utilise open-loop registration methods, correcting registration errors in open-loop systems is difficult because registration error has many sources throughout the AR pipeline. One possible solution to deal with registration error is implementing a closed-loop registration method within the AR system. While closed-loop registration is not widely used in AR systems, these methods have the advantage of being able to reduce or minimise registration error without having to explicitly define the sources of the registration error. Some closed-loop registration methods used in AR systems are derived from VVS within the robotics field, where visual differences in object motion throughout image frames is minimised through a feedback system, but these VVS methods require features within images to be predefined. Utilising AI-based methods within VVS makes for even more efficient closed-loop registration. Further, AI-based VVS methods are able to work without prior knowledge or features defined in the working space. In this way, these methods can be used in generic environments, such as those in manufacturing.

11. Rendering

Once the virtual objects are correctly registered with the real environment, both with respect to pose (detection and tracking) and appearance (inverse rendering), the next step is to render the virtual augmentations onto the screen of the main interface of the AR system. There are two main areas that must be considered in this case: what data to display and how to display this data (i.e. data visualisation). A hierarchical description of the topics covered within rendering can be found in Figure 12.

11.1. Current approaches for rendering in AR

When determining what information to display on the AR interface, many AR systems are context-aware and user-customisable. For instance, the Knowledge-based AR for Maintenance Assistance (KARMA) system by Feiner, Macintyre, and Seligmann (1993) uses a rule-based illustration generation system that incorporates user preferences into data visualisation. The AR system by Mengoni et al. (2015) utilises a 3D coordinate system to balance user preferences for the purposes of data visualisation.

A major issue in data visualisation on a screen is information overload (Van Krevelen and Poelman 2007). In order to combat information overload, two methods are recommended: filtering and visualisation layouts.

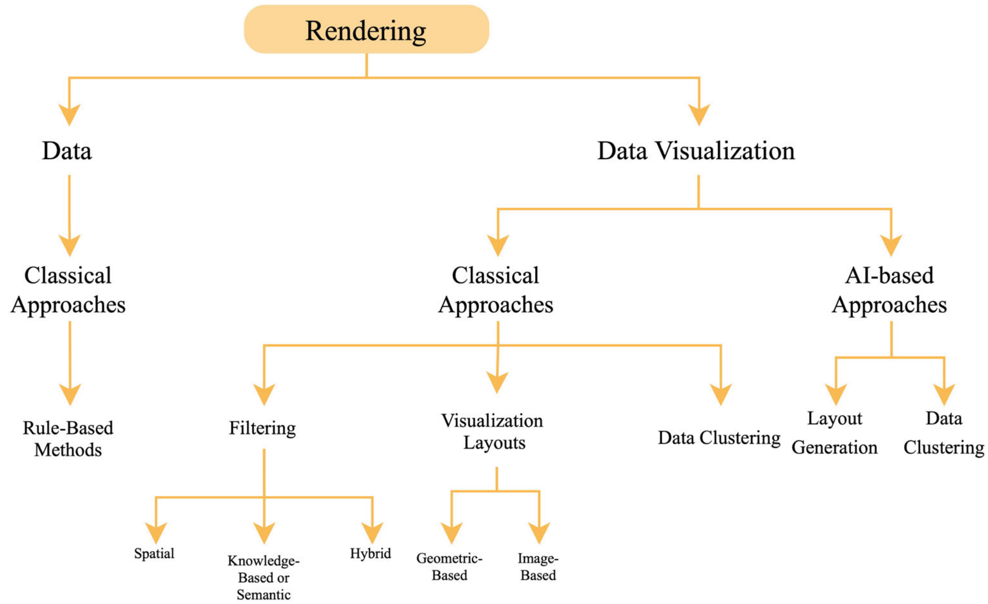


Figure 12. Visual depiction of the topics reviewed in rendering.

Filtering, as in Julier et al. (2000), reduces the number of augmentations given on the screen at any one time based on a metric such as spatial distance (Virkkunen 2018). There are two main types of filtering used in data visualisation: spatial filtering as in Bane and Hollerer (2004) and knowledge-based or semantic filtering as in Höllerer et al. (2001) (Tatzgern et al. 2016). Spatial filtering removes data based on physical distance from the user while knowledge-based or spatial filtering removes data based on user preferences or context (Tatzgern et al. 2016). There are also hybrid methods that combine the use of spatial and knowledge-based filters, as in Julier et al. (2000). With any filtering methods, data loss can occur, and the reduction of visual clutter is not guaranteed (Tatzgern et al. 2016).

Visualisation layouts, as in Bell and Feiner (2003), model the environment and track elements in the scene to make sure augmentations do not occlude important environmental features (Virkkunen 2018). In simple terms, visualisation layout methods reorganise data in a logical way (Tatzgern et al. 2016). The two main visualisation layout methods are geometric-based layout and image-based layout (Ichihashi and Fujinami 2019). Visualisation layout methods fail when the amount of data grows too large (Tatzgern et al. 2016).

Clustering, as in Tatzgern et al. (2016), is an alternative approach in data visualisation that combines information based on spatial and semantic attributes that not only reduces the amount of information shown in the display but also preserves all of the information, unlike filtering (Virkkunen 2018). Clustering is also robust with large amounts of data, unlike the visualisation layout methods. The classical data clustering methods include partition-based methods as in Boomija and Phil (2008), density-based methods as in Kriegel et al. (2011), and hierarchical methods as in Johnson (1967).

11.2. Limitations of current approaches for rendering in AR & potential for AI

Current approaches for determining what data to display to the user are rigid and are based on hard-coded rules. Innovative AI-based tools such as expert systems are available that can incorporate context-awareness into AR systems. These AI-based strategies not only include expert knowledge into the system but also provide more flexibility for the use of the system in varying contexts.

With regard to data visualisation, AI-based layout creation strategies can be used to aid in the development of efficient and effective AR rendering. Further, AI-based data clustering approaches have the ability to transform data into clustering-friendly nonlinear representations, thus easing the data clustering process (Min et al. 2018).

11.3. AI approaches for rendering in AR

Expert systems are AI-based computer programs that emulate the decision-making abilities of human experts (Patel, Virparia, and Patel 2012). Expert systems can aid in determining what data is appropriate to display to the user, thus enabling user customisation. Expert systems have already been incorporated into AR-assisted manufacturing systems. The AR solutions by Syberfeldt et al. (2016) and Holm et al. (2017) utilise expert systems to determine the level of detail displayed for AR-guided maintenance instructions.

One article was found that incorporates AI-based methods for data visualisation within AR-assisted manufacturing applications. Ivaschenko, Milutkin, and Sitnikov (2019) utilise a NN within the Navigator or context module of their AR system in order to handle context-awareness of visualised content.

11.4. Potential AI approaches for rendering in AR

A subset of DL research is dedicated to automating layout design for data visualisation using DL. For instance, Li et al., "Layoutgan" (2019) describe LayoutGAN, a GAN that takes as input a set of randomly placed 2D graphical elements and outputs a layout using the 2D elements that maps to a wireframe image. A CNN-based discriminator is used to optimise the layout within the image. Similarly, Nguyen et al. (2018) present DeepUI, which is a GAN that is trained on current phone apps and outputs optimised wireframe User Interface (UI) designs for the developer. The architecture is a GAN based on a generator, UIGenerator, and a discriminator, UIDiscriminator. Jia et al. (2019) utilise a novel feature map or *guidance map* to identify important image regions for label placement in AR. Zheng et al. (2019) incorporate context-awareness into graphic design layouts. A deep generative model learns complex layout structures from input images and keyword summaries. Li et al., "Auto Completion of User Interface" (2020) utilise transformer-based tree decoders for auto-completion of UIs.

Data clustering is another subset of DL research that can be applied to data visualisation in AR systems. Aljalbout et al. (2018) developed a taxonomy of DL-based data clustering and found that DNN-based clustering methods follow the same format of (1) implementing representation learning or feature mapping using DNNs, and (2) using the representations as input to a clustering method. Some classical unsupervised data clustering methods include k-means as in Tatzgern et al. (2016), Gaussian mixture model (GMM) clustering as in Biernacki, Celeux, and Govaert (2000), maximum-margin clustering as in Xu et al. (2005), agglomerative clustering as in Gowda and Krishna (1978), and information-theoretic clustering as in Li, Zhang, and Jiang (2004). Min et al. provided a review of deep clustering methods in 2018 and found that the current DNN-based clustering methods can be categorised into one of four types:

- Methods that utilise an AE to obtain a feasible feature space;
- Methods that utilise feed-forward networks, termed Clustering DNN (CDNN) methods;
- Methods that use VAEs; and
- Methods based on GANs.

For a full review of DL-based clustering methods, the reader is referred to Min et al. (2018).

11.5. Conclusion

With respect to the data displayed in AR systems, many researchers have identified the importance of adapting their systems to the user's environment and preferences. While classical methods of incorporating customisation into AR systems are strictly rule-based, AI-based methods are able to incorporate customisation into AR systems in a more flexible way either through DNNs or through AI-based expert systems. This adaptation of the AR system allows the operator to work in the manufacturing environment with their highest productivity.

With regards to the visualisation of this data, the majority of classical methods utilised are destructive, as in data filtering, or non-adaptive for large datasets, as in visualisation layouts. While classical clustering methods are effective, they are not as flexible as DL-based clustering methods. Further, DL research into layout optimisation is another strategy that can be utilised to better AR system data visualisation. Showing the operator concise information on intuitive interfaces will decrease their cognitive load and thus increase their efficiency in the manufacturing environment.

12. Future work

Based on the review, there are multiple innovative research directions that can be explored for parts of the AR system and for the entire AR system as a whole. These innovative future research directions are described next.

12.1. Reconstruction using shape primitives

Humans interpret a 2D image not as a projection of a 3D object from a specific viewpoint, rather they infer an abstract 3D representation of the actual 3D object (see Rai and Deshpande 2016). This inference is aided by the fact that the 2D images contain projections of a set of 3D objects. Directly inferring 3D objects/scenes from an image is a challenging problem. However, two basic facts aid human interpretation. Firstly, all the 3D objects are composed of a set of primitive geometric shapes called *geons* (Biederman 1987). The geons represent irreducible components like spheres, cylinders, cuboids, etc. Secondly, the contents of the image are not limited to monochromatic textureless projections of objects. The images contain textured details. The illumination characteristics of a cylinder differ from a cuboid, although both of them can project rectangles. These cues can aid in depth inference. Moreover, the real-time nature of AR can offer details from additional perspectives. Such information can be leveraged for inverse rendering and for an enhanced inference of the 3D environment from the live video feed.

12.2. Synergy of classical and AI-based CV

While AI offers several advantages over classical approa ches, it is not prudent to completely replace classical approaches with pure DL techniques in all aspects of the AR system. This is particularly evident in the task of estimating the absolute pose of the camera. Although Kendall, Grimes, and Cipolla (2015) argue that there exists an injective relationship between the camera pose and the images, Sattler et al. (2019) counter-argue that image-based pose estimation is indeed similar to image retrieval. The counter-argument is reflected in the sub-par performance of DL-based techniques in camera pose estimation. Several researchers have fused data from additional sensors (e.g. GPS, IMU) to increase accuracy. As the accuracy of DL-based methods is not guaranteed, this limits the applicability of DL in precision manufacturing applications. O'Mahony et al. (2019) listed AR as one of the problems that is not suitable for a differentiable implementation. Moreover, DL can sometimes be excessive for simpler tasks like accepting or rejecting objects of a certain color from a conveyor belt. While a DL algorithm would need an enormous amount of data to complete this task, a simple color thresholding algorithm can solve this problem easily. Additionally, due to the unavailability of dedicated comprehensive datasets for all of the various tasks in AR, DL techniques should be implemented in conjunction with classical methods (Rai and Sahu 2020).

12.3. Human-AR synergistic relationship

The typical AR system is designed to aid humans in performing tasks, and thus the AR system is very one-sided. One potential future advancement for AR systems includes utilising input from the user for the betterment of the computational aspects of the AR system itself. This is especially appealing because it is well known that DL algorithms are data hungry and require large training datasets. Data acquisition is an expensive process just because of the sheer amount of data or the limited capability of the equipment. Cases where humans perceive better than machines, like dynamic range captured by cameras and other perceptual cues like occlusion from both the real scene and its image, can heavily support ground truth data collection. AR can expedite the development of state-of-the-art AI techniques.

12.4. Reconfigurability

While the flexibility of procedure storage has been greatly improved through the use of ontologies instead of databases, the data that is stored within these structures is still very rigid as it is directly derived from procedural manuals or from expert developers. One major area of improvement within any procedure-based AR application is reducing the rigidity of these procedures so that they are able to adapt to changes in the use environment and the user. Particularly, current AR-assisted manufacturing applications have limited options for assistance if the operator encounters an unprecedented issue while performing a procedure. If modification of the procedures within the system is possible, the procedure can be modified off-site through a non-mobile workstation, on-site either by the operator themselves or remotely by an expert, or automatically by the AR system itself. Very few AR systems include strategies for automated reconfiguring of procedures, and the inclusion of automated strategies for procedure reconfiguration is recommended for the increased flexibility of the system for use in uncertain or inconsistent environments. Automated procedure reconfiguring can be brought to fruition through automated diagnostic methods (Chou and Yao 2009; Matei et al. 2020), disassembly sequence planning, or another complex representation of the relationships between different aspects of the procedures. Incorporation of domain knowledge into this process is of the utmost importance to ensure that the modified procedures are realistic and safe to perform.

12.5. Usability: human factors of AR

Depending upon the main interaction strategy utilised in the AR system, AR may assist or impede manufacturing activities. For instance, while the computational power of handheld AR devices like tablets and smartphones is enticing, these devices take up the hands of users and thus makes it difficult to perform actions relating to procedures. Thus, in cases of maintenance or assembly tasks, the use of VST, OST, or peripheral HMDs may be more beneficial as a trade-off between hands-free work and computational power. In the case of manual assembly tasks, user mental fatigue from performing the same task repetitively is a major safety hazard. When considering the future development of AR systems in the manufacturing environment, AR systems should be customised not only to environmental context and user preferences but also to current user status. This is especially important when considering the increase in the complexity of AR systems with the incorporation of AI-based algorithms. AI should be incorporated into AR-assisted manufacturing tasks with caution, as even the $\leq 1\%$ failure rate of AI can prove disastrous. Synthesizing this data for the user is not only essential for decreasing cognitive load and increasing focus, but it is also essential for ensuring safety in the manufacturing environment.

12.6. Multimodal AR

AR was first theorised as a technology that could augment all of the user's senses. Although for this review, we only focus on the visual modality of the AR system, a diverse set of input modalities can also be incorporated. Inputs from additional sensors like IMU and GPS can supplement vision for various tasks. This is especially true when considering the vast success of visual-inertial detection and tracking algorithms. The data acquired from such sensors can aid in training the DL techniques. While vision-based pose estimation is expensive and not robust with 3D transformations, inertial sensors are susceptible to noise and calibration errors. These modalities complement each other because inertial sensors allow for fast computation and vision can correct any drift errors from the inertial sensor (Bostanci et al. 2013).

12.7. Feedback in manufacturing

While AR can be utilised in all stages of PLM and impacts managers, employees, and consumers alike, the potential use for AR for each user in PLM is mostly separated. In order to implement a closed-loop system, feedback from consumers should be incorporated into the manufacturing cycle. For instance, AR can be utilised by consumers to troubleshoot issues with the final product. Active feedback and data that results from this process of troubleshooting can be collected and utilised to determine potential product weak points or defects, similar to how product design features are extracted from written customer feedback in Ali et al. (2020). While this information can be estimated using simulation strategies during the product development stage, incorporating direct consumer input would be beneficial in the creation of improved products. So, a centralised AR application with a customised UI for each user displaying only the information relevant to them can be an integral part of the PLM.

12.8. AR system flexibility

The computational pipeline of any AR system is rigidly set based on the requirements of the system, the user, and the use environment. Incorporating AI strategies into an AR system leads to increased overall flexibility of the system. For instance, if AI-based virtual object creation strategies are utilised, virtual objects or augmentations do not need to be created and loaded into the system before use. Similarly, if a closed-loop registration strategy is used, camera calibration is no longer needed (Espiau 1994). In the future, all AR systems should be closed-loop so that they automatically correct their intrinsic and extrinsic camera calibration parameters based on the feedback from the system. Future research on AR systems should be able to work without any parameters or data loaded into the system before use. This is a feasible goal if the AR system is computationally prepared and sufficiently interconnected with its use environment, current conditions, user actions and preferences, and domain knowledge. Further, any information that is collected with one AR device can be shared with other devices in the manufacturing facility, thus creating an increasingly robust and stable solution over time.

13. Conclusion

Our work identifies the scope of AI in AR-assisted manufacturing applications. We have explored the current manufacturing environment and the place for AR and AI in industrial processes. We have reviewed recent literature in both AR-assisted manufacturing and AI-driven manufacturing applications. As a result of this review, current key technical and managerial challenges for incorporating AI into AR-assisted manufacturing applications were outlined and described. The main limitations include issues with the handling of big data for AI within AR systems and larger organisational issues such as safety, regulations, and trust of AI technologies. However, the potential benefits of AI-driven AR-assisted manufacturing applications substantially outweigh these limitations. This review endorses the inclusion of AI as a tool within AR. This inclusion has been overlooked within manufacturing applications thus far.

We split the whole AR computational framework into *camera calibration* for acquiring the intrinsic parameters; *detection*, *tracking* and *camera pose estimation* for acquiring the extrinsic parameters of the camera; *inverse rendering* for inferring the illumination and depth characteristics of the scene for supporting non-intrusive rendering of virtual objects; *procedure storage* for bringing in the stored procedures for augmentation; *virtual object creation* for creating the augmentations; *registration* for registering the virtual objects into the scene; and *rendering* for optimally displaying the virtual objects in the display. In each of the sections, the current practices, their limitations and potential for AI, existing work incorporating

AI, and more importantly, additional AI techniques that can be adopted are detailed.

AI can delimit the AR applications from controlled objects (fiducials) and environments to generic objects and unrestricted work environments. Moreover, as AI can interpret the incoming video feed to simultaneously accomplish detection, tracking, and other tasks, the latency of the AR systems can be drastically reduced with parallel processing. By introducing AI in all stages of AR, we hope for an expedited adoption of AI in AR-assisted manufacturing applications.

Although there are challenges to incorporating AI strategies within AR-assisted manufacturing applications, further development and innovation will result in great gains in efficiency, reliability, and usefulness of AR systems. Although the ubiquitous use of AR applications in the manufacturing sector is not yet a reality, the idea of pervasive AR is quickly becoming tangible.

Notes

1. Cognitive load is the mental effort needed for processing information (Porter and Heppelmann 2017).
2. Short-term trackers were supplied with the position of the target in each frame and did not include a target re-detection module. So, they failed at the first instance of occlusion (Kristan et al. 2019).
3. In addition to tracking, long-term trackers had to identify tracking failure and trigger the included target re-detection module. So, long term trackers were evaluated for their performance under occlusion (Kristan et al. 2019).
4. Short-term trackers evaluated for real-time performance (Kristan et al. 2019).
5. https://motchallenge.net/data/MOT20Det/
6. https://www.votchallenge.net/vot2020/dataset.html

Disclosure statement

No potential conflict of interest was reported by the author(s).

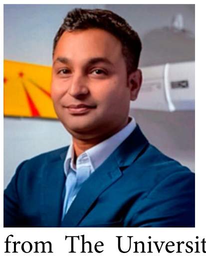
Funding

This work was supported by Naval Surface Warfare Center (NSWC) Naval Engineering Education Consortium (NEEC) award No. N00174-19-0025.

Notes on contributors

Chandan K. Sahu received his B.Tech. degree from National Institute of Technology Rourkela, India in 2014, and his M.S. degree from the State University of New York at Buffalo, USA, in 2020, both in mechanical engineering. He worked at Honda Motorcycle and Scooter India, from 2014 to 2017. He is currently a research assistant at the Geometric Reasoning and Artificial Intelligence Laboratory (GRAIL) at Clemson University's International Center for Automotive Research (CU-ICAR). He is interested in machine learning in general, with a focus on their geometric attributes and mechanical applications. His research is focussed on augmented reality, additive manufacturing, cyber-physical systems and autonomous vehicle applications. His current research work emphasises on supplementing neural networks with physics models for modelling complex systems where only limited data and limited physics are available.

Crystal Young received her B.S. degree in physics from Stony Brook University, Stony Brook, NY, in 2018, and her M.S. degree in mechanical engineering from the University at Buffalo, Buffalo, NY, in 2020. She also holds an Advanced Certificate in Advanced Manufacturing from the University at Buffalo, 2020. She recently completed an internship with the Naval Research Enterprise Internship Program in September 2020. Her current research interests include smart and sustainable manufacturing, optimal design, augmented reality, artificial intelligence, and assistive technologies. She is a member of Sigma Pi Sigma, The National Physics Honor Society.



Rahul Rai received the B.Tech. degree from the National Institute of Foundry and Forge Technology (NIFFT), Ranchi, India, in 2000, the M.S. degree in manufacturing engineering from the Missouri University of Science and Technology (Missouri S&T), USA, in 2002, and the Ph.D. degree in mechanical engineering from The University of Texas at Austin, USA, in 2006. He is currently a Dean's Distinguished Professor in the Department of Automotive Engineering at the International Center for Automotive Engineering (ICAR) at Clemson University. Dr Rai's research is focussed on developing computational tools for Manufacturing, Cyber-Physical System (CPS) Design, Autonomy, Collaborative Human-Technology Systems, Diagnostics and Prognostics, and Extended Reality (XR) domains. By combining engineering innovations with methods from machine learning, AI, statistics and optimisation, and geometric reasoning, his research strives to solve important problems in the above-mentioned domains. His research has been supported by NSF, DARPA, ONR, NSWC, DMDII, HP, NYSERDA, and NYSPII. He has authored over 100 articles to date in peer-reviewed conferences and journals covering a wide array of problems. He was a recipient of the 2009 HP Design Innovation and the 2017 ASME IDETC/CIE Young Engineer Award. He also received the 2019 PHM Society Conference Best Paper Award.

ORCID

Chandan K. Sahu http://orcid.org/0000-0001-5893-4031
Crystal Young http://orcid.org/0000-0001-5291-7468
Rahul Rai http://orcid.org/0000-0002-6478-4065

References

Ababsa, Fakhreddine. 2020. "Augmented Reality Application in Manufacturing Industry: Maintenance and Non-destructive Testing (NDT) Use Cases." International conference on Augmented Reality, Virtual Reality and Computer Graphics, Lecce, Italy, 333–344. Springer. doi:10.1007/978-3-030-58468-9_24

Ababsa, Fakhreddine, Jean-Yves Didier, Imane Zendjebil, and Malik Mallem. 2008. "Markerless Vision-based Tracking of Partially Known 3D Scenes for Outdoor Augmented Reality Applications." International symposium on Visual Computing, Las Vegas, NV, 498–507. Springer.

Ababsa, Fakhr-eddine, and Malik Mallem. 2004. "Robust Camera Pose Estimation Using 2d Fiducials Tracking for Real-Time Augmented Reality Systems." Proceedings of the 2004 ACM SIGGRAPH international conference on Virtual Reality Continuum and its Applications in Industry, Singapore, 431–435.

Ababsa, Fakhreddine, and Malik Mallem. 2011. "Robust Camera Pose Tracking for Augmented Reality Using Particle Filtering Framework." *Machine Vision and Applications* 22 (1): 181–195.

Abburu, Sunitha. 2012. "A Survey on Ontology Reasoners and Comparison." *International Journal of Computer Applications* 57 (17): 33–39.

Abdel-Hakim, Alaa E., and Aly A. Farag. 2006. "CSIFT: A SIFT Descriptor with Color Invariant Characteristics." 2006 IEEE computer society conference on Computer Vision and Pattern Recognition (CVPR'06), Hong Kong, Vol. 2, 1978–1983. IEEE.

Abdi, Lotfi, and Aref Meddeb. 2017. "Driver Information System: A Combination of Augmented Reality and Deep Learning." Proceedings of the symposium on Applied Computing, Marrakech, Morocco, 228–230.

Abraham, Magid, and Marco Annunziata. 2017. "Augmented Reality is Already Improving Worker Performance." Accessed 31 May 2020. https://hbr.org/2017/03/augmented-reality-is-already-improving-worker-performance.

Aggour, Kareem S., Vipul K. Gupta, Daniel Ruscitto, Leonardo Ajdelsztajn, Xiao Bian, Kristen H. Brosnan, Natarajan Chennimalai Kumar, et al. 2019. "Artificial Intelligence/Machine Learning in Manufacturing and Inspection: A GE Perspective." *MRS Bulletin* 44 (7): 545–558.

Akgul, Omer, H. Ibrahim Penekli, and Yakup Genc. 2016. "Applying Deep Learning in Augmented Reality Tracking." 12th international conference on Signal-Image Technology & Internet-based Systems (SITIS), Naples, Italy, 47–54. IEEE.

Alhaija, Hassan Abu, Siva Karthik Mustikovela, Lars Mescheder, Andreas Geiger, and Carsten Rother. 2017. "Augmented Reality Meets Deep Learning for Car Instance Segmentation in Urban Scenes." British Machine Vision Conference, London, UK, Vols. 1, 2.

Ali, Munira Mohd, Mamadou Bilo Doumbouya, Thierry Louge, Rahul Rai, and Mohamed Hedi Karray. 2020. "Ontology-based Approach to Extract Product's Design Features From Online Customers' Reviews." *Computers in Industry* 116: 103175.

Ali, Munira Mohd, Rahul Rai, J. Neil Otte, and Barry Smith. 2019. "A Product Life Cycle Ontology for Additive Manufacturing." *Computers in Industry* 105: 191–203.

Aljalbout, Elie, Vladimir Golkov, Yawar Siddiqui, Maximilian Strobel, and Daniel Cremers. 2018. "Clustering with Deep Learning: Taxonomy and New Methods." Preprint arXiv:1801.07648.

Álvarez, Hugo, Igor Lajas, Andoni Larrañaga, Luis Amozarrain, and Iñigo Barandiaran. 2019. "Augmented Reality System to Guide Operators in the Setup of Die Cutters." *The International Journal of Advanced Manufacturing Technology* 103 (1–4): 1543–1553.

Andriyenko, Anton, Konrad Schindler, and Stefan Roth. 2012. "Discrete-continuous Optimization for Multi-target Tracking." IEEE conference on Computer Vision and Pattern Recognition, Providence, RI, 1926–1933. IEEE.

Arinez, Jorge F., Qing Chang, Robert X. Gao, Chengying Xu, and Jianjing Zhang. 2020. "Artificial Intelligence in Advanced Manufacturing: Current Status and Future Outlook." *Journal of Manufacturing Science and Engineering* 142 (11): 110804.

Axholt, Magnus. 2011. "Pinhole Camera Calibration in the Presence of Human Noise." PhD diss., Linköping University Electronic Press.

Azadeh, A., M. Jeihoonian, B. Maleki Shoja, and S. H. Seyedmahmoudi. 2012. "An Integrated Neural Network–Simulation Algorithm for Performance Optimisation of the Bi-criteria Two-stage Assembly Flow-shop Scheduling Problem with Stochastic Activities." *International Journal of Production Research* 50 (24): 7271–7284.

Azuma, Ronald, Yohan Baillot, Reinhold Behringer, Steven Feiner, Simon Julier, and Blair MacIntyre. 2001. "Recent Advances in Augmented Reality." *IEEE Computer Graphics and Applications* 21 (6): 34–47.

Badrinarayanan, Vijay, Alex Kendall, and Roberto Cipolla. 2017. "Segnet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39 (12): 2481–2495.

Bai, Shuang, and Shan An. 2018. "A Survey on Automatic Image Caption Generation." *Neurocomputing* 311: 291–304.

Baig, Mirza Muhammad Ali, Mian Ihtisham Shah, Muhammad Abdullah Wajahat, Nauman Zafar, and Omar Arif. 2018. "Image Caption Generator with Novel Object Injection." Digital Image Computing: Techniques and Applications (DICTA), Canberra, Australia, 1–8. IEEE.

Bailey, Tim, and Hugh Durrant-Whyte. 2006. "Simultaneous Localization and Mapping (SLAM): Part II." *IEEE Robotics & Automation Magazine* 13 (3): 108–117.

Baillot, Yohan, Simon J. Julier, Dennis Brown, and Mark A. Livingston. 2003. "A Tracker Alignment Framework for Augmented Reality." Proceedings of the 2nd IEEE and ACM international symposium on Mixed and Augmented Reality, Tokyo, Japan, 142–150. IEEE.

Bajic, B., I. Cosic, M. Lazarevic, N. Sremcev, and A. Rikalovic. 2018. "Machine Learning Techniques for Smart Manufacturing: Applications and Challenges in Industry 4.0." 9th International Scientific and Expert Conference, TEAM 2018, Novi Sad, Serbia, 29.

Bajura, Michael, and Ulrich Neumann. 1995a. "Dynamic Registration Correction in Augmented-reality Systems." Proceedings Virtual Reality Annual International Symposium'95, Research Triangle Park, NC, 189–196. IEEE.

Bajura, Michael, and Ulrich Neumann. 1995b. “Dynamic Registration Correction in Video-based Augmented Reality Systems.” *IEEE Computer Graphics and Applications* 15 (5): 52–60.

Balntas, Vassileios, Shuda Li, and Victor Prisacariu. 2018. “Relocnet: Continuous Metric Learning Relocalisation using Neural Nets.” Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 751–767.

Bane, Ryan, and Tobias Hollerer. 2004. “Interactive Tools for Virtual X-ray Vision in Mobile Augmented Reality.” Third IEEE and ACM international symposium on Mixed and Augmented Reality, Arlington, VA, 231–239. IEEE.

Bar-Shalom, Yaakov, Thomas E. Fortmann, and Peter G. Cable. 1990. “Tracking and Data Association.” *The Journal of the Acoustical Society of America* 87: 918–919. doi:10.1121/1.398863.

Baryannis, George, Sahar Validi, Samir Dani, and Grigoris Antoniou. 2019. “Supply Chain Risk Management and Artificial Intelligence: State of the Art and Future Research Directions.” *International Journal of Production Research* 57 (7): 2179–2202.

Bateux, Quentin, Eric Marchand, Jürgen Leitner, François Chaumette, and Peter Corke. 2017. “Visual Servoing from Deep Neural Networks.” Preprint arXiv:1705.08940.

Bateux, Quentin, Eric Marchand, Jürgen Leitner, François Chaumette, and Peter Corke. 2018. “Training Deep Neural Networks for Visual Servoing.” IEEE International Conference on Robotics and Automation (ICRA), Brisbane, Australia, 1–8. IEEE.

Bauer, Martin. 2007. “Tracking Errors in Augmented Reality.” PhD diss., Technische Universität München.

Bay, Herbert, Tinne Tuytelaars, and Luc Van Gool. 2006. “Surf: Speeded Up Robust Features.” European Conference on Computer Vision, Graz, Austria, 404–417. Springer.

Behringer, Reinhold, Jun Park, and Venkataraman Sundareswaran. 2002. “Model-based Visual Tracking for Outdoor Augmented Reality Applications.” Proceedings of the international symposium on Mixed and Augmented Reality, Darmstadt, Germany, 277–322. IEEE.

Bell, B., and S. Feiner. 2003. “Augmented Reality for Collaborative Exploration of Unfamiliar Environments.” NFS Workshop on Collaborative Virtual Reality and Visualization, Lake Tahoe, NV, 26–28. Citeseer.

Benhimane, Selim, and Ezio Malis. 2007. “Homography-based 2D Visual Tracking and Servoing.” *The International Journal of Robotics Research* 26 (7): 661–676.

Berclaz, Jerome, Francois Fleuret, Engin Turetken, and Pascal Fua. 2011. “Multiple Object Tracking Using K-shortest Paths Optimization.” *IEEE Transactions on Pattern Analysis and Machine Intelligence*33 (9): 1806–1819.

Bertinetto, Luca, Jack Valmadre, Joao F. Henriques, Andrea Vedaldi, and Philip H. S. Torr. 2016. “Fully-convolutional Siamese Networks for Object Tracking.” European Conference on Computer Vision, Amsterdam, The Netherlands, 850–865. Springer.

Bewley, Alex, Zongyuan Ge, Lionel Ott, Fabio Ramos, and Ben Upcroft. 2016. “Simple Online and Realtime Tracking.” IEEE International Conference on Image Processing (ICIP), Phoenix, AZ, 3464–3468. IEEE.

Bian, Xiao, Ser Nam Lim, and Ning Zhou. 2016. “Multiscale Fully Convolutional Network with Application to Industrial Inspection.” IEEE Winter conference on Applications of Computer Vision (WACV), Lake Placid, NY, 1–8. IEEE.

Biederman, Irving. 1987. “Recognition-by-Components: A Theory of Human Image Understanding.” *Psychological Review* 94 (2): 115.

Biernacki, Christophe, Gilles Celeux, and Gérard Govaert. 2000. “Assessing a Mixture Model for Clustering with the Integrated Completed Likelihood.” *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22 (7): 719–725.

Bleser, Gabriele, Yulian Pastarmov, and Didier Stricker. 2005. “Real-time 3D Camera Tracking for Industrial Augmented Reality Applications.” The 13th international conference in Central Europe on Computer Graphics, Visualization and Computer Vision, University of West Bohemia, Campus Bory, Plzen-Bory, Czech Republic, 47–54.

Blostein, Dorothea, and Narendra Ahuja. 1989. “Shape From Texture: Integrating Texture-Element Extraction and Surface Estimation.” *IEEE Transactions on Pattern Analysis and Machine Intelligence*11 (12): 1233–1251.

Bogdan, Oleksandr, Viktor Eckstein, Francois Rameau, and Jean-Charles Bazin. 2018. “DeepCalib: A Deep Learning Approach for Automatic Intrinsic Calibration of Wide Field-of-View Cameras.” Proceedings of the 15th ACM SIGGRAPH European conference on Visual Media Production, Munich, Germany, 1–10.

Bolme, David S., J. Ross Beveridge, Bruce A. Draper, and Yui Man Lui. 2010. “Visual Object Tracking using Adaptive Correlation Filters.” IEEE computer society conference on Computer Vision and Pattern Recognition, San Francisco, CA, 2544–2550. IEEE.

Boomija, M. D., and M. Phil. 2008. “Comparison of Partition Based Clustering Algorithms.” *Journal of Computer Applications* 1 (4): 18–21.

Bostanci, Erkan, Nadia Kanwal, Shoaib Ehsan, and Adrian F. Clark. 2013. “User Tracking Methods for Augmented Reality.” *International Journal of Computer Theory and Engineering* 5 (1): 93.

Bouthemy, Patrick. 1989. “A Maximum Likelihood Framework for Determining Moving Edges.” *IEEE Transactions on Pattern Analysis and Machine Intelligence* 11 (5): 499–511.

Brachmann, Eric, Alexander Krull, Sebastian Nowozin, Jamie Shotton, Frank Michel, Stefan Gumhold, and Carsten Rother. 2017. “Dsac-differentiable RANSAC for Camera Localization.” Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, Honolulu, HI, 6684–6692.

Brachmann, Eric, and Carsten Rother. 2018. “Learning Less is More-6D Camera Localization via 3D Surface Regression.” Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, 4654–4662.

Breen, David E., Eric Rose, and Ross T. Whitaker. 1995. *Interactive Occlusion and Collision of Real and Virtual Objects in Augmented Reality*. Munich: European Computer Industry Research Center. Accessed 28 May 2020. http://www.cs.iupui.edu/ tuceryan/research/AR/ECRC-95-02.pdf.

Brunken, Hauke, and Clemens Gühmann. 2020. “Deep Learning Self-calibration from Planes.” 12th International Conference on Machine Vision (ICMV 2019), Amsterdam, The Netherlands, Vol. 11433, 114333L. International Society for Optics and Photonics.

Bruno, Fabio, Loris Barbieri, Emanuele Marino, Maurizio Muzzupappa, Luigi D'Oriano, and Biagio Colacino. 2019. "An Augmented Reality Tool to Detect and Annotate Design Variations in An Industry 4.0 Approach." *The International Journal of Advanced Manufacturing Technology* 105 (1–4): 875–887.

Byeon, Wonmin, Thomas M. Breuel, Federico Raue, and Marcus Liwicki. 2015. "Scene Labeling with LSTM Recurrent Neural Networks." Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, Boston, MA, 3547–3555.

Casser, Vincent, Soeren Pirk, Reza Mahjourian, and Anelia Angelova. 2019. "Depth Prediction Without the Sensors: Leveraging Structure for Unsupervised Learning from Monocular Videos." Proceedings of the AAAI conference on Artificial Intelligence, Honolulu, HI, Vol. 33, 8001–8008.

Castle, Robert Oliver, Darren J. Gawley, Georg Klein, and David W. Murray. 2007. "Towards Simultaneous Recognition, Localization and Mapping for Hand-held and Wearable Cameras." Proceedings of the IEEE international conference on Robotics and Automation, Roma, Italy, 4102–4107. IEEE.

Chang, M. M. L., A. Y. C. Nee, and S. K. Ong. 2020. "Interactive AR-assisted Product Disassembly Sequence Planning (ARDIS)." *International Journal of Production Research* 58: 4916–4931.

Chang, M. M. L., S. K. Ong, and A. Y. C. Nee. 2017. "AR-guided Product Disassembly for Maintenance and Remanufacturing, Kamakura, Japan." *Procedia Cirp* 61 (1): 299–304.

Chen, Weifeng, Zhao Fu, Dawei Yang, and Jia Deng. 2016. "Single-image Depth Perception in the Wild." Advances in Neural Information Processing Systems, Barcelona, Spain, 730–738.

Chen, Xinlei, Ross Girshick, Kaiming He, and Piotr Dollár. 2019. "Tensormask: A Foundation for Dense Object Segmentation." Proceedings of the IEEE international conference on Computer Vision, Seoul, Korea, 2061–2069.

Chen, Liang-Chieh, Alexander Hermans, George Papandreou, Florian Schroff, Peng Wang, and Hartwig Adam. 2018. "MaskLab: Instance Segmentation by Refining Object Detection with Semantic and Direction Features." Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, 4013–4022.

Chen, C. J., J. Hong, and S. F. Wang. 2015. "Automated Positioning of 3D Virtual Scene in AR-based Assembly and Disassembly Guiding System." *The International Journal of Advanced Manufacturing Technology* 76 (5–8): 753–764.

Chen, Liang-Chieh, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L. Yuille. 2017. "DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 40 (4): 834–848.

Chen, Longtao, Xiaojiang Peng, and Mingwu Ren. 2018. "Recurrent Metric Networks and Batch Multiple Hypothesis for Multi-object Tracking." *IEEE Access* 7: 3093–3105.

Chen, Bao Xin, and John K. Tsotsos. 2019. "Fast Visual Object Tracking with Rotated Bounding Boxes." Preprint arXiv:1907.03892.

Chen, Yajing, Shikui Tu, Yuqi Yi, and Lei Xu. 2017. "Sketch-pix2seq: A Model to Generate Sketches of Multiple Categories." Preprint arXiv:1709.04121.

Chen, Chu-Song, Chi-Kuo Yu, and Yi-Ping Hung. 1999. "New Calibration-free Approach for Augmented Reality based on Parameterized Cuboid Structure." Proceedings of the 7th IEEE international conference on Computer Vision, Kerkyra, Greece, Vol. 1, 30–37. IEEE.

Chendeb, Safwan, Mohamad Fawaz, and Vincent Guitteny. 2013. "Calibration of a Moving Zoom-lens Camera for Augmented Reality Applications." IEEE international symposium on Industrial Electronics, Taipei, Taiwan, 1–5. IEEE.

Cheng, Dachuan, Jian Shi, Yanyun Chen, Xiaoming Deng, and Xiaopeng Zhang. 2018. "Learning Scene Illumination by Pairwise Photos from Rear and Front Mobile Cameras." *Computer Graphics Forum* 37: 213–221.

Chi, Hung-Lin, Shih-Chung Kang, and Xiangyu Wang. 2013. "Research Trends and Opportunities of Augmented Reality Applications in Architecture, Engineering, and Construction." *Automation in Construction* 33: 116–122.

Chou, Ying-Chieh, and Leehter Yao. 2009. "Automatic Diagnostic System of Electrical Equipment using Infrared Thermography." International conference of Soft Computing and Pattern Recognition, Malacca, Malaysia, 155–160. IEEE.

Chum, Ondrej, and Jiri Matas. 2005. "Matching with PROSAC-progressive Sample Consensus." IEEE computer society conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, Vol. 1, 220–226. IEEE.

Collewet, Christophe, and Eric Marchand. 2011. "Photometric Visual Servoing." *IEEE Transactions on Robotics* 27 (4): 828–834.

Comaniciu, Dorin, and Peter Meer. 2002. "Mean Shift: A Robust Approach Toward Feature Space Analysis." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24 (5): 603–619.

Comaniciu, Dorin, Visvanathan Ramesh, and Peter Meer. 2000. "Real-time Tracking of Non-rigid Objects using Mean Shift." Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, CVPR 2000 (Cat. No. PR00662), Hilton Head Island, SC, Vol. 2, 142–149. IEEE.

Comport, Andrew I., Éric Marchand, and François Chaumette. 2003. "A Real-time Tracker for Markerless Augmented Reality." Proceedings of the 2nd IEEE and ACM international symposium on Mixed and Augmented Reality, Tokyo, Japan, 36–45. IEEE.

Comport, Andrew I., Eric Marchand, Muriel Pressigout, and Francois Chaumette. 2006. "Real-time Markerless Tracking for Augmented Reality: The Virtual Visual Servoing Framework." *IEEE Transactions on Visualization and Computer Graphics* 12 (4): 615–628.

Cordts, Marius, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. 2016. "The Cityscapes Dataset for Semantic Urban Scene Understanding." Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, Las Vegas, NV, 3213–3223.

Criminisi, Antonio, Patrick Perez, and Kentaro Toyama. 2003. "Object Removal by Exemplar-based Inpainting." Proceedings of the 2003 IEEE computer society conference on Computer Vision and Pattern Recognition, Madison, WI, Vol. 2, II–II. IEEE.

Crowder, Richard M., Max L. Wilson, David Fowler, Nigel Shadbolt, Gary Wills, and Sylvia Wong. 2009. "Navigation Over a Large Ontology for Industrial Web Applications." ASME 2009 International Design Engineering Technical

Conferences and Computers and Information in Engineering Conference, San Diego, CA, 1333–1340. American Society of Mechanical Engineers Digital Collection.

Cui, Le. 2016. "Robust Micro/Nano-Positioning by Visual Servoing." PhD diss., Institut de Recherche en Informatique et Système Aléatoires, Université de Rennes.

Dacko, Scott G. 2017. "Enabling Smart Retail Settings Via Mobile Augmented Reality Shopping Apps." *Technological Forecasting and Social Change* 124: 243–256.

Dagnaw, Gizealew. 2020. "Artificial Intelligence towards Future Industrial Opportunities and Challenges." The 6th Annual ACIST Proceedings, Addis Ababa University, Ethiopia. Accessed 11 May 2020. https://digitalcommons.kennesaw.edu/cgi/viewcontent.cgi?article=1027&context=acist.

Dai, Jifeng, Kaiming He, and Jian Sun. 2016. "Instance-aware Semantic Segmentation via Multi-task Network Cascades." Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, Las Vegas, NV, 3150–3158.

Dai, Jifeng, Yi Li, Kaiming He, and Jian Sun. 2016. "R-FCN: Object Detection via Region-based Fully Convolutional Networks." Advances in Neural Information Processing Systems, Barcelona, Spain, 379–387.

Dame, Amaury, and Eric Marchand. 2010. "Accurate Real-time Tracking Using Mutual Information." IEEE international symposium on Mixed and Augmented Reality, Seoul, Korea, 47–56. IEEE.

Danelljan, Martin, Goutam Bhat, Fahad Shahbaz Khan, and Michael Felsberg. 2017. "Eco: Efficient Convolution Operators for Tracking." Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, Honolulu, HI, 6638–6646.

Danelljan, Martin, Gustav Hager, Fahad Shahbaz Khan, and Michael Felsberg. 2015. "Convolutional Features for Correlation Filter based Visual Tracking." Proceedings of the IEEE international conference on Computer Vision Workshops, Santiago, Chile, 58–66.

Dash, Ajaya Kumar, Santosh Kumar Behera, Debi Prosad Dogra, and Partha Pratim Roy. 2018. "Designing of Marker-based Augmented Reality Learning Environment for Kids Using Convolutional Neural Network Architecture." *Displays* 55: 46–54.

Dave, Achal, Tarasha Khurana, Pavel Tokmakov, Cordelia Schmid, and Deva Ramanan. 2020. "TAO: A Large-Scale Benchmark for Tracking Any Object." Preprint arXiv:2005.10356.

Deac, Crina Narcisa, Gicu Calin Deac, Cicerone Laurentiu Popa, Mihalache Ghinea, and Costel Emil Cotet. 2017. "Using Augmented Reality in Smart Manufacturing." Proceedings of the 28th DAAAM International Symposium, Zadar, Croatia, 0727–0732.

Debevec, Paul. 2008. "Rendering Synthetic Objects into Real Scenes: Bridging Traditional and Image-based Graphics with Global Illumination and High Dynamic Range Photography." ACM SIGGRAPH 2008 Classes, Orlando, FL, 1–10.

Dementhon, Daniel F., and Larry S. Davis. 1995. "Model-based Object Pose in 25 Lines of Code." *International Journal of Computer Vision* 15 (1–2): 123–141.

Deng, Jia, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. "ImageNet: A Large-Scale Hierarchical Image Database." IEEE conference on Computer Vision and Pattern Recognition, Miami, FL, 248–255. IEEE.

Doil, Fabian, W. Schreiber, T. Alt, and C. Patron. 2003. "Augmented Reality for Manufacturing Planning." Proceedings of the workshop on Virtual Environments 2003, Zurich, Switzerland, 71–76.

Donné, Simon, Jonas De Vylder, Bart Goossens, and Wilfried Philips. 2016. "MATE: Machine Learning for Adaptive Calibration Template Detection." *Sensors* 16 (11): 1858.

Dorfmüller, Klaus. 1999. "Robust Tracking for Augmented Reality Using Retroreflective Markers." *Computers & Graphics* 23 (6): 795–800.

Doshi, Ashish, Ross T. Smith, Bruce H. Thomas, and Con Bouras. 2017. "Use of Projector Based Augmented Reality to Improve Manual Spot-welding Precision and Accuracy for Automotive Manufacturing." *The International Journal of Advanced Manufacturing Technology* 89 (5–8): 1279–1293.

Dosovitskiy, Alexey, Philipp Fischer, Eddy Ilg, Philip Hausser, Caner Hazirbas, Vladimir Golkov, Patrick Van Der Smagt, Daniel Cremers, and Thomas Brox. 2015. "FlowNet: Learning Optical Flow with Convolutional Networks." Proceedings of the IEEE international conference on Computer Vision, Santiago, Chile, 2758–2766.

Drummond, Tom, and Roberto Cipolla. 1999. "Visual Tracking and Control using Lie Algebras." Proceedings of the 1999 IEEE computer society conference on Computer Vision and Pattern Recognition (Cat. No PR00149), Fort Collins, CO, Vol. 2, 652–657. IEEE.

Drummond, Tom, and Roberto Cipolla. 2002. "Real-time Visual Tracking of Complex Structures." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24 (7): 932–946.

Durrant-Whyte, Hugh, and Tim Bailey. 2006. "Simultaneous Localization and Mapping: Part I." *IEEE Robotics & Automation Magazine* 13 (2): 99–110.

Ebert, Frederik, Sudeep Dasari, Alex X. Lee, Sergey Levine, and Chelsea Finn. 2018. "Robustness via Retrying: Closed-loop Robotic Manipulation with Self-supervised Learning." Preprint arXiv:1810.03043.

Ebert, Frederik, Chelsea Finn, Alex X. Lee, and Sergey Levine. 2017. "Self-supervised Visual Planning with Temporal Skip Connections." Preprint arXiv:1710.05268.

Eigen, David, and Rob Fergus. 2015. "Predicting Depth, Surface Normals and Semantic Labels with a Common Multi-scale Convolutional Architecture." Proceedings of the IEEE international conference on Computer Vision, Santiago, Chile, 2650–2658.

Eigen, David, Christian Puhrsch, and Rob Fergus. 2014. "Depth Map Prediction from a Single Image using a Multi-scale Deep Network." Advances in Neural Information Processing Systems, Montreal, Canada, 2366–2374.

Erhan, Dumitru, Christian Szegedy, Alexander Toshev, and Dragomir Anguelov. 2014. "Scalable Object Detection Using Deep Neural Networks." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, 2147–2154.

Escobar, Carlos A., and Ruben Morales-Menendez. 2018. "Machine Learning Techniques for Quality Control in High Conformance Manufacturing Environment." *Advances in Mechanical Engineering* 10 (2): 1687814018755519.

Espiau, Bernard. 1994. "Effect of Camera Calibration Errors on Visual Servoing in Robotics." Experimental Robotics III, Kyoto, Japan, 182–192. Springer.

Everingham, Mark, Luc Van Gool, Christopher K. I. Williams, John Winn, and Andrew Zisserman. 2010. "The Pascal

Visual Object Classes (VOC) Challenge." *International Journal of Computer Vision* 88 (2): 303–338.

Facil, Jose M., Benjamin Ummenhofer, Huizhong Zhou, Luis Montesano, Thomas Brox, and Javier Civera. 2019. "CAM-Convs: Camera-aware Multi-scale Convolutions for Single-view Depth." Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, Long Beach, CA, 11826–11835.

Fan, Haoqiang, Hao Su, and Leonidas J. Guibas. 2017. "A Point Set Generation Network for 3D Object Reconstruction from a Single Image." Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, Honolulu, HI, 605–613.

Fang, H. C., S. K. Ong, and A. Y. C. Nee. 2013. "Orientation Planning of Robot End-effector Using Augmented Reality." *The International Journal of Advanced Manufacturing Technology* 67 (9–12): 2033–2049.

Faugeras, Olivier D., Q.-T. Luong, and Stephen J. Maybank. 1992. "Camera Self-calibration: Theory and Experiments." European conference on Computer Vision, Santa Margherita Ligure, Italy, 321–334. Springer.

Feiner, Steven, Blair Macintyre, and Dorée Seligmann. 1993. "Knowledge-based Augmented Reality." *Communications of the ACM* 36 (7): 53–62.

Feng, Tuo, and Dongbing Gu. 2019. "Sganvo: Unsupervised Deep Visual Odometry and Depth Estimation with Stacked Generative Adversarial Networks." *IEEE Robotics and Automation Letters* 4 (4): 4431–4437.

Feng, Qianyu, Yu Wu, Hehe Fan, Chenggang Yan, Mingliang Xu, and Yi Yang. 2020. "Cascaded Revision Network for Novel Object Captioning." *IEEE Transactions on Circuits and Systems for Video Technology*.

Fiala, Mark. 2005. "ARTag, A Fiducial Marker System using Digital Techniques." IEEE computer society conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, Vol. 2, 590–596. IEEE.

Finn, Chelsea, Ian Goodfellow, and Sergey Levine. 2016. "Unsupervised Learning for Physical Interaction Through Video Prediction." Advances in Neural Information Processing Systems, Barcelona, Spain, 64–72.

Finn, Chelsea, and Sergey Levine. 2017. "Deep Visual Foresight for Planning Robot Motion." IEEE international conference on Robotics and Automation (ICRA), Mariana Bay Sands, Singapore, 2786–2793. IEEE.

Finn, Chelsea, Xin Yu Tan, Yan Duan, Trevor Darrell, Sergey Levine, and Pieter Abbeel. 2016. "Deep Spatial Autoencoders for Visuomotor Learning." IEEE International Conference on Robotics and Automation (ICRA), Stockholm, Sweden, 512–519. IEEE.

Fischer, Jan, Dirk Bartz, and Wolfgang Straßer. 2004. "Occlusion Handling for Medical Augmented Reality using a Volumetric Phantom Model." Proceedings of the ACM symposium on Virtual Reality Software and Technology, Hong Kong, 174–177.

Fischler, Martin A., and Robert C. Bolles. 1981. "Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography." *Communications of the ACM* 24 (6): 381–395.

Flatt, Holger, Nils Koch, Carsten Röcker, Andrei Günter, and Jürgen Jasperneite. 2015. "A Context-aware Assistance System for Maintenance Applications in Smart Factories based on Augmented Reality and Indoor Localization." IEEE 20th conference on Emerging Technologies & Factory Automation (ETFA), Luxembourg, 1–4. IEEE.

Fu, Cheng-Yang, Wei Liu, Ananth Ranga, Ambrish Tyagi, and Alexander C. Berg. 2017. "Dssd: Deconvolutional Single Shot Detector." Preprint arXiv:1701.06659.

Furini, Francesco, Rahul Rai, Barry Smith, Giorgio Colombo, and Venkat Krovi. 2016. "Development of a Manufacturing Ontology for Functionally Graded Materials." ASME 2016 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, Charlotte, NC, American Society of Mechanical Engineers Digital Collection.

Garcia-Garcia, Alberto, Sergio Orts-Escolano, Sergiu Oprea, Victor Villena-Martinez, and Jose Garcia-Rodriguez. 2017. "A Review on Deep Learning Techniques Applied to Semantic Segmentation." Preprint arXiv:1704.06857.

Gardner, Marc-André, Kalyan Sunkavalli, Ersin Yumer, Xiaohui Shen, Emiliano Gambaretto, Christian Gagné, and Jean-François Lalonde. 2017. "Learning to Predict Indoor Illumination from a Single Image." Preprint arXiv:1704.00090.

Garg, Ravi, B. G. Vijay Kumar, Gustavo Carneiro, and Ian Reid. 2016. "Unsupervised CNN for Single View Depth Estimation: Geometry to the Rescue." European conference on Computer Vision, Amsterdam, The Netherlands, 740–756. Springer.

Garon, Mathieu, Kalyan Sunkavalli, Sunil Hadap, Nathan Carr, and Jean-François Lalonde. 2019. "Fast Spatially-Varying Indoor Lighting Estimation." Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, Long Beach, CA, 6908–6917.

Gavish, Nirit, Teresa Gutiérrez, Sabine Webel, Jorge Rodríguez, Matteo Peveri, Uli Bockholt, and Franco Tecchia. 2015. "Evaluating Virtual Reality and Augmented Reality Training for Industrial Maintenance and Assembly Tasks." *Interactive Learning Environments* 23 (6): 778–798.

Geiger, Andreas, Philip Lenz, Christoph Stiller, and Raquel Urtasun. 2013. "Vision Meets Robotics: The Kitti Dataset." *The International Journal of Robotics Research* 32 (11): 1231–1237.

Geiger, Andreas, Philip Lenz, and Raquel Urtasun. 2012. "Are We Ready for Autonomous Driving? The Kitti Vision Benchmark Suite." IEEE conference on Computer Vision and Pattern Recognition, Providence, RI, 3354–3361. IEEE.

Gheisari, Masoud, Shane Goodman, Justin Schmidt, Graceline Williams, and Javier Irizarry. 2014. "Exploring BIM and Mobile Augmented Reality Use in Facilities Management." Construction Research Congress 2014: Construction in a Global Network, Atlanta, GA, 1941–1950.

Girshick, Ross. 2015. "Fast R-CNN." Proceedings of the IEEE international conference on Computer Vision, Santiago, Chile, 1440–1448.

Girshick, Ross, Jeff Donahue, Trevor Darrell, and Jitendra Malik. 2014. "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation." Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, Columbus, OH, 580–587.

Godard, Clément, Oisin Mac Aodha, Michael Firman, and Gabriel J. Brostow. 2019. "Digging into Self-supervised Monocular Depth Estimation." Proceedings of the IEEE international conference on Computer Vision, Seoul, Korea, 3828–3838.

Gomez, J.-F. V., Gilles Simon, and M.-O. Berger. 2005. "Calibration Errors in Augmented Reality: A Practical Study." 4th IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR'05), Vienna, Austria, 154–163. IEEE.

Gonzalez-Franco, Mar, Julio Cermeron, Katie Li, Rodrigo Pizarro, Jacob Thorn, Windo Hutabarat, Ashutosh Tiwari, and Pablo Bermell-Garcia. 2016. "Immersive Augmented Reality Training for Complex Manufacturing Scenarios." Preprint arXiv:1602.01944.

Goodfellow, Ian, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. "Generative Adversarial Nets." Advances in Neural Information Processing Systems, Montreal, Canada, 2672–2680.

Gordon, Gaile, Mark Billinghurst, Melanie Bell, John Woodfill, Bill Kowalik, Alex Erendi, and Janet Tilander. 2002. "The Use of Dense Stereo Range Data in Augmented Reality." Proceedings of the international symposium on Mixed and Augmented Reality, Darmstadt, Germany, 14–23. IEEE.

Gordon, Ariel, Hanhan Li, Rico Jonschkowski, and Anelia Angelova. 2019. "Depth from Videos in the Wild: Unsupervised Monocular depth Learning from Unknown Cameras." Proceedings of the IEEE international conference on Computer Vision, Seoul, Korea, 8977–8986.

Gordon, Iryna, and David G. Lowe. 2006. "What and Where: 3D Object Recognition with Accurate Pose." In *Toward Category-level Object Recognition*, 67–82. Springer.

Gowda, K. Chidananda, and G. Krishna. 1978. "Agglomerative Clustering Using the Concept of Mutual Nearest Neighbourhood." *Pattern Recognition* 10 (2): 105–112.

Green, Scott A., Mark Billinghurst, XiaoQi Chen, and J. Geoffrey Chase. 2008. "Human-robot Collaboration: A Literature Review and Augmented Reality Approach in Design." *International Journal of Advanced Robotic Systems* 5 (1): 1–18.

Gruber, Thomas R. 1993. "A Translation Approach to Portable Ontology Specifications." *Knowledge Acquisition* 5 (2): 199–221.

Gruber, Lukas, Thomas Richter-Trummer, and Dieter Schmalstieg. 2012. "Real-time Photometric Registration from Arbitrary Geometry." IEEE International Symposium on Mixed and Augmented Reality (ISMAR), Atlanta, GA, 119–128. IEEE.

Guo, Qing, Wei Feng, Ce Zhou, Rui Huang, Liang Wan, and Song Wang. 2017. "Learning Dynamic Siamese Network for Visual Object Tracking." Proceedings of the IEEE international conference on Computer Vision, Venice, Italy, 1763–1771.

Gwn Lore, Kin, Kishore Reddy, Michael Giering, and Edgar A. Bernal. 2018. "Generative Adversarial Networks for Depth Map Estimation from RGB Video." Proceedings of the IEEE conference on Computer Vision and Pattern Recognition Workshops, Salt Lake City, UT, 1177–1185.

Ha, David, and Douglas Eck. 2017. "A Neural Representation of Sketch Drawings." Preprint arXiv:1704.03477.

Ha, Jaewon, Jinki Jung, ByungOk Han, Kyusung Cho, and Hyun S. Yang. 2011. "Mobile Augmented Reality using Scalable Recognition and Tracking." IEEE Virtual Reality Conference, Singapore, 211–212. IEEE.

Hahne, Uwe, and Marc Alexa. 2009. "Depth Imaging by Combining Time-of-Flight and On-demand Stereo." Workshop on Dynamic 3D Imaging, Jena, Germany, 70–83. Springer.

Hamid Rezatofighi, Seyed, Anton Milan, Zhen Zhang, Qinfeng Shi, Anthony Dick, and Ian Reid. 2015. "Joint Probabilistic Data Association Revisited." Proceedings of the IEEE international conference on Computer Vision, Santiago, Chile, 3047–3055.

Hariharan, Bharath, Pablo Arbeláez, Ross Girshick, and Jitendra Malik. 2014. "Simultaneous Detection and Segmentation." European conference on Computer Vision, Zurich, Switzerland, 297–312. Springer.

Hariharan, Bharath, Pablo Arbelaez, Ross Girshick, and Jitendra Malik. 2016. "Object Instance Segmentation and Fine-grained Localization Using Hypercolumns." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39 (4): 627–639.

Harris, Chris. 1993. "Tracking with Rigid Models." In *Active Vision*, 59–73.

Hayder, Zeeshan, Xuming He, and Mathieu Salzmann. 2017. "Boundary-aware Instance Segmentation." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, 5696–5704.

He, Kaiming, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. 2017. "Mask R-CNN." Proceedings of the IEEE international conference on Computer Vision, Venice, Italy, 2961–2969.

He, Anfeng, Chong Luo, Xinmei Tian, and Wenjun Zeng. 2018a. "Towards a Better Match in Siamese Network based Visual Object Tracker." Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 1–16.

He, Anfeng, Chong Luo, Xinmei Tian, and Wenjun Zeng. 2018b. "A Twofold Siamese Network for Real-Time Object Tracking." Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, 4834–4843.

He, Lei, Guanghui Wang, and Zhanyi Hu. 2018. "Learning Depth From Single Images with Deep Neural Network Embedding Focal Length." *IEEE Transactions on Image Processing* 27 (9): 4676–4689.

He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2015. "Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 37 (9): 1904–1916.

He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. "Deep Residual Learning for Image Recognition." Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, Las Vegas, NV, 770–778.

Henderson, Steven J., and Steven K. Feiner. 2007. *Augmented Reality for Maintenance and Repair (ARMAR)*. Technical Report. Columbia Univ New York Dept of Computer Science.

Hendricks, Lisa Anne, Subhashini Venugopalan, Marcus Rohrbach, Raymond Mooney, Kate Saenko, and Trevor Darrell. 2016. "Deep Compositional Captioning: Describing Novel Object Categories Without Paired Training Data." Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, Las Vegas, NV, 1–10. IEEE.

Heyden, Anders, and Kalle Åkström. 1998. "Minimal Conditions on Intrinsic Parameters for Euclidean Reconstruction." Asian conference on Computer Vision, Hong Kong, 169–176. Springer.

Heyden, Anders, and Kalle Astrom. 1997. "Euclidean Reconstruction from Image Sequences with Varying and Unknown

Focal Length and Principal Point." Proceedings of IEEE computer society conference on Computer Vision and Pattern Recognition, San Juan, Puerto Rico, 438–443. IEEE.

Hinton, Geoffrey E., Nitish Srivastava, Alex Krizhevsky, Ilya Sutskever, and Ruslan R. Salakhutdinov. 2012. "Improving Neural Networks by Preventing Co-adaptation of Feature Detectors." Preprint arXiv:1207.0580.

Hold-Geoffroy, Yannick, Kalyan Sunkavalli, Jonathan Eisenmann, Matthew Fisher, Emiliano Gambaretto, Sunil Hadap, and Jean-François Lalonde. 2018. "A Perceptual Measure for Deep Single Image Camera Calibration." Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, 2354–2363.

Höllerer, Tobias, Steven Feiner, Drexel Hallaway, Blaine Bell, Marco Lanzagorta, Dennis Brown, Simon Julier, Yohan Baillot, and Lawrence Rosenblum. 2001. "User Interface Management Techniques for Collaborative Mobile Augmented Reality." *Computers & Graphics* 25 (5): 799–810.

Holloway, Richard L. 1997. "Registration Error Analysis for Augmented Reality." *Presence: Teleoperators & Virtual Environments* 6 (4): 413–432.

Holm, Magnus, Oscar Danielsson, Anna Syberfeldt, Philip Moore, and Lihui Wang. 2017. "Adaptive Instructions to Novice Shop-floor Operators Using Augmented Reality." *Journal of Industrial and Production Engineering* 34 (5): 362–374.

Holynski, Aleksander, and Johannes Kopf. 2018. "Fast Depth Densification for Occlusion-aware Augmented Reality." *ACM Transactions on Graphics (TOG)* 37 (6): 1–11.

Horn, Berthold K. P., and Michael J. Brooks. 1989. *Shape from Shading*. Vol. 2. Cambridge, MA: MIT Press. ISBN: 9780262519175.

Howard, Andrew G., Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. 2017. "Mobilenets: Efficient Convolutional Neural Networks for Mobile Vision Applications." Preprint arXiv:1704.04861.

Huber, Peter J. 1992. "Robust Estimation of a Location Parameter." In *Breakthroughs in Statistics*, edited by S. Kotz and N. L. Johnson, 492–518. New York, NY: Springer. doi:10.1007/978-1-4612-4380-9_35.

Hung, Wei-Chih, Yi-Hsuan Tsai, Yan-Ting Liou, Yen-Yu Lin, and Ming-Hsuan Yang. 2018. "Adversarial Learning for Semi-supervised Semantic Segmentation." Preprint arXiv:1802.07934.

Huo, Jiage, Jianghua Zhang, and Felix T. S. Chan. 2020. "A Fuzzy Control System for Assembly Line Balancing with a Three-state Degradation Process in the Era of Industry 4.0." *International Journal of Production Research* 58: 1–18.

Iandola, Forrest N., Song Han, Matthew W. Moskewicz, Khalid Ashraf, William J. Dally, and Kurt Keutzer. 2016. "SqueezeNet: AlexNet-level Accuracy with 50× Fewer Parameters and < 0.5 MB Model Size." Preprint arXiv:1602.07360.

Ichihashi, Keita, and Kaori Fujinami. 2019. "Estimating Visibility of Annotations for View Management in Spatial Augmented Reality Based on Machine-Learning Techniques." *Sensors* 19 (4): 939–966.

Irani, Michal, and P. Anandan. 1998. "Robust Multi-sensor Image Alignment." 6th International conference on Computer Vision (IEEE Cat. No. 98CH36271), Bombay, India, 959–966. IEEE.

Ivaschenko, Anton, Michael Milutkin, and Pavel Sitnikov. 2019. "Industrial Application of Accented Visualization Based on Augmented Reality." 24th conference of Open Innovations Association (FRUCT), Moscow, Russia, 123–129. IEEE.

Järvenpää, Eeva, Niko Siltala, Otto Hylli, and Minna Lanz. 2019. "The Development of An Ontology for Describing the Capabilities of Manufacturing Resources." *Journal of Intelligent Manufacturing* 30 (2): 959–978.

Jia, Jianqing, Semir Elezovikj, Heng Fan, Shuojin Yang, Jing Liu, Wei Guo, Chiu C. Tan, and Haibin Ling. 2019. "Semantic-Aware Label Placement for Augmented Reality in Street View." Preprint arXiv:1912.07105.

Jiang, Wei, Juan Camilo Gamboa Higuera, Baptiste Angles, Weiwei Sun, Mehrsan Javan, and Kwang Moo Yi. 2020. "Optimizing through Learned Errors for Accurate Sports Field Registration." The IEEE Winter conference on Applications of Computer Vision, Snowmass Village, Colorado, 201–210.

Jiddi, Salma, Philippe Robert, and Eric Marchand. 2016. "Reflectance and Illumination Estimation for Realistic Augmentations of Real Scenes." IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct), Merida, Yucatan, Mexico, 244–249. IEEE.

Jiddi, Salma, Philippe Robert, and Eric Marchand. 2017. "[POSTER] Illumination Estimation Using Cast Shadows for Realistic Augmented Reality Applications." IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct), Nantes, France, 192–193. IEEE.

Jo, Geun-Sik, Kyeong-Jin Oh, Inay Ha, Kee-Sung Lee, Myung-Duk Hong, Ulrich Neumann, and Suya You. 2014. "A Unified Framework for Augmented Reality and Knowledge-based Systems in Maintaining Aircraft." 26th IAAI Conference, Toronto, Canada, 2990–2997.

Johnson, Stephen C.. 1967. "Hierarchical Clustering Schemes." *Psychometrika* 32 (3): 241–254.

Jordan, Michael I., and Tom M. Mitchell. 2015. "Machine Learning: Trends, Perspectives, and Prospects." *Science* 349 (6245): 255–260.

Julier, Simon, Marco Lanzagorta, Yohan Baillot, Lawrence Rosenblum, Steven Feiner, Tobias Hollerer, and Sabrina Sestito. 2000. "Information Filtering for Mobile Augmented Reality." Proceedings IEEE and ACM International Symposium on Augmented Reality (ISAR 2000), Munich, Germany, 3–11. IEEE.

Kalman, Rudolph Emil. 1960. "A New Approach to Linear Filtering and Prediction Problems." *Journal of Fluids Engineering* 82 (1): 35–45.

Kán, Peter, and Hannes Kaufmann. 2013. "Differential Irradiance Caching for Fast High-quality Light Transport between Virtual and Real Worlds." IEEE International Symposium on Mixed and Augmented Reality (ISMAR), Adelaide, S.A., Australia, 133–141. IEEE.

Kán, Peter, Johannes Unterguggenberger, and Hannes Kaufmann. 2015. "High-quality Consistent Illumination in Mobile Augmented Reality by Radiance Convolution on the GPU." International symposium on Visual Computing, Las Vegas, NV, 574–585. Springer.

Kanade, Takeo, and Masatoshi Okutomi. 1994. "A Stereo Matching Algorithm with An Adaptive Window: Theory and Experiment." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 16 (9): 920–932.

Kanbara, Masayuki, Takashi Okuma, Haruo Takemura, and Naokazu Yokoya. 2000. "A Stereoscopic Video See-through Augmented Reality System based on Real-time Vision-based Registration." Proceedings IEEE Virtual Reality 2000 (Cat. No. 00CB37048), New Brunswick, NJ, 255–262. IEEE.

Kang, Seokho, Eunji Kim, Jaewoong Shim, Wonsang Chang, and Sungzoon Cho. 2018. "Product Failure Prediction with Missing Data." *International Journal of Production Research* 56 (14): 4849–4859.

Kato, Hirokazu, and Mark Billinghurst. 1999. "Marker Tracking and HMD Calibration for a Video-based Augmented Reality Conferencing System." Proceedings 2nd IEEE and ACM International Workshop on Augmented Reality (IWAR'99), Bellevue, WA, 85–94. IEEE.

Ke, Yan, and Rahul Sukthankar. 2004. "PCA-SIFT: A More Distinctive Representation for Local Image Descriptors." Proceedings of the 2004 IEEE computer society conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004, Washington, DC, Vol. 2, II–II. IEEE.

Kendall, Alex, and Roberto Cipolla. 2016. "Modelling Uncertainty in Deep Learning for Camera Relocalization." IEEE International Conference on Robotics and Automation (ICRA), Stockholm, Sweden, 4762–4769. IEEE.

Kendall, Alex, and Roberto Cipolla. 2017. "Geometric Loss Functions for Camera Pose Regression with Deep Learning." Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, Honolulu, HI, 5974–5983.

Kendall, Alex, Matthew Grimes, and Roberto Cipolla. 2015. "Posenet: A Convolutional Network for Real-time 6-DOF Camera Relocalization." Proceedings of the IEEE international conference on Computer Vision, Santiago, Chile, 2938–2946.

Kendall, Alex, Hayk Martirosyan, Saumitro Dasgupta, Peter Henry, Ryan Kennedy, Abraham Bachrach, and Adam Bry. 2017. "End-to-End Learning of Geometry and Context for Deep Stereo Regression." Proceedings of the IEEE international conference on Computer Vision, Venice, Italy, 66–75.

Kieritz, Hilke, Wolfgang Hubner, and Michael Arens. 2018. "Joint Detection and Online Multi-object Tracking." Proceedings of the IEEE conference on Computer Vision and Pattern Recognition Workshops, Salt Lake City, UT, 1459–1467.

Kim, Minyoung, Stefano Alletto, and Luca Rigazio. 2016. "Similarity Mapping with Enhanced Siamese Network for Multi-object Tracking." Preprint arXiv:1609.09156.

Kim, Chanran, Younkyoung Lee, Jong-Il Park, and Jaeha Lee. 2018. "Diminishing Unwanted Objects based on Object Detection using Deep Learning and Image Inpainting." International Workshop on Advanced Image Technology (IWAIT), Chiang Mai, Thailand, 1–3. IEEE.

Kim, Kiyoung, Vincent Lepetit, and Woontack Woo. 2010. "Keyframe-based Modeling and Tracking of Multiple 3D Objects." IEEE international symposium on Mixed and Augmented Reality, Seoul, Korea, 193–198. IEEE.

Kim, Chanho, Fuxin Li, and James M Rehg. 2018. "Multi-object Tracking with Neural Gating using Bilinear LSTM." Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 200–215.

Klein, Georg, and David W. Murray. 2006. "Full-3D Edge Tracking with a Particle Filter." BMVC, Edinburgh, UK, 1119–1128.

Klein, Georg, and David Murray. 2007. "Parallel Tracking and Mapping for Small AR Workspaces." 6th IEEE and ACM international symposium on Mixed and Augmented Reality, Nara, Japan, 225–234. IEEE.

Knorr, Sebastian B., and Daniel Kurz. 2014. "Real-time Illumination Estimation from Faces for Coherent Rendering." IEEE international symposium on Mixed and Augmented Reality (ISMAR), Munich, Germany, 113–122. IEEE.

Kriegel, Hans-Peter, Peer Kröger, Jörg Sander, and Arthur Zimek. 2011. "Density-based Clustering." *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 1 (3): 231–240.

Kristan, Matej, Ales Leonardis, Jiri Matas, Michael Felsberg, Roman Pflugfelder, Luka Cehovin Zajc, and Tomas Vojir. 2018. "The Sixth Visual Object Tracking VOT2018 Challenge Results." Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 3–53.

Kristan, Matej, Jiri Matas, Ales Leonardis, Michael Felsberg, Roman Pflugfelder, Joni-Kristian Kamarainen, and Luka Cehovin Zajc. 2019. "The Seventh Visual Object Tracking VOT2019 Challenge Results." Proceedings of the IEEE international conference on Computer Vision Workshops, Seoul, Korea, 2206–2241.

Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. 2012. "Imagenet Classification with Deep Convolutional Neural Networks." Advances in Neural Information Processing Systems, Lake Tahoe, NV, 1097–1105.

Kruppa, Erwin. 1913. "Zur Ermittlung eines Objektes aus zwei Perspektiven mit innerer Orientierung." *Sitzungsberichte der Mathematisch-Naturwissenschaftlichen Kaiserlichen Akademie der Wissenschaften* 122: 1939–1948.

Kudan Inc. 2020. "Home — Kudan." https://www.kudan.io/.

Kuhn, Harold W. 1955. "The Hungarian Method for the Assignment Problem." *Naval Research Logistics Quarterly* 2 (1–2): 83–97. Accessed 14 May 2020. http://www.bioinfo.org.cn/~dbu/AlgorithmCourses/Lectures/50YearsIP.pdf#page=46.

Kuo, Cheng-Hao, and Ram Nevatia. 2011. "How Does Person Identity Recognition Help Multi-person Tracking?" CVPR 2011, Colorado Springs, CO, 1217–1224. IEEE.

Kutulakos, Kiriakos N., and James R. Vallino. 1998. "Calibra tion-free Augmented Reality." *IEEE Transactions on Visualization and Computer Graphics* 4 (1): 1–20.

Kuznietsov, Yevhen, Jorg Stuckler, and Bastian Leibe. 2017. "Semi-supervised Deep Learning for Monocular Depth Map Prediction." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, 6647–6655.

Laina, Iro, Christian Rupprecht, Vasileios Belagiannis, Federico Tombari, and Nassir Navab. 2016. "Deeper Depth Prediction with Fully Convolutional Residual Networks." 4th international conference on 3D Vision (3DV), Stanford, CA, 239–248. IEEE.

Lalonde, Jean-François. 2018. "Deep Learning for Augmented Reality." 17th Workshop on Information Optics (WIO), Quebec, QC, Canada, 1–3. IEEE.

Lampe, Thomas, and Martin Riedmiller. 2013. "Acquiring Visual Servoing Reaching and Grasping Skills using Neural Reinforcement Learning." International Joint Conference on Neural Networks (IJCNN), Dallas, TX, 1–8. IEEE.

Laskar, Zakaria, Iaroslav Melekhov, Surya Kalia, and Juho Kannala. 2017. "Camera Relocalization by Computing Pairwise

Relative Poses using Convolutional Neural Network." Proceedings of the IEEE international conference on Computer Vision Workshops, Venice, Italy, 929–938.

Leal-Taixé, Laura, Cristian Canton-Ferrer, and Konrad Schindler. 2016. "Learning by Tracking: Siamese CNN for Robust Target Association." Proceedings of the IEEE conference on Computer Vision and Pattern Recognition Workshops, Las Vegas, NV, 33–40.

LeCun, Yann, Léon Bottou, Yoshua Bengio, and Patrick Haffner. 1998. "Gradient-based Learning Applied to Document Recognition." *Proceedings of the IEEE* 86 (11): 2278–2324.

Lee, Taehee, and Tobias Hollerer. 2008. "Hybrid Feature Tracking and user Interaction for Markerless Augmented Reality." IEEE Virtual Reality Conference, Reno, NV, 145–152. IEEE.

Lee, Sangyun, and Euntai Kim. 2018. "Multiple Object Tracking Via Feature Pyramid Siamese Networks." *IEEE Access* 7: 8181–8194.

Lee, Alex X., Sergey Levine, and Pieter Abbeel. 2017. "Learning Visual Servoing with Deep Features and Fitted Q-iteration." Preprint arXiv:1703.11000.

Lee, Jae Yeol, and Guewon Rhee. 2008a. "Context-aware 3D Visualization and Collaboration Services for Ubiquitous Cars Using Augmented Reality." *The International Journal of Advanced Manufacturing Technology* 37 (5–6): 431–442.

Lee, Jae Yeol, and Guewon Rhee. 2008b. "Context-aware 3D Visualization and Collaboration Services for Ubiquitous Cars Using Augmented Reality." *The International Journal of Advanced Manufacturing Technology* 37 (5–6): 431–442.

LeGendre, Chloe, Wan-Chun Ma, Graham Fyffe, John Flynn, Laurent Charbonnel, Jay Busch, and Paul Debevec. 2019. "Deeplight: Learning Illumination for Unconstrained Mobile Mixed Reality." Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, Long Beach, CA, 5918–5928.

Lepetit, Vincent, and Pascal Fua. 2005. "Monocular Model-based 3D Tracking of Rigid Objects." *Foundations and Trends in Computer Graphics and Vision* 1 (1): 1–89. doi:10.1561/0600000001.

Lepetit, Vincent, and Pascal Fua. 2006. "Keypoint Recognition Using Randomized Trees." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 28 (9): 1465–1479.

Lepetit, Vincent, Luca Vacchetti, Daniel Thalmann, and Pascal Fua. 2003. "Fully Automated and Stable Registration for Augmented Reality Applications." Proceedings of the 2nd IEEE and ACM international symposium on Mixed and Augmented Reality, Tokyo, Japan, 93–102. IEEE.

Li, Yang, Julien Amelot, Xin Zhou, Samy Bengio, and Si Si. 2020. "Auto Completion of User Interface Layout Design Using Transformer-Based Tree Decoders." Preprint arXiv:2001.05308.

Li, Yuanyuan, Stefano Carabelli, Edoardo Fadda, Daniele Manerba, Roberto Tadei, and Olivier Terzo. 2020. "Machine Learning and Optimization for Production Rescheduling in Industry 4.0." *The International Journal of Advanced Manufacturing Technology* 110: 2445–2463.

Li, Mengtian, Zhe Lin, Radomir Mech, Ersin Yumer, and Deva Ramanan. 2019. "Photo-sketching: Inferring Contour Drawings from Images." IEEE Winter conference on Applications of Computer Vision (WACV), Waikoloa Village, Hawaii, 1403–1412. IEEE.

Li, Zhengqin, Mohammad Shafiei, Ravi Ramamoorthi, Kalyan Sunkavalli, and Manmohan Chandraker. 2020. "Inverse Rendering for Complex Indoor Scenes: Shape, Spatially-varying Lighting and SVBRDF from a Single Image." Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition, Long Beach, CA, 2475–2484.

Li, Bo, Chunhua Shen, Yuchao Dai, Anton Van Den Hengel, and Mingyi He. 2015. "Depth and Surface Normal Estimation from Monocular Images using Regression on Deep Features and Hierarchical CRFs." Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, Boston, MA, 1119–1127.

Li, Bo, Wei Wu, Qiang Wang, Fangyi Zhang, Junliang Xing, and Junjie Yan. 2019. "Siamrpn++: Evolution of Siamese Visual Tracking with Very Deep Networks." Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, Long Beach, CA, 4282–4291.

Li, Bo, Junjie Yan, Wei Wu, Zheng Zhu, and Xiaolin Hu. 2018. "High Performance Visual Tracking with Siamese Region Proposal Network." Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, 8971–8980.

Li, Jianan, Jimei Yang, Aaron Hertzmann, Jianming Zhang, and Tingfa Xu. 2019. "Layoutgan: Generating Graphic Layouts with Wireframe Discriminators." Preprint arXiv:1901.06767.

Li, Yehao, Ting Yao, Yingwei Pan, Hongyang Chao, and Tao Mei. 2019. "Pointing Novel Objects in Image Captioning." Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, Long Beach, CA, 12497–12506.

Li, Haifeng, Keshu Zhang, and Tao Jiang. 2004. "Minimum Entropy Clustering and Applications to Gene Expression Analysis." Proceedings of the IEEE Computational Systems Bioinformatics Conference, CSB 2004, Stanford, CA, 142–151. IEEE.

Liang, Yiming, and Yue Zhou. 2018. "LSTM Multiple Object Tracker Combining Multiple Cues." 25th IEEE International Conference on Image Processing (ICIP), Athens, Greece, 2351–2355. IEEE.

Lin, Tsung-Yi, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. 2017. "Feature Pyramid Networks for Object Detection." Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, Honolulu, HI, 2117–2125.

Liu, Wei, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C. Berg. 2016. "SSD: Single Shot multibox Detector." European conference on Computer Vision, Amsterdam, The Netherlands, 21–37. Springer.

Liu, Jingshu, and Yuan Li. 2019. "An Image Based Visual Servo Approach with Deep Learning for Robotic Manipulation." Preprint arXiv:1909.07727.

Liu, Daquan, Chengjiang Long, Hongpan Zhang, Hanning Yu, Xinzhi Dong, and Chunxia Xiao. 2020. "ARShadowGAN: Shadow Generative Adversarial Network for Augmented Reality in Single Light Scenes." Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition, Long Beach, CA, 8139–8148.

Liu, Shu, Lu Qi, Haifang Qin, Jianping Shi, and Jiaya Jia. 2018. "Path Aggregation Network for Instance Segmentation." Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, 8759–8768.

Liu, Wei, Andrew Rabinovich, and Alexander C. Berg. 2015. "Parsenet: Looking Wider to See Better." Preprint arXiv:1506.04579.

Liu, Ying, Rahul Rai, Anurag Purwar, Bin He, and Mahesh Mani. 2020. "Machine Learning Applications in Manufacturing." *Journal of Computing and Information Science in Engineering* 20 (2): 020301. doi:10.1115/1.4046427.

Livingston, Mark A., Lawrence J. Rosenblum, Dennis G. Brown, Gregory S. Schmidt, Simon J. Julier, Yohan Baillot, J. Edward Swan, Zhuming Ai, and Paul Maassel. 2011. "Military Applications of Augmented Reality." In *Handbook of Augmented Reality* edited by Borko Furht, 671–706. New York, NY: Springer.

Long, Jonathan, Evan Shelhamer, and Trevor Darrell. 2015. "Fully Convolutional Networks for Semantic Segmentation." Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, Boston, MA, 3431–3440.

Lopez, Manuel, Roger Mari, Pau Gargallo, Yubin Kuang, Javier Gonzalez-Jimenez, and Gloria Haro. 2019. "Deep Single Image Camera Calibration with Radial Distortion." Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, Long Beach, CA, 11817–11825.

Lowe, David G. 1991. "Fitting Parameterized Three-dimensio nal Models to Images." *IEEE Transactions on Pattern Analysis & Machine Intelligence* 13 (5): 441–450.

Lowe, David G. 1999. "Object Recognition from Local Scale-invariant Features." Proceedings of the 7th IEEE international conference on Computer Vision, Kerkyra, Greece, Vol. 2, 1150–1157. IEEE.

Lowe, David G. 2004. "Distinctive Image Features From Scale-invariant Keypoints." *International Journal of Computer Vision* 60 (2): 91–110.

Lu, C.-P., Gregory D. Hager, and Eric Mjolsness. 2000. "Fast and Globally Convergent Pose Estimation From Video Images." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22 (6): 610–622.

Lu, Boun Vinh, Tetsuya Kakuta, Rei Kawakami, Takeshi Oishi, and Katsushi Ikeuchi. 2010. "Foreground and Shadow Occlusion Handling for Outdoor Augmented Reality." IEEE international symposium on Mixed and Augmented Reality, Seoul, Korea, 109–118. IEEE.

Lu, Yongyi, Cewu Lu, and Chi-Keung Tang. 2017. "Online Video Object Detection using Association LSTM." Proceedings of the IEEE international conference on Computer Vision, Venice, Italy, 2344–2352.

Luc, Pauline, Camille Couprie, Soumith Chintala, and Jakob Verbeek. 2016. "Semantic Segmentation using Adversarial Networks." Preprint arXiv:1611.08408.

Lucas, Bruce D., and Takeo Kanade. 1981. "An Iterative Image Registration Technique with an Application to Stereo Vision." Proceedings of the 7th international joint conference on Artificial Intelligence, Vancouver, BC, Canada, Vol. 2, 674–679. Morgan Kaufmann Publishers.

Lytridis, Chris, Avgoustos Tsinakos, and Ioannis Kazanidis. 2018. "ARTutor–An Augmented Reality Platform for Interactive Distance Learning." *Education Sciences* 8 (1): 6.

MacIntyre, Blair, and E. Machado Coelho. 2000. "Adapting to Dynamic Registration Errors using Level of Error (LOE) Filtering." Proceedings IEEE and ACM International Symposium on Augmented Reality (ISAR 2000), Munich, Germany, 85–88. IEEE.

MacIntyre, Blair, Enylton Machado Coelho, and Simon J. Julier. 2002. "Estimating and Adapting to Registration Errors in Augmented Reality Systems." Proceedings IEEE Virtual Reality 2002, Orlando, FL, 73–80. IEEE.

MacKay, David J. C. 1992. "A Practical Bayesian Framework for Backpropagation Networks." *Neural Computation* 4 (3): 448–472.

Maidi, Madjid, Jean-Yves Didier, Fakhreddine Ababsa, and Malik Mallem. 2010. "A Performance Study for Camera Pose Estimation Using Visual Marker Based Tracking." *Machine Vision and Applications*21 (3): 365–376.

Makris, Sotiris, Panagiotis Karagiannis, Spyridon Koukas, and Aleksandros-Stereos Matthaiakis. 2016. "Augmented Reality System for Operator Support in Human–Robot Collaborative Assembly." *CIRP Annals* 65 (1): 61–64.

Malek, S., N. Zenati-Henda, M. Belhocine, and S. Benbelkacem. 2008. "Calibration Method for An Augmented Reality System." *World Academy of Science, Engineering and Technology* 45: 309–314.

Mancini, Michele, Gabriele Costante, Paolo Valigi, and Thomas A. Ciarfuglia. 2016. "Fast Robust Monocular Depth Estimation for Obstacle Detection with Fully Convolutional Networks." IEEE/RSJ international conference on Intelligent Robots and Systems (IROS), Daejeon, Korea, 4296–4303. IEEE.

Mao, Junhua, Xu Wei, Yi Yang, Jiang Wang, Zhiheng Huang, and Alan L. Yuille. 2015. "Learning Like a Child: Fast Novel Visual Concept Learning from Sentence Descriptions of Images." Proceedings of the IEEE international conference on Computer Vision, Santiago, Chile, 2533–2541.

Marchand, Eric, Patrick Bouthemy, and François Chaumette. 2001. "A 2D–3D Model-based Approach to Real-time Visual Tracking." *Image and Vision Computing* 19 (13): 941–955.

Marchand, Éric, and François Chaumette. 2002. "Virtual Visual Servoing: A Framework for Real-Time Augmented Reality." *Computer Graphics Forum* 21: 289–297.

Marchand, Éric, and François Chaumette. 2005. "Feature Tracking for Visual Servoing Purposes." *Robotics and Autonomous Systems* 52 (1): 53–70.

Marques, Bruno Augusto Dorta, Esteban Walter Gonzalez Clua, and Cristina Nader Vasconcelos. 2018. "Deep Spherical Harmonics Light Probe Estimator for Mixed Reality Games." *Computers & Graphics*76: 96–106.

Marques, Bruno A. D., Rafael Rego Drumond, Cristina Nader Vasconcelos, and Esteban Clua. 2018. "Deep Light Source Estimation for Mixed Reality." VISIGRAPP (1: GRAPP), Funchal, Madeira, Portugal, 303–311.

Martinetti, Alberto, Henrique Costa Marques, Sarbjeet Singh, and Leo van Dongen. 2019. "Reflections on the Limited Pervasiveness of Augmented Reality in Industrial Sectors." *Applied Sciences* 9 (16): 3382.

Masood, Tariq, and Johannes Egger. 2020. "Adopting Augmented Reality in the Age of Industrial Digitalisation." *Computers in Industry* 115: 103112.

Matei, Ion, Johan de Kleer, Alexander Feldman, Rahul Rai, and Souma Chowdhury. 2020. "Hybrid Modeling: Applications in Real-time Diagnosis." Preprint arXiv:2003.02671.

Meka, Abhimitra, Maxim Maximov, Michael Zollhoefer, Avishek Chatterjee, Hans-Peter Seidel, Christian Richardt, and Christian Theobalt. 2018. "Lime: Live Intrinsic Material

Estimation." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, 6315–6324.

Melekhov, Iaroslav, Juha Ylioinas, Juho Kannala, and Esa Rahtu. 2017. "Image-based Localization using Hourglass Networks." Proceedings of the IEEE international conference on Computer Vision Workshops, Venice, Italy, 879–886.

Mengoni, Maura, Matteo Iualè, Margherita Peruzzini, and Michele Germani. 2015. "An Adaptable AR User Interface to Face the Challenge of Ageing Workers in Manufacturing." International conference on Human Aspects of IT for the Aged Population, Los Angeles, CA, 311–323. Springer.

Milan, Anton, S. Hamid Rezatofighi, Anthony Dick, Ian Reid, and Konrad Schindler. 2017. "Online Multi-target Tracking using Recurrent Neural Networks." 31st AAAI conference on Artificial Intelligence, San Francisco, CA, 4225–4232.

Min, Erxue, Xifeng Guo, Qiang Liu, Gen Zhang, Jianjing Cui, and Jun Long. 2018. "A Survey of Clustering with Deep Learning: From the Perspective of Network Architecture." *IEEE Access* 6: 39501–39514.

Mohring, Mathias, Christian Lessig, and Oliver Bimber. 2004. "Video See-through AR on Consumer Cell-phones." 3rd IEEE and ACM international symposium on Mixed and Augmented Reality, Arlington, VA, 252–253. IEEE.

Morel, Jean-Michel, and Guoshen Yu. 2009. "ASIFT: A New Framework for Fully Affine Invariant Image Comparison." *SIAM Journal on Imaging Sciences* 2 (2): 438–469.

Moreno-Noguer, Francesc, Vincent Lepetit, and Pascal Fua. 2007. "Accurate Non-Iterative O(n) Solution to the PnP Problem." IEEE 11th international conference on Computer Vision, Rio de Janeiro, Brazil, 1–8. IEEE.

Mortensen, Eric N., Hongli Deng, and Linda Shapiro. 2005. "A SIFT Descriptor with Global Context." IEEE computer society conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, Vol. 1, 184–190. IEEE.

Motoyama, Yuichi, Kazuyo Iwamoto, Hitoshi Tokunaga, and Toshimitsu Okane. 2020. "Measuring Hand-pouring Motion in Casting Process Using Augmented Reality Marker Tracking." *The International Journal of Advanced Manufacturing Technology* 106 (11): 5333–5343.

Mourtzis, Dimitris, Vasilios Zogopoulos, and Fotini Xanthi. 2019. "Augmented Reality Application to Support the Assembly of Highly Customized Products and to Adapt to Production Re-scheduling." *The International Journal of Advanced Manufacturing Technology* 105 (9): 3899–3910.

Munir, Kamran, and M. Sheraz Anjum. 2018. "The Use of Ontologies for Effective Knowledge Modelling and Information Retrieval." *Applied Computing and Informatics* 14 (2): 116–126.

Mur-Artal, Raul, Jose Maria Martinez Montiel, and Juan D. Tardos. 2015. "ORB-SLAM: A Versatile and Accurate Monocular SLAM System." *IEEE Transactions on Robotics* 31 (5): 1147–1163.

Naik, Hemal, Federico Tombari, Christoph Resch, Peter Keitler, and Nassir Navab. 2015. "A Step Closer To Reality: Closed Loop Dynamic Registration Correction in SAR." IEEE international symposium on Mixed and Augmented Reality, Fukuoka, Japan, 112–115. IEEE.

Naimark, Leonid, and Eric Foxlin. 2002. "Circular Data Matrix Fiducial System and Robust Image Processing for a Wearable Vision-inertial Self-tracker." Proceedings. international symposium on Mixed and Augmented Reality, Darmstadt, Germany, 27–36. IEEE.

Nair, Vinod, and Geoffrey E. Hinton. 2010. "Rectified Linear Units Improve Restricted Boltzmann Machines." Proceedings of the 27th International Conference on Machine Learning (ICML-10), Haifa, Israel, 807–814.

Najibi, Mahyar, Mohammad Rastegari, and Larry S. Davis. 2016. "G-CNN: An Iterative Grid based Object Detector." Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, Las Vegas, NV, 2369–2377.

Nam, Hyeonseob, and Bohyung Han. 2016. "Learning Multi-domain Convolutional Neural Networks for Visual Tracking." Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, Las Vegas, NV, 4293–4302.

Naseer, Tayyab, and Wolfram Burgard. 2017. "Deep Regression for Monocular Camera-based 6-DOF Global Localization in Outdoor Environments." IEEE/RSJ international conference on Intelligent Robots and Systems (IROS), Vancouver, BC, Canada, 1525–1530. IEEE.

Neges, Matthias, Mario Wolf, and Michael Abramovici. 2015. "Secure Access Augmented Reality Solution for Mobile Maintenance Support Utilizing Condition-oriented Work Instructions." *Procedia CIRP* 38: 58–62.

Neumann, Ulrich, and Youngkwan Cho. 1996. "A Self-tracking Augmented Reality System." Proceedings of the ACM symposium on Virtual Reality Software and Technology, Hong Kong, 109–115.

Neumann, Ulrich, and Suya You. 1998. "Integration of Region Tracking and Optical Flow for Image Motion Estimation." Proceedings 1998 International Conference on Image Processing, ICIP98 (Cat. No. 98CB36269), Chicago, IL, 658–662. IEEE.

Newcombe, Richard A., and Andrew J. Davison. 2010. "Live Dense Reconstruction with a Single Moving Camera." IEEE computer society conference on Computer Vision and Pattern Recognition, San Francisco, CA, 1498–1505. IEEE.

Nguyen, Tam, Phong Vu, Hung Pham, and Tung Nguyen. 2018. "Deep Learning UI Design Patterns of Mobile Apps." IEEE/ACM 40th International Conference on Software Engineering: New Ideas and Emerging Technologies Results (ICSE-NIER), Gothenburg, Sweden, 65–68. IEEE.

Noh, Hyeonwoo, Seunghoon Hong, and Bohyung Han. 2015. "Learning Deconvolution Network for Semantic Segmentation." Proceedings of the IEEE international conference on Computer Vision, Santiago, Chile, 1520–1528.

Noy, Natalya F., and Deborah L. McGuinness. 2000. "What is an Ontology and Why We Need It." Accessed 10 May 2020. https://protege.stanford.edu/publications/ontology_development/ontology101-noy-mcguinness.html.

Oberkampf, Denis, Daniel F. DeMenthon, and Larry S. Davis. 1996. "Iterative Pose Estimation Using Coplanar Feature Points." *Computer Vision and Image Understanding* 63 (3): 495–511.

O'Mahony, Niall, Sean Campbell, Anderson Carvalho, Suman Harapanahalli, Gustavo Velasco Hernandez, Lenka Krpalkova, Daniel Riordan, and Joseph Walsh. 2019. "Deep Learning vs. Traditional Computer Vision." Science and Information Conference, Las Vegas, NV, 128–144. Springer.

Ong, S. K., M. L. Yuan, and A. Y. C. Nee. 2008. "Augmented Reality Applications in Manufacturing: A Survey." *International Journal of Production Research* 46 (10): 2707–2742.

Osti, Francesco, Alessandro Ceruti, Alfredo Liverani, and Gianni Caligiana. 2017. "Semi-automatic Design for Disassembly Strategy Planning: An Augmented Reality Approach." *Procedia Manufacturing* 11: 1481–1488.

Ozuysal, Mustafa, Michael Calonder, Vincent Lepetit, and Pascal Fua. 2009. "Fast Keypoint Recognition Using Random Ferns." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32 (3): 448–461.

Ozuysal, Mustafa, Pascal Fua, and Vincent Lepetit. 2007. "Fast Keypoint Recognition in Ten Lines of Code." IEEE conference on Computer Vision and Pattern Recognition, Minneapolis, MN, 1–8. IEEE.

Palmer, Claire, Zahid Usman, Osiris Canciglieri Junior, Andreia Malucelli, and Robert I. M. Young. 2018. "Interoperable Manufacturing Knowledge Systems." *International Journal of Production Research* 56 (8): 2733–2752.

Panagopoulos, Alexandros, Chaohui Wang, Dimitris Samaras, and Nikos Paragios. 2011. "Illumination Estimation and Cast Shadow Detection through a Higher-order Graphical Model." CVPR 2011, Colorado Springs, CO, 673–680. IEEE.

Pang, Yanwei, Wei Li, Yuan Yuan, and Jing Pan. 2012. "Fully Affine Invariant SURF for Image Matching." *Neurocomputing* 85: 6–10.

Pang, Y., M. L. Yuan, Andrew Y. C. Nee, Soh-Khim Ong, and Kamal Youcef-Toumi. 2006. "A Markerless Registration Method for Augmented Reality based on Affine Properties." Proceedings of the 7th Australasian User Interface Conference, Hobart, Tasmania, Vol. 50, 25–32. Citeseer.

Panin, Giorgio, and Alois Knoll. 2008. "Mutual Information-based 3D Object Tracking." *International Journal of Computer Vision* 78 (1): 107–118.

Park, Kyeong-Beom, Minseok Kim, Sung Ho Choi, and Jae Yeol Lee. 2020. "Deep Learning-based Smart Task Assistance in Wearable Augmented Reality." *Robotics and Computer-Integrated Manufacturing*63: 101887.

Park, Youngmin, Vincent Lepetit, and Woontack Woo. 2008. "Multiple 3D Object Tracking for Augmented Reality." 7th IEEE/ACM international symposium on Mixed and Augmented Reality, Cambridge, UK, 117–120. IEEE.

Park, Joonsuk, and Jun Park. 2010. "3DOF Tracking Accuracy Improvement for Outdoor Augmented Reality." IEEE international symposium on Mixed and Augmented Reality, Seoul, Korea, 263–264. IEEE.

Park, Jun, Suya You, and Ulrich Neumann. 1999. "Natural Feature Tracking for Extendible Robust Augmented Realities." Proceedings of the international workshop on Augmented Reality: Placing Artificial Objects in Real Scenes: Placing Artificial Objects in Real Scenes, Bellevue, WA, 209–217. AK Peters, Ltd.

Paszke, Adam, Abhishek Chaurasia, Sangpil Kim, and Eugenio Culurciello. 2016. "Enet: A Deep Neural Network Architecture for Real-time Semantic Segmentation." Preprint arXiv:1606.02147.

Patel, Maitri, Paresh Virparia, and Dharmendra Patel. 2012. "Web Based Fuzzy Expert System and its Applications: A Survey." *International Journal of Applied Information Systems* 1 (7): 11–15.

Peniche, Amaury, Christian Diaz, Helmuth Trefftz, and Gabriel Paramo. 2012. "Combining Virtual and Augmented Reality to Improve the Mechanical Assembly Training Process in Manufacturing." Proceedings of the 6th WSEAS international conference on Computer Engineering and Applications, and Proceedings of the 2012 American conference on Applied Mathematics, Cambridge, MA, 292–297. World Scientific and Engineering Academy and Society (WSEAS).

Piekarski, Wayne, and Bruce Thomas. 2002. "ARQuake: The Outdoor Augmented Reality Gaming System." *Communications of the ACM* 45 (1): 36–38.

Pilet, Julien, and Hideo Saito. 2010. "Virtually Augmenting Hundreds of Real Pictures: An Approach based on Learning, Retrieval, and Tracking." IEEE Virtual Reality Conference (VR), Waltham, MA, 71–78. IEEE.

Pinheiro, Pedro H. O., and Ronan Collobert. 2014. "Recurrent Convolutional Neural Networks for Scene Labeling." 31st International Conference on Machine Learning (ICML), Beijing, China, Vol I, 82–90.

Pinheiro, Pedro O. O., Ronan Collobert, and Piotr Dollár. 2015. "Learning to Segment Object Candidates." Advances in Neural Information Processing Systems, Montreal, Canada, 1990–1998.

Pinheiro, Pedro O., Tsung-Yi Lin, Ronan Collobert, and Piotr Dollár. 2016. "Learning to Refine Object Segments." European conference on Computer Vision, Amsterdam, The Netherlands, 75–91. Springer.

Pollefeys, Marc, Reinhard Koch, and Luc Van Gool. 1999. "Self-calibration and Metric Reconstruction Inspite of Varying and Unknown Intrinsic Camera Parameters." *International Journal of Computer Vision* 32 (1): 7–25.

Pontes, Jhony K., Chen Kong, Sridha Sridharan, Simon Lucey, Anders Eriksson, and Clinton Fookes. 2018. "Image2Mesh: A Learning Framework for Single Image 3D Reconstruction." Asian Conference on Computer Vision, Perth, WA, Australia, 365–381. Springer.

Porter, Michael E., and James E. Heppelmann. 2017. "A Manager's Guide to Augmented Reality." November–December. Accessed 27 May 2020. https://hbr.org/2017/11/a-managers-guide-to-augmented-reality.

Pressigout, Muriel, and Eric Marchand. 2004. "Model-free Augmented Reality by Virtual Visual Servoing." Proceedings of the 17th International Conference on Pattern Recognition, ICPR 2004, Cambridge, England, UK, Vol. 2, 887–890. IEEE.

PTC. 2020. "Vuforia: Market-Leading Enterprise AR — PTC." Accessed 19 May 2020. https://www.ptc.com/en/products/augmented-reality/vuforia.

Pupilli, Mark, and Andrew Calway. 2006. "Real-time Camera Tracking Using Known 3D Models and a Particle Filter." 18th International Conference on Pattern Recognition (ICPR'06), Hong Kong, Vol. 1, 199–203. IEEE.

Qin, S. Joe, and Leo H. Chiang. 2019. "Advances and Opportunities in Machine Learning for Process Data Analytics." *Computers & Chemical Engineering* 126: 465–473.

Quandt, Moritz, Benjamin Knoke, Christian Gorldt, Michael Freitag, and Klaus-Dieter Thoben. 2018. "General Requirements for Industrial Augmented Reality Applications." *Procedia Cirp* 72 (1): 1130–1135.

Radu, Iulian. 2012. "Why Should My Students Use AR? A Comparative Review of the Educational Impacts of Augmented-reality." IEEE International Symposium on Mixed and Augmented Reality (ISMAR), Atlanta, GA, 313–314. IEEE.

Radwan, Noha, Abhinav Valada, and Wolfram Burgard. 2018. "Vlocnet++: Deep Multitask Learning for Semantic Visual Localization and Odometry." *IEEE Robotics and Automation Letters* 3 (4): 4407–4414.

Rai, Rahul, and Akshay V. Deshpande. 2016. "Fragmentary Shape Recognition: A BCI Study." *Computer-Aided Design* 71: 51–64.

Rai, Rahul, and Chandan K. Sahu. 2020. "Driven by Data Or Derived Through Physics? A Review of Hybrid Physics Guided Machine Learning Techniques With Cyber-Physical System (CPS) Focus." *IEEE Access* 8: 71050–71073.

Rao, Jinmeng, Yanjun Qiao, Fu Ren, Junxing Wang, and Qingyun Du. 2017. "A Mobile Outdoor Augmented Reality Method Combining Deep Learning Object Detection and Spatial Relationships for Geovisualization." *Sensors* 17 (9): 1951.

Redmon, Joseph, Santosh Divvala, Ross Girshick, and Ali Farhadi. 2016. "You Only Look Once: Unified, Real-time Object Detection." Proceedings of the IEEE conference on Computer Vision and Pattern recognition, Las Vegas, NV, 779–788.

Redmon, Joseph, and Ali Farhadi. 2017. "YOLO9000: Better, Faster, Stronger." Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, Honolulu, HI, 7263–7271.

Redmon, Joseph, and Ali Farhadi. 2018. "YOLOv3: An Incremental Improvement." Preprint arXiv:1804.02767.

Regenbrecht, Holger, Gregory Baratoff, and Wilhelm Wilke. 2005. "Augmented Reality Projects in the Automotive and Aerospace Industries." *IEEE Computer Graphics and Applications* 25 (6): 48–56.

Reid, Donald. 1979. "An Algorithm for Tracking Multiple Targets." *IEEE Transactions on Automatic Control* 24 (6): 843–854.

Reitinger, Bernhard, Christopher Zach, and Dieter Schmalstieg. 2007. "Augmented Reality Scouting for Interactive 3D Reconstruction." IEEE Virtual Reality Conference, Charlotte, NC, 219–222. IEEE.

Reitmayr, Gerhard, and Tom W. Drummond. 2006. "Going Out: Robust Model-based Tracking for Outdoor Augmented Reality." IEEE/ACM international symposium on Mixed and Augmented Reality, Santa Barbara, CA, 109–118. IEEE.

Reitmayr, Gerhard, Ethan Eade, and Tom W. Drummond. 2007. "Semi-automatic Annotations in Unknown Environments." 6th IEEE and ACM international symposium on Mixed and Augmented Reality, Nara, Japan, 67–70. IEEE.

Reitmayr, Gerhard, Tobias Langlotz, Daniel Wagner, Alessandro Mulloni, Gerhard Schall, Dieter Schmalstieg, and Qi Pan. 2010. "Simultaneous Localization and Mapping for Augmented Reality." International symposium on Ubiquitous Virtual Reality, Gwangji, South Korea, 5–8. IEEE.

Reitmayr, Gerhard, and Dieter Schmalstieg. 2003. "Data Management Strategies for Mobile Augmented Reality." Proceedings of the international workshop on Software Technology for Augmented Reality Systems, Tokyo, Japan, 47–52.

Ren, Shaoqing, Kaiming He, Ross Girshick, and Jian Sun. 2015. "Faster R-CNN: Towards Real-time Object Detection with Region Proposal Networks." Advances in Neural Information Processing Systems, Montreal, Canada, 91–99.

Ren, Liangliang, Jiwen Lu, Zifeng Wang, Qi Tian, and Jie Zhou. 2018. "Collaborative Deep Reinforcement Learning for Multi-object Tracking." Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 586–602.

Rentzos, Loukas, Stergios Papanastasiou, Nikolaos Papakostas, and George Chryssolouris. 2013. "Augmented Reality for Human-based Assembly: Using Product and Process Seman tics." *IFAC Proceedings Volumes* 46 (15): 98–101.

Riaz Muhammad, Umar, Yongxin Yang, Yi-Zhe Song, Tao Xiang, and Timothy M. Hospedales. 2018. "Learning Deep Sketch Abstraction." Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, 8014–8023.

Rogério, Richa, Raphael Sznitman, Russell Taylor, and Gregory Hager. 2011. "Visual Tracking using the Sum of Conditional Variance." IEEE/RSJ international conference on Intelligent Robots and Systems, San Francisco, CA, 2953–2958. IEEE.

Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. 2015. "U-net: Convolutional Networks for Biomedical Image Segmentation." International conference on Medical Image Computing and Computer-assisted Intervention, Munich, Germany, 234–241. Springer.

Rosello, Pol, and Mykel J. Kochenderfer. 2018. "Multi-agent Reinforcement Learning for Multi-object Tracking." Proceedings of the 17th international conference on Autonomous Agents and MultiAgent Systems, Munich, Germany, 1397–1404. International Foundation for Autonomous Agents and Multiagent Systems.

Rousseeuw, Peter J., and Annick M. Leroy. 2005. *Robust Regression and Outlier Detection*. Vol. 589. Hoboken, NJ: John Wiley & Sons. doi:10.1002/0471725382.

Roy, Anirban, and Sinisa Todorovic. 2016. "A Multi-scale CNN for Affordance Segmentation in RGB Images." European Conference on Computer Vision, Amsterdam, The Netherlands, 186–201. Springer.

Rüßmann, Michael, Markus Lorenz, Philipp Gerbert, Manuela Waldner, Jan Justus, Pascal Engel, and Michael Harnisch. 2015. "Industry 4.0: The Future of Productivity and Growth in Manufacturing Industries." *Boston Consulting Group* 9 (1): 54–89. Accessed 14 May 2020. https://www.bcg.com/publications/2015/engineered_products_project_business_industry_4_future_productivity_growth_manufacturing_industries.

Sadeghian, Amir, Alexandre Alahi, and Silvio Savarese. 2017. "Tracking the Untrackable: Learning to Track Multiple Cues with Long-term Dependencies." Proceedings of the IEEE international conference on Computer Vision, Venice, Italy, 300–311.

Sampedro, Carlos, Alejandro Rodriguez-Ramos, Ignacio Gil, Luis Mejias, and Pascual Campoy. 2018. "Image-Based Visual Servoing Controller for Multirotor Aerial Robots Using Deep Reinforcement Learning." IEEE/RSJ international conference on Intelligent Robots and Systems (IROS), Madrid, Spain, 979–986. IEEE.

Sanches, Silvio R. R., Daniel M. Tokunaga, Valdinei F. Silva, Antonio C. Sementille, and Romero Tori. 2012. "Mutual Occlusion between Real and Virtual Elements in Augmented Reality based on Fiducial Markers." IEEE Workshop on the Applications of Computer Vision (WACV), Breckenridge, CO, 49–54. IEEE.

Sanderson, David, Jack C. Chaplin, and Svetan Ratchev. 2019. "A Function-Behaviour-Structure Design Methodology for Adaptive Production Systems." *The International Journal of Advanced Manufacturing Technology* 105 (9): 3731–3742.

Sandler, Mark, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. 2018. "MobileNetV2: Inverted Residuals and Linear Bottlenecks." Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, 4510–4520.

Sanna, Andrea, Federico Manuri, Fabrizio Lamberti, Gianluca Paravati, and Pietro Pezzolla. 2015. "Using Handheld Devices to Support Augmented Reality-based Maintenance and Assembly Tasks." IEEE International Conference on Consumer Electronics (ICCE), Las Vegas, NV, 178–179. IEEE.

Sarlin, Paul-Edouard, Cesar Cadena, Roland Siegwart, and Marcin Dymczyk. 2019. "From Coarse to Fine: Robust Hierarchical Localization at Large Scale." Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, Long Beach, CA, 12716–12725.

Sato, Imari, Yoichi Sato, and Katsushi Ikeuchi. 2003. "Illumination From Shadows." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 25 (3): 290–300.

Sattler, Torsten, Qunjie Zhou, Marc Pollefeys, and Laura Leal-Taixe. 2019. "Understanding the Limitations of CNN-based Absolute Camera Pose Regression." Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, Long Beach, CA, 3302–3312.

Saxena, Aseem, Harit Pandya, Gourav Kumar, Ayush Gaud, and K. Madhava Krishna. 2017. "Exploring Convolutional Networks for End-to-End Visual Servoing." IEEE International Conference on Robotics and Automation (ICRA), Mariana Bay Sands, Singapore, 3817–3823. IEEE.

Schall, Gerhard, Helmut Grabner, Michael Grabner, Paul Wohlhart, Dieter Schmalstieg, and Horst Bischof. 2008. "3D Tracking in Unknown Environments using On-line Keypoint Learning for Mobile Augmented Reality." IEEE computer society conference on Computer Vision and Pattern Recognition Workshops, Anchorage, AK, 1–8. IEEE.

Schmalstieg, Dieter, and Daniel Wagner. 2007. "Experiences with Handheld Augmented Reality." 6th IEEE and ACM international symposium on Mixed and Augmented Reality, Nara, Japan, 3–18. IEEE.

Scholz, Joachim, and Andrew N. Smith. 2016. "Augmented Reality: Designing Immersive Experiences that Maximize Consumer Engagement." *Business Horizons* 59 (2): 149–161.

Schwerdtfeger, Björn, Daniel Pustka, Andreas Hofhauser, and Gudrun Klinker. 2008. "Using Laser Projectors for Augmented Reality." Proceedings of the 2008 ACM symposium on Virtual Reality Software and Technology, Bordeaux, France, 134–137.

Seo, Yongduek, Min-Ho Ahn, and Ki Sang Hong. 1998. "Video Augmentation by Image-based Rendering Under the Perspective Camera Model." Proceedings of the 14th international conference on Pattern Recognition (Cat. No. 98EX170), Brisbane, Queensland, Australia, Vol. 2, 1694–1696. IEEE.

Seo, Yongduek, and Ki Sang Hong. 2000. "Calibration-free Augmented Reality in Perspective." *IEEE Transactions on Visualization and Computer Graphics* 6 (4): 346–359.

Sermanet, Pierre, David Eigen, Xiang Zhang, Michaël Mathieu, Rob Fergus, and Yann LeCun. 2013. "Overfeat: Integrated Recognition, Localization and Detection using Convolutional Networks." Preprint arXiv:1312.6229.

Setti, Amedeo, Paolo Bosetti, and Matteo Ragni. 2016. "ARTool-Augmented Reality Platform for Machining Setup and Maintenance." Proceedings of SAI Intelligent Systems Conference, London, UK, 457–475. Springer.

Shen, Zhiqiang, Zhuang Liu, Jianguo Li, Yu-Gang Jiang, Yurong Chen, and Xiangyang Xue. 2017. "DSOD: Learning Deeply Supervised Object Detectors from Scratch." Proceedings of the IEEE international conference on Computer Vision, Venice, Italy, 1919–1927.

Shim, Hyunjung. 2012. "Faces As Light Probes for Relighting." *Optical Engineering* 51 (7): 077002.

Siew, C. Y., S. K. Ong, and A. Y. C. Nee. 2019. "A Practical Augmented Reality-assisted Maintenance System Framework for Adaptive User Support." *Robotics and Computer-Integrated Manufacturing* 59: 115–129.

Silberman, Nathan, Derek Hoiem, Pushmeet Kohli, and Rob Fergus. 2012. "Indoor Segmentation and Support Inference from RGBD Images." European conference on Computer Vision, Firenze, Italy, 746–760. Springer.

Simões, Francisco Paulo Magalhãws, Rafael Alves Roberto, Lucas Silva Figueiredo, João Paulo Silva do Monte Lima, Mozart William Almeida, and Veronica Teichrieb. 2013. "3D Tracking in Industrial Scenarios: A Case Study at the ISMAR Tracking Competition." XV symposium on Virtual and Augmented Reality, Cuiaba, Brazil, 97–106. IEEE.

Simon, Gilles. 2011. "Tracking-by-Synthesis using Point Features and Pyramidal Blurring." 10th IEEE international symposium on Mixed and Augmented Reality, Basel, Switzerland, 85–92. IEEE.

Simonyan, Karen, and Andrew Zisserman. 2014. "Very Deep Convolutional Networks for Large-scale Image Recognition." Preprint arXiv:1409.1556.

Skrypnyk, Iryna, and David G. Lowe. 2004. "Scene Modelling, Recognition and Tracking with Invariant Image Features." 3rd IEEE and ACM international symposium on Mixed and Augmented Reality, Arlington, VA, 110–119. IEEE.

Song, Shuran, and Thomas Funkhouser. 2019. "Neural Illumination: Lighting Prediction for Indoor Environments." Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, Long Beach, CA, 6918–6926.

Song, Jifei, Kaiyue Pang, Yi-Zhe Song, Tao Xiang, and Timothy M. Hospedales. 2018. "Learning to Sketch with Shortcut Cycle Consistency." Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, 801–810.

Souly, Nasim, Concetto Spampinato, and Mubarak Shah. 2017. "Semi supervised semantic segmentation using generative adversarial network." Proceedings of the IEEE international conference on Computer Vision, Venice, Italy, 5688–5696.

Srinivasan, Pratul P., Ben Mildenhall, Matthew Tancik, Jonathan T. Barron, Richard Tucker, and Noah Snavely. 2020. "Lighthouse: Predicting Lighting Volumes for Spatially-Coherent Illumination." Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition, Long Beach, CA, 8080–8089.

Steinbis, John, William Hoff, and Tyrone L. Vincent. 2008. "3D Fiducials for Scalable AR Visual Tracking." 7th IEEE/ACM international symposium on Mixed and Augmented Reality, Cambridge, UK, 183–184. IEEE.

Stocker, Cosima, Marc Schmid, and Gunther Reinhart. 2019. "Reinforcement Learning–based Design of Orienting Devi

ces for Vibratory Bowl Feeders." *The International Journal of Advanced Manufacturing Technology* 105 (9): 3631–3642.
Subakti, Hanas, and Jehn-Ruey Jiang. 2018. "Indoor Augmented Reality using Deep Learning for Industry 4.0 Smart Factories." IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC), Tokyo, Japan, Vol. 2, 63–68. IEEE.
Sundareswaran, Venkataraman, and Reinhold Behringer. 1999. "Visual Servoing-based Augmented Reality." IWAR'98: Proceedings of the International Workshop on Augmented Reality: Placing Artificial Objects in Real Scenes, Bellevue, WA, 193–200. Natick, MA: AK Peters, Ltd.
Sutherland, Ivan E. 1968. "A Head-mounted Three Dimensional Display." Proceedings of the December 9–11, 1968, Fall Joint Computer Conference, San Francisco, CA, Part I, 757–764.
Swan, J. Edward, Adam Jones, Eric Kolstad, Mark A. Livingston, and Harvey S. Smallman. 2007. "Egocentric Depth Judgments in Optical, See-through Augmented Reality." *IEEE Transactions on Visualization and Computer Graphics* 13 (3): 429–442.
Syberfeldt, Anna, Oscar Danielsson, and Patrik Gustavsson. 2017. "Augmented Reality Smart Glasses in the Smart Factory: Product Evaluation Guidelines and Review of Available Products." *IEEE Access*5: 9118–9130.
Syberfeldt, Anna, Oscar Danielsson, Magnus Holm, and Lihui Wang. 2016. "Dynamic Operator Instructions based on Augmented Reality and Rule-based Expert Systems." CIRP CMS 2015, 48th CIRP Conference on Manufacturing Systems, Research and Innovation in Manufacturing: Key Enabling Technologies for the Factories of the Future, Naples, Italy, 24–26 June 2015, Ischia (Naples), Italy, Vol. 41, 346–351. Elsevier.
Szegedy, Christian, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. 2015. "Going Deeper with Convolutions." Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, Boston, MA, 1–9.
Szegedy, Christian, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. 2016. "Rethinking the Inception Architecture for Computer Vision." Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, Las Vegas, NV, 2818–2826.
Takacs, Gabriel, Vijay Chandrasekhar, Natasha Gelfand, Yingen Xiong, Wei-Chao Chen, Thanos Bismpigiannis, Radek Grzeszczuk, Kari Pulli, and Bernd Girod. 2008. "Outdoors Augmented Reality on Mobile Phone using Loxel-based Visual Feature Organization." Proceedings of the 1st ACM international conference on Multimedia Information Retrieval, Vancouver, BC, Canada, 427–434.
Taketomi, Takafumi, Kazuya Okada, Goshiro Yamamoto, Jun Miyazaki, and Hirokazu Kato. 2014. "Camera Pose Estimation Under Dynamic Intrinsic Parameter Change for Augmented Reality." *Computers & Graphics* 44: 11–19.
Tang, Arthur, Charles Owen, Frank Biocca, and Weimin Mou. 2003. "Comparative Effectiveness of Augmented Reality in Object Assembly." Proceedings of the SIGCHI conference on Human Factors in Computing Systems, Ft. Lauderdale, FL, 73–80.
Tao, Ran, Efstratios Gavves, and Arnold W. M. Smeulders. 2016. "Siamese Instance Search for Tracking." Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, Las Vegas, NV, 1420–1429.
Tao, Fei, Qinglin Qi, Ang Liu, and Andrew Kusiak. 2018. "Data-driven Smart Manufacturing." *Journal of Manufacturing Systems* 48: 157–169.
Tatzgern, Markus, Valeria Orso, Denis Kalkofen, Giulio Jacucci, Luciano Gamberini, and Dieter Schmalstieg. 2016. "Adaptive Information Density for Augmented Reality Displays." IEEE Virtual Reality (VR), Greenville, SC, 83–92. IEEE.
Teichrieb, Veronica, J. P. S. M. Lima, E. Apolinário, T. S. M. C. Farias, Márcio Bueno, Judith Kelner, and Ismael Santos. 2007. "A Survey of Online Monocular Markerless Augmented Reality." *International Journal of Modeling and Simulation for the Petroleum Industry* 1 (1): 1–7.
Teng, Chin-Hung, and Bing-Shiun Wu. 2012. "Developing QR Code based Augmented Reality using SIFT Features." 9th international conference on Ubiquitous Intelligence and Computing and 9th international conference on Autonomic and Trusted Computing, Fukuoka, Japan, 985–990. IEEE.
Terenzi, Graziano, and Giuseppe Basile. 2013. "Smart Maintenance." In *An Augmented Reality Platform for Training and Field Operations in the Manufacturing Industry*. ARMedia, 2nd White Paper of Inglobe Technologies Srl.
Thomas, P. C., and W. M. David. 1992. "Augmented Reality: An Application of Heads-up Display Technology to Manual Manufacturing Processes." Hawaii international conference on System Sciences, Kauai, HI, 659–669.
tom Dieck, M. Claudia, and Timothy Jung. 2018. "A Theoretical Model of Mobile Augmented Reality Acceptance in Urban Heritage Tourism." *Current Issues in Tourism* 21 (2): 154–174.
Tonnis, Marcus. 2003. "Data Management for Augmented Reality Applications." Master's thesis, Technische Universitat Munchen.
Toro, Carlos, Javier Vaquero, and Jorge Posada. 2009. "User Experience, Ambient Intelligence and Virtual Reality in an Industrial Maintenance domain using Protégé.", Amsterdam, The Netherlands.
Tosi, Fabio, Filippo Aleotti, Matteo Poggi, and Stefano Mattoccia. 2019. "Learning Monocular Depth Estimation Infusing Traditional Stereo Knowledge." Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, Long Beach, CA, 9799–9809.
Tuceryan, Mihran, Yakup Genc, and Nassir Navab. 2002. "Single-point Active Alignment Method (SPAAM) for Optical See-through HMD Calibration for Augmented Reality." *Presence: Teleoperators & Virtual Environments, Munich, Germany* 11 (3): 259–276.
Uchiyama, Hideaki, and Eric Marchand. 2012. "Object Detection and Pose Tracking for Augmented Reality: Recent Approaches." 18th Korea-Japan Joint Workshop on Frontiers of Computer Vision (FCV) Conference, Kawasaki, Japan, 1–8.
Uijlings, Jasper R. R., Koen E. A. Van De Sande, Theo Gevers, and Arnold W. M. Smeulders. 2013. "Selective Search for Object Recognition." *International Journal of Computer Vision* 104 (2): 154–171.
Uva, Antonio E., Michele Gattullo, Vito M. Manghisi, Daniele Spagnulo, Giuseppe L. Cascella, and Michele Fiorentino. 2018. "Evaluating the Effectiveness of Spatial Augmented Reality in Smart Manufacturing: A Solution for Manual

Working Stations." *The International Journal of Advanced Manufacturing Technology* 94 (1–4): 509–521.
Vacchetti, Luca, Vincent Lepetit, and Pascal Fua. 2004. "Combining Edge and Texture Information for Real-time Accurate 3D Camera Tracking." 3rd IEEE and ACM international symposium on Mixed and Augmented Reality, Arlington, VA, 48–56. IEEE.
Valada, Abhinav, Noha Radwan, and Wolfram Burgard. 2018. "Deep Auxiliary Learning for Visual Localization and Odometry." IEEE International Conference on Robotics and Automation (ICRA), Brisbane, Australia, 6939–6946 . IEEE.
Van Krevelen, D., and R. Poelman. 2007. "Augmented Reality: Technologies, Applications, and Limitations." Vrije Univ. Amsterdam, Dep. Comput. Sci., 1–25. doi:10.13140/RG.2.1.1874.7929.
Venugopalan, Subhashini, Lisa Anne Hendricks, Marcus Rohrbach, Raymond Mooney, Trevor Darrell, and Kate Saenko. 2017. "Captioning Images with Diverse Objects." Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, Honolulu, HI, 5753–5761.
Viola, Paul, and William M. Wells III. 1997. "Alignment by Maximization of Mutual Information." *International Journal of Computer Vision* 24 (2): 137–154.
Virkkunen, Anja. 2018. "Automatic Speech Recognition for the Hearing Impaired in an Augmented Reality Application." Master's thesis, School of Science, Aalto University.
Visin, Francesco, Marco Ciccone, Adriana Romero, Kyle Kastner, Kyunghyun Cho, Yoshua Bengio, Matteo Matteucci, and Aaron Courville. 2016. "Reseg: A Recurrent Neural Network-based Model for Semantic Segmentation." Proceedings of the IEEE conference on Computer Vision and Pattern Recognition Workshops, Las Vegas, NV, 41–48.
Visin, Francesco, Kyle Kastner, Kyunghyun Cho, Matteo Matteucci, Aaron Courville, and Yoshua Bengio. 2015. "Renet: A Recurrent Neural Network based Alternative to Convolutional Networks." Preprint arXiv:1505.00393.
Wagner, Daniel, Gerhard Reitmayr, Alessandro Mulloni, Tom Drummond, and Dieter Schmalstieg. 2008. "Pose Tracking from Natural Features on Mobile Phones." 7th IEEE/ACM international symposium on Mixed and Augmented Reality, Cambridge, UK, 125–134. IEEE.
Wagner, Daniel, Gerhard Reitmayr, Alessandro Mulloni, Tom Drummond, and Dieter Schmalstieg. 2009. "Real-time Detection and Tracking for Augmented Reality on Mobile Phones." *IEEE Transactions on Visualization and Computer Graphics* 16 (3): 355–368.
Wagner, Daniel, and Dieter Schmalstieg. 2007. "ARToolKitPlus for POSE Tracking on Mobile Devices." Proceedings of 12th Computer Vision Winter Workshop, St. Lambrecht, Austria.
Walch, Florian, Caner Hazirbas, Laura Leal-Taixe, Torsten Sattler, Sebastian Hilsenbeck, and Daniel Cremers. 2017. "Image-based Localization using LSTMs for Structured Feature Correlation." Proceedings of the IEEE international conference on Computer Vision, Venice, Italy, 627–637.
Wan, Xingyu, Jinjun Wang, and Sanping Zhou. 2018. "An Online and Flexible Multi-object Tracking Framework using Long Short-term Memory." Proceedings of the IEEE conference on Computer Vision and Pattern Recognition Workshops, Venice, Italy, 1230–1238.
Wang, Chaoyang, José Miguel Buenaposada, Rui Zhu, and Simon Lucey. 2018. "Learning Depth from Monocular Videos using Direct Methods." Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, 2022–2030.
Wang, Panqu, Pengfei Chen, Ye Yuan, Ding Liu, Zehua Huang, Xiaodi Hou, and Garrison Cottrell. 2018. "Understanding Convolution for Semantic Segmentation." IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Tahoe, NV, 1451–1460. IEEE.
Wang, Naiyan, Siyi Li, Abhinav Gupta, and Dit-Yan Yeung. 2015. "Transferring Rich Feature Hierarchies for Robust Visual Tracking." Preprint arXiv:1501.04587.
Wang, Ting-Chun, Ming-Yu Liu, Jun-Yan Zhu, Andrew Tao, Jan Kautz, and Bryan Catanzaro. 2018. "High-resolution Image Synthesis and Semantic Manipulation with Conditional Gans." Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, 8798–8807.
Wang, Z. B., L. X. Ng, S. K. Ong, and A. Y. C. Nee. 2013. "Assembly Planning and Evaluation in An Augmented Reality Environment." *International Journal of Production Research* 51 (23–24): 7388–7404.
Wang, Z. B., S. K. Ong, and A. Y. C. Nee. 2013. "Augmented Reality Aided Interactive Manual Assembly Design." *The International Journal of Advanced Manufacturing Technology* 69 (5–8): 1311–1321.
Wang, X., S. K. Ong, and A. Y. C. Nee. 2014. "Augmented Reality Interfaces for Industrial Assembly Design and Planning." Proceedings of Eighth International Conference on Interfaces and Human Computer Interaction, Lisbon, Portugal, 83–90.
Wang, Lijun, Wanli Ouyang, Xiaogang Wang, and Huchuan Lu. 2015. "Visual Tracking with Fully Convolutional Networks." Proceedings of the IEEE international conference on Computer Vision, Santiago, Chile, 3119–3127.
Wang, Li, Nam Trung Pham, Tian-Tsong Ng, Gang Wang, Kap Luk Chan, and Karianto Leman. 2014. "Learning Deep Features for Multiple Object Tracking by using a Multi-task Learning Strategy." IEEE International Conference on Image Processing (ICIP), Paris, France, 838–842. IEEE.
Wang, Peng, Xiaohui Shen, Zhe Lin, Scott Cohen, Brian Price, and Alan L. Yuille. 2015. "Towards Unified Depth and Semantic Prediction from a Single Image." Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, Boston, MA, 2800–2809.
Wang, Qiang, Zhu Teng, Junliang Xing, Jin Gao, Weiming Hu, and Stephen Maybank. 2018. "Learning Attentions: Residual Attentional Siamese Network for High Performance Online Visual Tracking." Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, 4854–4863.
Wang, Xiang, Kai Wang, and Shiguo Lian. 2019. "Deep Consistent Illumination in Augmented Reality." IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct), Beijing, China, 189–194. IEEE.
Wang, Bing, Li Wang, Bing Shuai, Zhen Zuo, Ting Liu, Kap Luk Chan, and Gang Wang. 2016. "Joint Learning of Convolutional Neural Networks and Temporally Constrained Metrics for Tracklet Association." Proceedings of the IEEE conference on Computer Vision and Pattern Recognition Workshops, Las Vegas, NV, 1–8.
Wang, Peng, Ruigang Yang, Binbin Cao, Wei Xu, and Yuanqing Lin. 2018. "DeLS-3D: Deep Localization and Segmentation

with a 3D Semantic Map." Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, 5860–5869.

Wang, X., A. W. W. Yew, S. K. Ong, and A. Y. C. Nee. 2020. "Enhancing Smart Shop Floor Management with Ubiquitous Augmented Reality." *International Journal of Production Research* 58 (8): 2352–2367.

Wang, Qiang, Li Zhang, Luca Bertinetto, Weiming Hu, and Philip H. S. Torr. 2019. "Fast Online Object Tracking and Segmentation: A Unifying Approach." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, 1328–1338.

Wei, Side, Gang Ren, and Eamonn O'Neill. 2014. "Haptic and Audio Displays for Augmented Reality Tourism Applications." IEEE Haptics Symposium (HAPTICS), Houston, TX, 485–488. IEEE.

Wen, Longyin, Dawei Du, Zhaowei Cai, Zhen Lei, Ming-Ching Chang, Honggang Qi, Jongwoo Lim, Ming-Hsuan Yang, and Siwei Lyu. 2015. "UA-DETRAC: A New Benchmark and Protocol for Multi-object Detection and Tracking." Preprint arXiv:1511.04136.

Weng, Juyang, Thomas S. Huang, and Narendra Ahuja. 1989. "Motion and Structure From Two Perspective Views: Algorithms, Error Analysis, and Error Estimation." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 11 (5): 451–476.

Werrlich, Stefan, Kai Nitsche, and Gunther Notni. 2017. "Demand Analysis for an Augmented Reality based Assembly Training." Proceedings of the 10th international conference on PErvasive Technologies Related to Assistive Environments, Island of Rhodes, Greece, 416–422.

Wikitude GmbH. 2018. "Wikitude Augmented Reality: The World's Leading Cross-Platform AR SDK." Accessed 9 May 2020. https://www.wikitude.com/.

Wu, Hsin-Kai, Silvia Wen-Yu Lee, Hsin-Yi Chang, and Jyh-Chong Liang. 2013. "Current Status, Opportunities and Challenges of Augmented Reality in Education." *Computers & Education* 62: 41–49.

Wu, Yi, Jongwoo Lim, and Ming-Hsuan Yang. 2013. "Online Object Tracking: A Benchmark." IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Portland, OR.

Wu, Jian, Liwei Ma, and Xiaolin Hu. 2017. "Delving Deeper into Convolutional Neural Networks for Camera Relocalization." IEEE International Conference on Robotics and Automation (ICRA), Mariana Bay Sands, Singapore, 5644–5651. IEEE.

Wu, Xingyuan, Yonggang Qi, Jun Liu, and Jie Yang. 2018. "SketchsegNet: A RNN Model for Labeling Sketch Strokes." IEEE 28th international workshop on Machine Learning for Signal Processing (MLSP), Aalborg, Denmark, 1–6. IEEE.

Wu, Xiongwei, Doyen Sahoo, and Steven C. H. Hoi. 2020. "Recent Advances in Deep Learning for Object Detection." *Neurocomputing* 396: 39–64.

Wu, Jiajun, Yifan Wang, Tianfan Xue, Xingyuan Sun, Bill Freeman, and Josh Tenenbaum. 2017. "MarrNet: 3D Shape Reconstruction via 2.5 D Sketches." Advances in Neural Information Processing Systems, Long Beach, CA, 540–550.

Wu, Yu, Linchao Zhu, Lu Jiang, and Yi Yang. 2018. "Decoupled Novel Object Captioner." Proceedings of the 26th ACM international conference on Multimedia, Seoul, Korea, 1029–1037.

Wuest, Harald, Florent Vial, and Didier Strieker. 2005. "Adaptive Line Tracking with Multiple Hypotheses for Augmented Reality." 4th IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR'05), Vienna, Austria, 62–69. IEEE.

Wuest, Thorsten, Daniel Weimer, Christopher Irgens, and Klaus-Dieter Thoben. 2016. "Machine Learning in Manufacturing: Advantages, Challenges, and Applications." *Production & Manufacturing Research*4 (1): 23–45.

Xiao, Jianxiong, Krista A. Ehinger, Aude Oliva, and Antonio Torralba. 2012. "Recognizing Scene Viewpoint using Panoramic Place Representation." IEEE conference on Computer Vision and Pattern Recognition, Providence, RI, 2695–2702. IEEE.

Xie, Chun, Hidehiko Shishido, Yoshinari Kameda, Kenji Suzuki, and Itaru Kitahara. 2018. "A Calibration Method for Large-scale Projection based Floor Display System." IEEE conference on Virtual Reality and 3D User Interfaces (VR), Reutlingen, Germany, 725–726. IEEE.

Xu, Di, Zhen Li, and Yanning Zhang. 2020. "Real-time Illumination Estimation for Mixed Reality on Mobile Devices." IEEE conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW), Atlanta, GA, 703–704. IEEE.

Xu, Linli, James Neufeld, Bryce Larson, and Dale Schuurmans. 2005. "Maximum Margin Clustering." Advances in Neural Information Processing Systems, Vancouver, BC, Canada, 1537–1544.

Yacob, Filmon, Daniel Semere, and Erik Nordgren. 2019. "Anomaly Detection in Skin Model Shapes Using Machine Learning Classifiers." *The International Journal of Advanced Manufacturing Technology*105 (9): 3677–3689.

Yamamoto, Tomonori, Niki Abolhassani, Sung Jung, Allison M. Okamura, and Timothy N. Judkins. 2012. "Augmented Reality and Haptic Interfaces for Robot-assisted Surgery." *The International Journal of Medical Robotics and Computer Assisted Surgery* 8 (1): 45–56.

Yang, Tianyu, and Antoni B. Chan. 2018. "Learning Dynamic Memory Networks for Object Tracking." Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 152–167.

Yang, Ming-Der, Chih-Fan Chao, Kai-Siang Huang, Liang-You Lu, and Yi-Ping Chen. 2013. "Image-based 3D Scene Reconstruction and Exploration in Augmented Reality." *Automation in Construction* 33: 48–60.

Yang, Ruo-Yu, and Rahul Rai. 2019. "Machine Auscultation: Enabling Machine Diagnostics Using Convolutional Neural Networks and Large-scale Machine Audio Data." *Advances in Manufacturing* 7 (2): 174–187.

Yang, Ruoyu, Shubhendu Kumar Singh, Mostafa Tavakkoli, Nikta Amiri, Yongchao Yang, M. Amin Karami, and Rahul Rai. 2020. "CNN-LSTM Deep Learning Architecture for Computer Vision-based Modal Frequency Detection." *Mechanical Systems and Signal Processing* 144: 106885.

Yao, Ting, Yingwei Pan, Yehao Li, and Tao Mei. 2017. "Incorporating Copying Mechanism in Image Captioning for Learning Novel Objects." Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, Honolulu, HI, 6580–6588.

Yao, Xifan, Jiajun Zhou, Jiangming Zhang, and Claudio R. Boër. 2017. "From Intelligent Manufacturing to Smart Manufacturing for Industry 4.0 Driven by Next Generation Artificial

Intelligence and Further On." 5th International conference on Enterprise Systems (ES), Beijing, China, 311–318. IEEE.

Ye, Menglong, Edward Johns, Ankur Handa, Lin Zhang, Philip Pratt, and Guang-Zhong Yang. 2017. "Self-supervised Siamese Learning on Stereo Image Pairs for Depth Estimation in Robotic Surgery." Preprint arXiv:1705.08260.

Yi, Renjiao, Chenyang Zhu, Ping Tan, and Stephen Lin. 2018. "Faces as Lighting Probes via Unsupervised Deep Highlight Extraction." Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 317–333.

Yin, Zhichao, and Jianping Shi. 2018. "GeoNet: Unsupervised Learning of Dense Depth, Optical Flow and Camera Pose." Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, 1983–1992.

Yoo, Donggeun, Sunggyun Park, Joon-Young Lee, Anthony S. Paek, and In So Kweon. 2015. "AttentionNet: Aggregating Weak Directions for Accurate Object Detection." Proceedings of the IEEE international conference on Computer Vision, Santiago, Chile, 2659–2667.

Yu, Fisher, and Vladlen Koltun. 2015. "Multi-scale Context Aggregation by Dilated Convolutions." Preprint arXiv:1511.07122.

Yuan, M. L., Soh-Khim Ong, and Andrew Y. C. Nee. 2006. "Registration Using Natural Features for Augmented Reality Systems." *IEEE Transactions on Visualization and Computer Graphics* 12 (4): 569–580.

Zabih, Ramin, and John Woodfill. 1994. "Non-parametric Local Transforms for Computing Visual Correspondence." European conference on Computer Vision, Berlin, Germany, 151–158. Springer.

Zamir, Amir Roshan, Afshin Dehghan, and Mubarak Shah. 2012. "GMCP-tracker: Global Multi-object Tracking using Generalized Minimum Clique Graphs." European conference on Computer Vision, Firenze, Italy, 343–356. Springer.

Zhang, Mi, Xiangyun Hu, Jian Yao, Like Zhao, and Jiancheng Li, Jianya Gong. 2019. "Line-Based Geometric Consensus Rectification and Calibration From Single Distorted Manhattan Image." *IEEE Access* 7: 156400–156412.

Zhang, Binbin, Prakhar Jaiswal, Rahul Rai, Paul Guerrier, and George Baggs. 2019. "Convolutional Neural Network-based Inspection of Metal Additive Manufacturing Parts." *Rapid Prototyping Journal*25: 530–540.

Zhang, Yunbo, and Tsz-Ho Kwok. 2018. "Design and Interaction Interface Using Augmented Reality for Smart Manufacturing." *Procedia Manufacturing* 26: 1278–1286.

Zhang, Li, Yuan Li, and Ramakant Nevatia. 2008. "Global Data Association for Multi-object Tracking using Network Flows." IEEE conference on Computer Vision and Pattern Recognition, Anchorage, AK, 1–8. IEEE.

Zhang, Xianyu, Xinguo Ming, Zhiwen Liu, Dao Yin, Zhihua Chen, and Yuan Chang. 2019. "A Reference Framework and Overall Planning of Industrial Artificial Intelligence (I-AI) for New Application Scenarios." *The International Journal of Advanced Manufacturing Technology* 101 (9–12): 2367–2389.

Zhang, Zhipeng, and Houwen Peng. 2019. "Deeper and Wider Siamese Networks for Real-time Visual Tracking." Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, Long Beach, CA, 4591–4600.

Zhang, Aijia, Yan Zhao, and Shigang Wang. 2019. "Illumination Estimation for Augmented Reality Based on a Global Illumination Model." *Multimedia Tools and Applications* 78 (23): 33487–33503.

Zhang, Xiangyu, Xinyu Zhou, Mengxiao Lin, and Jian Sun. 2018. "ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices." Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, 6848–6856.

Zhao, Shuai, and Quan Qian. 2017. "Ontology Based Heterogeneous Materials Database Integration and Semantic Query." *AIP Advances* 7 (10): 105325.

Zhao, Hengshuang, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, and Jiaya Jia. 2017. "Pyramid Scene Parsing Network." Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, Honolulu, HI, 2881–2890.

Zheng, Feng. 2015. "Spatio-temporal Registration in Augmented Reality." PhD diss., Department of Computer Science, University of North Carolina at Chapel Hill. doi:10.17615/pw64-4p19.

Zheng, Xinru, Xiaotian Qiao, Ying Cao, and Rynson W. H. Lau. 2019. "Content-aware Generative Modeling of Graphic Design Layouts." *ACM Transactions on Graphics (TOG)* 38 (4): 1–15.

Zheng, Feng, Dieter Schmalstieg, and Greg Welch. 2014. "Pixel-wise Closed-loop Registration in Video-based Augmented Reality." IEEE International Symposium on Mixed and Augmented Reality (ISMAR), Munich, Germany, 135–143. IEEE.

Zheng, Feng, Ryan Schubert, and Greg Weich. 2013. "A General Approach for Closed-loop Registration in AR." IEEE Virtual Reality (VR), Lake Buena Vista, FL, 47–50. IEEE.

Zhou, Huiyu, Yuan Yuan, and Chunmei Shi. 2009. "Object Tracking Using SIFT Features and Mean Shift." *Computer Vision and Image Understanding* 113 (3): 345–352.

Zhou, Guanghui, Chao Zhang, Zhi Li, Kai Ding, and Chuang Wang. 2020. "Knowledge-driven Digital Twin Manufacturing Cell Towards Intelligent Manufacturing." *International Journal of Production Research* 58 (4): 1034–1051.

Zhu, Jiahong, Sohkhim K. Ong, and Andrew Y. C. Nee. 2013. "An Authorable Context-aware Augmented Reality System to Assist the Maintenance Technicians." *The International Journal of Advanced Manufacturing Technology* 66 (9–12): 1699–1714.

Zhu, Jiejie, and Zhigeng Pan. 2008. "Occlusion Registration in Video-based Augmented Reality." Proceedings of The 7th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and Its Applications in Industry, Singapore, 1–6.

Zhu, Jiejie, Zhigeng Pan, Chao Sun, and Wenzhi Chen. 2010. "Handling Occlusions in Video-based Augmented Reality Using Depth Information." *Computer Animation and Virtual Worlds* 21 (5): 509–521.

Zhu, Zheng, Qiang Wang, Bo Li, Wei Wu, Junjie Yan, and Weiming Hu. 2018. "Distractor-aware Siamese Networks for Visual Object Tracking." Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 101–117.

Zhuang, Bingbing, Quoc-Huy Tran, Pan Ji, Gim Hee Lee, Loong Fah Cheong, and Manmohan Chandraker. 2019. "Degeneracy in Self-Calibration Revisited and a Deep Learning Solution for Uncalibrated SLAM." Preprint arXiv:1907.13185.

Zollmann, Stefanie, and Gerhard Reitmayr. 2012. "Dense Depth Maps from Sparse Models and Image Coherence for Augmented Reality." Proceedings of the 18th ACM

Symposium on Virtual Reality Software and Technology, Toronto, Ontario, Canada, 53–60.
Zou, Yuliang, Zelun Luo, and Jia-Bin Huang. 2018. "DF-Net: Unsupervised Joint Learning of Depth and Flow using Cross-task Consistency." Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 36–53.
Zubizarreta, Jon, Iker Aguinaga, and Aiert Amundarain. 2019. "A Framework for Augmented Reality Guidance in Industry." *The International Journal of Advanced Manufacturing Technology* 102 (9–12): 4095–4108.
Zwald, Laurent, and Sophie Lambert-Lacroix. 2012. "The BerHu Penalty and the Grouped Effect." Preprint arXiv: 1207.6868.