Association for Information Systems
AIS Electronic Library (AISeL)

MCIS 2018 Proceedings

Mediterranean Conference on Information Systems (MCIS)

2018

A short survey on modern virtual environments that utilize AI and synthetic data

Michalis Korakakis
Ionian University Corfu, michalis.korakakis@gmail.com

Phivos Mylonas
Ionian University Corfu, fmylonas@ionio.gr

Evaggelos Spyrou
NCSR - "Demokritos", espyrou@iit.demokritos.gr

A short survey on modern virtual environments that utilize AI and synthetic data

Research full-length paper

General Track

Korakakis, Michalis, Ionian University, Corfu, Greece, michalis.korakakis@gmail.com

Mylonas, Phivos, Ionian University, Corfu, Greece, fmylonas@ionio.gr

Spyrou, Evaggelos, NCSR – "Demokritos", Athens, Greece, espyrou@iit.demokritos.gr

Abstract

Within a rather abstract computational framework Artificial Intelligence (AI) may be defined as intelligence exhibited by machines. In computer science, though, the field of AI research defines itself as the study of "intelligent agents." In this context, interaction with popular virtual environments, as for instance in virtual game playing, has gained a lot of focus recently in the sense that it provides innovative aspects of AI perception that did not occur to researchers until now. Such aspects are typically formed by the computational intelligent behavior captured through interaction with the virtual environment, as well as the study of graphic models and biologically inspired learning techniques, like, for instance, evolutionary computation, neural networks, and reinforcement learning. In this short survey paper, we attempt to provide an overview of the most recent research works on such novel, yet quite interesting, research domains. We feel that this topic forms an attractive candidate for fellow researchers that came into sight over the last years. Thus, we initiate our study by presenting a brief overview of our motivation and continue with some basic information on recent virtual graphic models utilization and the state-of-the-art on virtual environments, which constitutes two clearly identifiable components of the herein attempted summarization. We then continue, by briefly reviewing the interesting video games territory, and by discerning and discriminating its useful types, thus envisioning possible further utilization scenarios for the collected information. A short discussion on the identified trends and a couple of future research directions conclude the paper.

Keywords

Keywords: computational intelligence, virtual environment, graphic models, video games, survey

I. Motivation & Introduction

During the last few years, recent advances in the fields of machine learning and AI have led to the development of deep learning algorithms. Such algorithms typically require large datasets which may be difficult and tedious both to collect and to annotate. Moreover, there are certain limitations that may arise when using such a dataset. First, manual labeling may be ambiguous. When many users are involved, they may have different backgrounds, limited knowledge on the specific goals of the annotation (i.e., on *how* will the dataset be used) and even limited understanding of the concepts they aim to annotate. These may lead to poor annotations, prone to errors. Second, in real-life datasets occlusion of objects is common and limited viewpoints are included. Thus, it is difficult or even impossible to create high-quality datasets without any such flaws. A notable solution when dealing with the aforementioned limitations is to develop graphic and real-like models of real-life objects. Boosted by the increasing computational processing power of late GPUs, many approaches have been lately presented that aim to create realistic models for typical objects and many synthetic datasets have emerged. It is now feasible to have access to large, automatically generated and annotated datasets of 2D and/or 3D objects from any viewpoint, e.g., by exploiting CAD models. Of course, ambiguity is eliminated and semantic descriptions are accurate. This approach has been used for tasks such as Simultaneous Localization and Mapping (SLAM), tracking, gesture recognition, etc.

In other tasks, such as surveillance or autonomous driving, it is necessary to create more complex virtual environments. It is also time-consuming to collect annotated data, whilst annotation errors should be avoided. Two of the areas that have attracted the interest of the related research community are the *virtual worlds* and *video gaming*. The former mainly deal with simulated environments that are accommodated within networked computers, allowing users to participate via personalized avatars. In many cases, such virtual worlds are real-like, i.e., they mimic real-life. Users are able to navigate within them and interact with other users. This way, one is able to capture scenes from such worlds (e.g., pedestrians crossing a street) and use them to train real-life applications (e.g., collision warning with pedestrian detection systems in cars). As for video games, it is easier, e.g., to segment and annotate synthetic data sets for pattern recognition, or test the performance of agents.

In this position paper we attempt to provide a short survey on the aforementioned research areas. Motivated by the growth of synthetic datasets and the exploitation of data generated within virtual environments & video games, we aim to identify the main trends and provide a tabular summarization of relevant research works in order to facilitate a better understanding of the

Work	Task(s)	Method(s)	Pros	Cons	Dataset(s)
[1]	flight controller	deep reinforcement learning	collision avoidance	manual training	synthetic
[2]	benchmark & model	dataset generation	robust ground truth, evaluation	simulation parameters	synthetic
[3]	hand gestures analysis	dataset generation	comparative evaluation	narrow focus	natural & synthetic
[4]	semantic scene understanding	dataset generation	contextual information	limited dataset	1002 image classes
[5]	multi-view object class detector	3D objects recognition	performance	limited novelty small dataset	custom (car) dataset
[6]	fine-pose estimation	CAD models exploitation	performance	limited dataset	50 real images
[7]	fine-pose estimation	exact 3D models exploitation	performance	limited dataset	800 images, 225 models
[8]	2D object detection	3D model rendering	small-scale dataset required	limited object types	virtual data
[9]	viewpoints synthesis and re-synthesis	3D model utilization	comparative evaluation	limited dataset	synthetic
[10]	commonsense assertion classification	visual abstraction	novel approach	subjective evaluation	synthetic (AMT)
[11]	deep CNNs object detection	crowd-sourced 3D CAD model utilization	novel approach	preliminary results	synthetic
[12]	multi-view object class detection	3D geometric model utilization	novel approach	weak evaluation	synthetic
[18]	multi-view object class detection	3D representations of object classes	novel approach	limited object categories	PASCAL 2006 car dataset 58 synthetic models
[13]	viewpoint estimation	3D model rendering	convolutional neural networks training	complex approach	5 datasets
[14]	object detection	3D CAD models utilization	visual elements matching	false positives	Google 3D Warehouse
[16]	CNNs features analysis	3D CAD models utilization	quantitative & qualitative approach	complexity	3 datasets
[15]	vision algorithm training	on-the-fly rendering	CNNs utilization	limited testing	NYU dataset
[17]	indoor scene understanding	training data generation	noise incorporation	limited evaluation	synthetic, NYUv2 dataset

Table 1. Approaches utilizing graphic models.

field for future research studies and applications. In this manner, all tables to follow include six columns: The first contains each works bibliographic reference number in order of appearance; the second describes the main task the particular work attempts to tackle; the third focuses on the depicted methodology the authors propose or utilize in order to solve the particular research task at hand; the fourth column presents a representative set of positive characteristics (if applicable); the fifth column presents a representative set of negative characteristics (if applicable); the sixth column provides information on the utilized dataset/game (if any).

The structure of the rest of this paper is as follows: in Section II we present research works exploiting graphic models within computational efficient tasks. Section III deals with recent research efforts aiming at virtual worlds utilization, whereas Section IV details the application of information derived from video games for the purpose of resolving related research problems in an innovative manner. Finally, in Section V we conclude the paper, while also mentioning our plans for future work.

II. Graphic Models

The rather unexpected recent advance on graphic models utilization in a variety of computational tasks boosted related research and constituted the latter as the main building block, on top of which several, quite distinct approaches, originating from different computer science domains, were based. Towards this end, a number of studies have tried to take advantage of graphic modeling characteristics in order to utilize them as a means to provide unbiased and accurate solutions to traditional computational problems. Among the pioneers in this field, Sadeghi [1] proposed a flight controller scheme trained entirely in a 3D CAD model simulator, based on collision avoidance via a deep reinforcement learning algorithm. Handa et al. [2] presented a benchmark and a model primarily aimed at RGB-D visual odometry and 3D reconstruction. Their work provided a means of quantitatively evaluating the quality of the final map or surface reconstruction produced in the process. In [3] authors introduced a corpus for benchmarking hand gesture recognition systems. In another interesting approach that attempts to bridge two rather distinct research fields, Zitnick et al. [4] tackled the semantic scene understanding problem via the utilization of abstract images created from collections of clip art. More specifically, their main research question was whether photorealistic images are necessary for scene understanding. They demonstrated the potential of using abstract images to study high-level semantic understanding, since such images may allow for the easy creation of huge noiseless data sets.

In other early works, researchers revisited the idea of utilizing learning shape models for object class recognition purely from 3D data [5]; the main novelty lies on the fact that authors did not use any natural training images of the object class of interest in the process. In the same sense, Lim et al. introduced a novel approach to the problem of localizing objects in an image and estimating their fine-pose ([6], [7]). In an alternative approach, Sun et al. [8] attempted 2D object detection by training on virtual data rendered from 3D models, avoiding the need for manual labeling. Rematas et al. [9] utilized 3D models towards an image-based synthesis and re-synthesis of viewpoints by successfully addressing the challenging problem of filling in disocclusion areas.

Vedantam et al. [10] took the aforementioned interpretations a step further and investigated whether vision provides a complementary source of commonsense knowledge to text. Hence, in addition to reasoning about the similarity between tuples based on text, they proposed to ground commonsense assertions in the visual world and evaluate similarity between assertions using visual features. Crowd-sourced 3D CAD models are utilized in [11] in order to augment the training data of contemporary

Work	Task(s)	Method(s)	Pros	Cons	Dataset(s)
[20]	synthetic data framework	driving simulator	easily adoptable	lack on comparable evaluation	synthetic
[21]	pedestrian detection	SVM-based approach	virtual & real world comparison	false positives	synthetic
[22]	pedestrian recognition	classifier	performance	no high-level semantic interpretation	Towncenter, PETS 2006, CMUSRD
[29]	pedestrian detection	aspect clustering & part alignment	detailed approach	occluded pedestrians detection accuracy	4 datasets
[23]	video surveillance evaluation	visual surveillance simulation test-bed	flexibility	weak evaluation	synthetic
[24]	autonomous driving	vision-based semantic segmentation	real-world urban images	manual annotations	synthetic
[25]	agent training test-bed	virtual world racing simulation	modular & portable	stability	synthetic
[26]	data-driven robotic simulator	real-world simulator	cross-platform	no evaluation	synthetic
[27]	low-level features evaluation	photo-realistic synthetic world	innovative approach	performance	2 virtual datasets
[28]	virtual- to real-world domain adaptation	vehicle detection	novel approach comparative evaluation	narrow application domain	virtual world data
[30]	simulation platform	graphics exploitation to boost simulation	innovative approach	narrow application domain	synthetic

Table 2. Approaches utilizing graphic models.

deep convolutional neural network models (CNNs). In [12] authors used a part model which discriminatively learns the object appearance with spatial pyramids from a database of real images, and encodes the 3D geometry of the object class with a generative representation built from a database of synthetic models. In a similar approach [18], they presented a 3D approach to multi-view object class detection by building 3D representations of object classes which allow to handle viewpoint changes and intra-class variability.

In [13], authors tackled the object viewpoint estimation problem by utilizing 3D - instead of traditional 2D - models, i.e., they used images rendered from 3D models to train convolutional neural networks for viewpoint estimation on real-world images. Aubrey et al. [14] translated the problem of object category detection in images as a 2D-to-3D alignment problem, whereas in [16] they further analyzed CNN feature responses corresponding to different scene factors via utilizing a large database of suitable 3D CAD models. Papon et al. [15] addressed the problem of indoor scene understanding from RGB-D images through a multi-output convolutional neural network that utilizes an on-the-fly rendering pipeline generating realistic synthetic training datasets. Finally, in [17], authors tackled the indoor scene understanding problem by carefully synthesizing training data with appropriate noise models with promising results. In other words they used deep learning as their computational framework for semantic segmentation, a rather interesting and yet to be fully analyzed research approach. Table 1. provides a detailed overview of the discussed research efforts by categorizing them accordingly, illustrates their advantages and disadvantages and reasons on their suitability within the broader research field.

III. Virtual Environments

From the definition point of view, a virtual world forms “a computer-based simulated environment populated by users who may explore the virtual world, participate in its activities and communicate with other users” [19]. Such virtual environments can serve a variety of research and educational goals and may be useful for examining human and computational behaviour. Haltakov et al. [20] presented a virtual world, based on the open-source framework VDRIFT, towards the simplification of collecting images for driving-related applications. Furthermore, the authors highlighted the efficiency of their proposed framework through its evaluation upon a set of multi-class segmentation algorithms. Using a SVM-based approach Marin et al. [21] proposed a classifier able to discern from an image the presence of a pedestrian. To achieve this, the authors exploited data generated by a virtual world. Their findings indicated the suitability of synthetic data for the aforementioned task. Similarly, Hattori et al. [22] using data produced by a virtual world, trained a classifier for the task of recognizing pedestrians. The authors stated that their proposed approach was able to outperform conventional pedestrian detection systems. In addition, Xu et al. [29] discussed in detail how virtual world data may be used for learning pedestrian deformable part-based models, i.e., popular state-of-the-art pedestrian/object detectors. By utilizing a popular game engine Taylor et al. [23] presented ObjectVideo Virtual Video, an open-source framework for generating realistic video surveillance footage. Through the presentation of various applicable scenarios along with the implementation of a set of vision-based algorithms the authors highlighted the applicability of their surveillance test-bed. Ros et al. [24] created SYNTHIA, a synthetic annotated dataset concerning autonomous driving applications. Through the application of a convolutional neural network upon a training corpus comprised of synthetic data for the task of semantic segmentation, the authors concluded that the model was able to operate efficiently on real world data.

Wymann et al. [25] developed TORCS, a test-bed for training agents in the context of a racing simulation. The authors stressed that the constructed platform is ideal for a multitude of both high- and low-level tasks concerning autonomous driving. In [26], authors present an easy-to-use simulator aiming to enable seamless generation of training data. It allows for enabling rapid training and development of data-driven robotic systems and enables high fidelity simulation which in turn can be used to collect training data for building machine learning models. Kaneva et al. in [27] investigated the utilization of a synthetic photo-realistic virtual world in order to gain complete and repeatable control of the environment so as to evaluate low-level computer vision image features. Furthermore, Lopez et al. in [28] introduce the notion of a deformable part-based model as an exemplifying case of virtual- to real-world domain adaptation. As a use case, they addressed the challenge of vehicle detection

for driver assistance, using different publicly available virtual world data. Finally, Veeravasarapu et al. [30] described a simulation platform that incorporates latest graphics advances and use it for systematic performance characterization and trade-off analysis for vision system design. Their approach establishes the link between alternative viewpoints, involving models with physics based semantics and signal and perturbation semantics and confirms insights in literature on robust change detection. They verified the utility of their simulation platform in several case studies dealing with various situations such as illumination changes, noise and the effect of weather. Table 2. presents the herein discussed approaches according to their type and illustrates each one's main features and characteristics.

IV. Video Games

In the herein discussed framework, video games are nowadays gaining a lot of attention towards one of the long-standing challenges of AI, namely the task of successfully learning to control agents directly from typical computational inputs, provided by computer vision. Most successful reinforcement learning applications that operate on such a domain have relied on manual features and representations. Clearly, the performance of such systems heavily relies on the quality of features representation, where video games may indeed come to the rescue. Recent advances in deep learning have made it possible to extract high-level features from raw sensory data, leading to important breakthroughs and novel approaches. Among the ones worth mentioning is the one by

Work	Task(s)	Method(s)	Pros	Cons	Game(s)
[31]	machine learning algorithms training	photo-realistic simulation images	unsupervised method	dataset bias	GTA
[32]	machine learning algorithms efficiency	neural network model	technique validation	narrow application domain	N/A
[33]	ground truth creation	pixel-accurate semantic label maps	adaptability of approach	small dataset (25K frames)	GTA V
[34]	visual-based reinforcement learning	convolutional deep neural networks	3D, semi-realistic	narrow application domain	Doom
[35]	artificial intelligence agents benchmarking	reinforcement learning and planning	extensive evaluation	platform dependent	55 games
[36]	recursively decomposable factorizations	QTF model	good evaluation	platform dependent	Atari 2600 games
[37]	reinforcement learning toolkit	web platform	collaborative effort	preliminary implementation	POMDPs collection
[38]	autonomous artificial agents learning	facilitates creative task development	flexible API	lack on evaluation	Quake III Arena
[39]	deep learning algorithms training	client-server architecture	flexible API	lack on evaluation	StarCraft: Brood War
[41]	autonomous artificial agents learning	utilization of 3D world	plethora of use-cases	lack on evaluation	Minecraft
[42]	autonomous artificial agents learning	reinforcement learning environment	detailed approach	platform dependent	StarCraft II

Table 3. Video games exploitation works.

Johnson-Roberson et al. [31], who presented an unsupervised approach for the annotation of objects found in synthetic images generated by the popular Grand Theft Auto V open-world video game. By training a deep learning model using the aforementioned data, the authors concluded that the network was able to achieve better results than those obtained with a real-world dataset. Similarly, by providing similar environment conditions to the ones found in real-world datasets, Shafaei et al. [32] investigated whether synthetically generated RGB images from a video game can augment the efficiency of machine learning algorithms responsible for performing image segmentation and depth perception. By training a neural network model on synthetic and real-world data they were able to achieve comparable results and moreover they stated that the synthetically generated RGB images may even provide better results compared to the real-world datasets, if a simple domain adaptation technique is applied. Because the creation of an adequate annotated dataset is a rather resource-intensive and complex process, Richter et al. [33] aimed to examine the exploitation of video games for the purpose of training a semantic segmentation system. The authors stressed that models trained on both types of data are able to achieve better results than those relying solely on real-world images, noting, in addition, that the object variety found in the synthetically generated images is rather satisfactory.

Based on the need for the creation of a research-oriented reinforcement learning platform, Kempka et al. [34] developed ViZDoom, a visual learning platform that is based on Doom, a first-person shooter (FPS) video game and may be highly customizable via user scenarios. Moreover, through the realization of two experiments concerning deep reinforcement learning, the authors highlighted the effectiveness of their proposed platform into training competent bots exhibiting human-like behaviors. Bellemare et al. ([35], [36]) presented the Arcade Learning Environment (ALE), a multipurpose video game platform for benchmarking AI agents using Atari 2600 video games. More specifically, upon an experimental evaluation of the presented platform concerning reinforcement learning and planning algorithms, the authors highlighted the suitability of their proposed platform for assessing agents. Having in mind the significant impact of recent advances in reinforcement learning, Brockman et al. [37] developed OpenAI Gym, a platform for reinforcement learning research. More specifically, the aforementioned toolkit allows the development and assessment of agents in a plethora of supported video games.Moreover, the OpenAI team presented Universe,[1][2] a tool built on top of OpenAI Gym that offers the ability to create and test reinforcement learning algorithms in a multitude of both video games (e.g., Flash and Atari video games), as well as real-world tasks such as web browsing.

Synnaeve et al. [39] developed TorchCraft, a library that facilitates the application of deep learning algorithms in video games within the Real-Time Strategy (RTS) domain (e.g., StarCraft), through the utilization of Torch [40], a well-known machine-learning library. Similarly, Vinyals et al. [42] created a machine learning API that constitutes possible to utilize information extracted from StarCraft in real-time so as to train and test both deep learning and reinforcement learning algorithms.Johnson et al. [41]

[1] https://blog.openai.com/universe/
[2] https://github.com/openai/universe

presented Malmo, a platform that utilizes Minecraft, a sandbox video game, in order to provide a test-bed for the creation of agents aimed for fundamental research in artificial general intelligence (AGI). In particular, Malmo supports the design of agents that are able to perform a multitude of complex tasks, as offered in Minecraft's detailed 3D world. By using as a basis the Quake III Arena, a first-person shooter video game, Beattie et al. [38] presented DeepMind Lab, a 3D environment designed for research and development of general AI and machine learning systems. A summary of the aforementioned studies is provided in Table 3., which classifies them accordingly. Finally, we should herein mention several AI competitions that use popular video games. Typically, the goal is to develop some kind of AI software agent, able to "play" non-deterministic games. Popular examples of such competitions include Ms Pac-Man Competition[3], the StarCraft AI Competition [4], the Simulated Car Racing Championship [5] and the Mario AI championship[6]. Such competitions take place within popular international AI conferences and attract the interest of the research community.

V. Conclusions & Future Work

In this work, we presented several studies focusing on utilizing new, evolving gaming and computational intelligence techniques for improving and/or gaining insight into a multitude of different tasks within the so-called "virtual world" domain. We organized the aforementioned studies into three major categories, namely providing information on recent works on graphic modeling, using virtual environments as the means of implementing numerous computational tasks, and finally, video gaming with a special focus on reinforcement learning approaches. Having in mind the impact of virtual gaming in both everyday life of youngsters and research advances of the respective research community, our motivation was to identify the main trends within this particular application domain in order to facilitate a better understanding of the emerged field for future studies. We believe and hope that based on the tabular organization and interpretation of each category we provided, future useful research directions may be identified by interested fellow researchers and that they may be able to use this survey work as a future point of reference.

According to the findings of this survey, one may identify a clear trend: *current related research is dominated by an active, ongoing and steadily increasing utilization of information and synthetic data derived from virtual environments towards the simplification of computational tasks*, that were never before treated nor interpreted in this manner by researchers. The most aided domain at the moment seems to be (modern) computer vision approaches that exploit the aforementioned techniques to depend less on expensive data acquisition and accurate manual labeling. Still, we think this is an evolving and currently shaping research territory with many more application domains to be identified in the process. So, among our future work lies the task of further monitoring this evolving research community so as to broaden this study into other identifiable sub-domains, while in addition exploring traditional relationships between virtual world production and consumption.

References

[1] Sadeghi, F., Levine, S. (2016). *(CAD)2RL: Real Single-Image Flight without a Single Real Image*. arXiv:1611.04201.

[2] Handa, A., Whelan, T., McDonald, J., Davison, A. J. (2014). *A Benchmark for RGB-D Visual Odometry, 3D Reconstruction and SLAM*. 2014 IEEE International Conference on Robotics and Automation (ICRA). Hong Kong, China.

[3] Molina, J., Pajuelo, J. A., Escudero-Violo, M., Bescs, J., Martnez, J. M. (2014). *A natural and synthetic corpus for benchmarking of hand gesture recognition systems*. Machine Vision and Applications. vol. 25, no. 4. pp. 943–954.

[4] Zitnick, C. L., Vedantam, R., Parikh, D. (2016). *Adopting Abstract Images for Semantic Scene Understanding*. IEEE Transactions on Pattern Analysis and Machine Intelligence. vol. 38, no. 4. pp. 627–638.

[5] Stark, M., Goesele, M., Schiele, B. (2010). *Back to the Future: Learning Shape Models from 3D CAD Data*. British Machine Vision Conference (BMVC). Aberystwyth, Wales, UK.

[6] Lim, J. J., Khosla, ., Torralba, . (2014). *FPM: Fine pose Parts-based Model with 3D CAD models*. European Conference on Computer Vision (ECCV). Zurich, Switzerland. September 6-12.

[7] Lim, J. J., Pirsiavash, ., Torralba, . (2013). *Parsing IKEA Objects: Fine Pose Estimation*. IEEE International Conference on Computer Vision (ICCV). Sydney, Australia. April.

[8] Sun, B., Saenko, K. (2014). *From Virtual to Reality: Fast Adaptation of Virtual Object Detectors to Real Domains*. British Machine Vision Conference (BMVC). Nottingham, UK. September 1-5.

[9] Rematas, K., Ritschel, T., Fritz, M., Tuytelaars, T. (2014). *Image-based Synthesis and Re-Synthesis of Viewpoints Guided by 3D Models*. Computer Vision and Pattern Recognition. Columbus, Ohio, USA.

[10] Vedantam, R., Lin, X., Batra, T., Zitnick, C.L., Parikh, D. (2015). *Learning Common Sense Through Visual Abstraction*. IEEE International Conference on Computer Vision (ICCV). Santiago, Chile.

[11] Peng, X., Sun, B., Ali, K., Saenko, K. (2015). *Learning Deep Object Detectors from 3D Models*. IEEE International Conference on Computer Vision (ICCV). Santiago, Chile.

[12] Liebelt, J., Schmid, C. (2010). *Multi-View Object Class Detection with a 3D Geometric Model*. Computer Vision and Pattern Recognition (CVPR). San Francisco, CA, USA.

[13] Su, H., Qi, C. R., Li, Y., Guibas, L. J. (2015). *Render for CNN: Viewpoint Estimation in Images Using CNNs Trained with Rendered 3D Model Views*. IEEE International Conference on Computer Vision (ICCV). Santiago, Chile.

[3] http://cswww.essex.ac.uk/staff/sml/pacman/PacManContest.html
[4] https://cilab.sejong.ac.kr/sc_competition/
[5] http://cig.dei.polimi.it/
[6] http://www.marioai.org/

[14] Aubry, M., Maturana, D., Efros, A. A., Russell, B. C., Sivic, J. (2014). *Seeing 3D chairs: exemplar part-based 2D-3D alignment using a large dataset of CAD models*. Computer Vision and Pattern Recognition (CVPR). Columbus, Ohio, USA.

[15] Papon, J., Schoeler, M. (2015). *Semantic Pose using Deep Networks Trained on Synthetic RGB-D*. IEEE International Conference on Computer Vision (ICCV). Santiago, Chile.

[16] Aubry, M., Russell, B. C. (2015). *Understanding deep features with computer-generated imagery*. IEEE International Conference on Computer Vision (ICCV). Santiago, Chile.

[17] Handa, A., Patraucean, V., Badrinarayanan, V., Stent, S., Cipolla, R. (2015). *Understanding Real World Indoor Scenes With Synthetic Data*. Computer Vision and Pattern Recognition (CVPR). Boston, MA, USA.

[18] Liebelt, J., Schmid, C., Schertler, K. (2008). *Viewpoint-Independent Object Class Detection using 3D Feature Maps*. IEEE Computer Vision and Pattern Recognition (CVPR). Anchorage, Alaska, USA. 24-26 June.

[19] Aichner, T., Jacob, F. (2015). *Measuring the Degree of Corporate Social Media Use*. International Journal of Market Research. 57 (2). pp. 257-275. March.

[20] Haltakov, V., Unger, C., Ilic, S. (2013). *Framework for generation of synthetic ground truth data for driver assistance applications*. In German Conference on Pattern Recognition (GCPR). Saarbrcken, Germany.

[21] Marin, J., Vazquez, D., Geronimo, D., Lopez, A. M. (2010). *Learning Appearance in Virtual Scenarios for Pedestrian Detection*. Computer Vision and Pattern Recognition (CVPR). San Francisco, CA, USA.

[22] Hattori, H., Boddeti, V. N., Kitani, K., Kanade, T. (2015). *Learning Scene-Specific Pedestrian Detectors without Real Data*. Computer Vision and Pattern Recognition (CVPR). Boston, MA, USA. June.

[23] Taylor, G. R., Chosak, A. J., Brewer, P. C. (2007). *OVVV: Using Virtual Worlds to Design and Evaluate Surveillance Systems*. Computer Vision and Pattern Recognition (CVPR). Minneapolis, MN, USA. June.

[24] Ros, G., Sellart, L., Materzynska, J., Vazquez, D., Lopez, A. M. (2016). *The SYNTHIA Dataset: A Large Collection of Synthetic Images for Semantic Segmentation of Urban Scenes*. Computer Vision and Pattern Recognition (CVPR). Las Vegas, Nevada, USA. June.

[25] Wymann, B., Dimitrakakis, C., Sumner, A., Espie, E., Guionneau, C. (2013). *TORCS: The open racing car simulator*. http://torcs.sourceforge.net/.

[26] Shah, S., Dey, D., Lovett, C., Kapoor, A. (2017). *Aerial Informatics and Robotics Platform*. https://www.microsoft.com/en-us/research/project/aerial-informatics-robotics-platform/.

[27] Kaneva, B., Torralba, A., Freeman, W. T. (2011). *Evaluation of Image Features Using a Photorealistic Virtual World*. IEEE International Conference on Computer Vision (ICCV). Barcelona, Spain. November.

[28] Lopez, A. M., Xu, J., Gomez, J. L., Vazquez, D., Ros, G. (2016). *From Virtual to Real World Visual Perception using Domain Adaptation - The DPM as Example*. arXiv:1612.09134.

[29] Xu, J., Vzquez, D., Lpez, A. M., Marn, J., Ponsa, D. (2014). *Learning a Part-Based Pedestrian Detector in a Virtual World*. IEEE Transactions on Intelligent Transportation Systems. vol. 15, no. 5. pp. 2121–2131.

[30] Veeravasarapu, V. S. R., Hotay, R. N., Rothkopf, C., Visvanathan, R. (2015). *Simulations for Validation of Vision Systems*. arXiv:1512.01030.

[31] Johnson-Roberson, M., Barto, C., Mehta, R., Sridhar, S. N., Vasudevan, R. (2016). *Driving in the matrix: Can virtual worlds replace human-generated annotations for real world tasks?*. arXiv:1610.01983.

[32] Shafaei, A., Little, J., Schmidt, M. (2016). *Play and learn: Using video games to train computer vision models*. British Machine Vision Conference (BMVC). York, UK.

[33] Richter, S. R., Vineet, V., Roth, S., Koltun, V. (2016). *Playing for data: Ground truth from computer games*. European Conference on Computer Vision (ECCV). Amsterdam, The Netherlands. October.

[34] Kempka, M., Wydmuch, M., Runc, G., Toczek, J., Jaskowski, W. (2016). *Vizdoom: A doom-based ai research platform for visual reinforcement learning*. arXiv:1605.02097.

[35] Bellemare, M. G., Naddaf, Y., Veness, J., Bowling, M. (2012). *The arcade learning environment: An evaluation platform for general agents*. Journal of Artificial Intelligence Research. vol. 47, no. 1. pp. 253–279.

[36] Bellemare, M. G., Veness, J., Bowling, M. (2013). *Bayesian Learning of Recursively Factored Environments*. 30th International Conference on Machine Learning (ICML). Atlanta, USA. June 1621.

[37] Brockman, G., Cheung, V., Pettersson, L., Schneider, J., Schulman, J., Tang, J., Zaremba, W. (2016). *OpenAI Gym*. arXiv:1606.01540.

[38] Beattie, C., Leibo, J. Z., Teplyashin, D., Ward, T., Wainwright, M., Küttler, H., Lefrancq, A., Green, S., Valds, V., Sadik, A., Schrittwieser, J., Anderson, K., York, S., Cant, M., Cain, A., Bolton, A., Gaffney, S., King, H., Hassabis, D., Legg, S., Petersen, S. (2016). *DeepMind Lab*. arXiv:1612.03801.

[39] Synnaeve, G., Nardelli, N., Auvolat, A., Chintala, S., Lacroix, T., Lin, Z., Richoux, F., Usunier, N. (2016). *TorchCraft: a Library for Machine Learning Research on Real-Time Strategy Games*. arXiv:1611.00625.

[40] Collobert, R., Bengio, S., Marithoz, J. (2002). *Torch: A Modular Machine Learning Software Library*.

[41] Johnson, M., Hofmann, K., Hutton, T., Bignell, D. (2016). *The Malmo Platform for Artificial Intelligence Experimentation*. International Joint Conference on Artificial Intelligence (IJCAI). New York, USA.

[42] Vinyals, O., Ewalds, T., Bartunov, S., Georgiev, P., Vezhnevets, A. S., Yeo, M., Makhzani, A., Küttler, H., Agapiou, J., Schrittwieser, J., Quan, J., Gaffney, S., Petersen, S., Simonyan, K., Schaul, T., van Hasselt, H., Silver, D., Lillicrap, T., Calderone, K., Keet, P., Brunasso, A., Lawrence, D., Ekermo, A., Repp, J., Tsing, R. (2017). *StarCraft II: A New Challenge for Reinforcement Learning*. arXiv:1708.04782.