AFRL-RH-WP-TR-2018-0058

HUMAN ACTIVITY SYNTHETIC DATA GENERATION

Zhiqing Cheng
Timothy MtCastle
Todd Huster
Max Grattan

Infoscitex Corporation, a DCS Company

John Camp
Huaining Cheng
Monique Brisson

Human Signatures Branch

August 2018
Interim Report

AIR FORCE RESEARCH LABORATORY
711TH HUMAN PERFORMANCE WING,
AIRMAN SYSTEMS DIRECTORATE,
WRIGHT-PATTERSON AIR FORCE BASE, OH 45433
AIR FORCE MATERIEL COMMAND
UNITED STATES AIR FORCE

NOTICE AND SIGNATURE PAGE

Using Government drawings, specifications, or other data included in this document for any purpose other than Government procurement does not in any way obligate the U.S. Government. The fact that the Government formulated or supplied the drawings, specifications, or other data does not license the holder or any other person or corporation; or convey any rights or permission to manufacture, use, or sell any patented invention that may relate to them.

This report was cleared for public release by the 88th Air Base Wing Public Affairs Office and is available to the general public, including foreign nationals. Copies may be obtained from the Defense Technical Information Center (DTIC) (http://www.dtic.mil).

AFRL-RH-WP-TR-2018-0058 HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

CAMP.JOHN.L.1102680240 Digitally signed by CAMP.JOHN.L.1102680240 Date: 2019.04.17 12:04:24 -04'00'

JOHN CAMP, Ph.D., Work Unit Manager
Human Signatures Branch
Airman Systems Directorate
711th Human Performance Wing
Air Force Research Laboratory

SIMPSON.RICHARD.D.1091443429 Digitally signed by SIMPSON.RICHARD.D.1091443429 Date: 2019.05.06 11:32:02 -04'00'

RICHARD D. SIMPSON, DR-IV, DAF
Chief, Human Centered ISR Division]
Airman Systems Directorate
711th Human Performance Wing
Air Force Research Laboratory

This report is published in the interest of scientific and technical information exchange, and its publication does not constitute the Government's approval or disapproval of its ideas or findings.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
10-08-2018	Interim	1 January 2017 - 31 May 2017

4. TITLE AND SUBTITLE

Human Activity Synthetic Data Generation

5a. CONTRACT NUMBER
FA8650-12-D-6354 T03

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Zhiqing Cheng[a], Timothy MtCastle[a], Todd Huster[a], Max Grattan[a], John Camp[b], Huaining Cheng[b], and Monique Brisson[b]

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER
H07U

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

[a]Infoscitex Corporation, a DCS Company
4027 Colonel Glenn Hwy
Dayton, Ohio

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

[b]Air Force Materiel Command
Air Force Research Laboratory
711th Human Performance Wing
Airman Systems Directorate
Human-Centered ISR Division
Human Signatures Branch
Wright-Patterson AFB, OH 45433

10. SPONSORING/MONITORING AGENCY ACRONYM(S)
711 HPW/RHXB

11. SPONSORING/MONITORING AGENCY REPORT NUMBER(S)
AFRL-RH-WP-TR-2018-0058

12. DISTRIBUTION/AVAILABILITY STATEMENT
Distribution A: Approved for public release.

13. SUPPLEMENTARY NOTES
Report contains color. 88ABW-2017-2676 26May17
Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC) 2017

14. ABSTRACT

In this paper, using human activity modeling and simulation to generate synthetic human activity data for machine learning is investigated. The needs for synthetic data are identified from the perspective of human centric, computer-vision based technology development. The basic requirements of synthetic data are defined in light of machine learning. Factors that contribute to the fidelity and applicability of synthetic data are analyzed. In particular, two factors related to human activity modeling and simulation, bio-fidelity and variability are investigated. Several modeling and simulation tools and game engines (e.g., 3dsMAX, Unity, and NVIG) are used for data generation, and their performances are compared and evaluated. Synthetic full motion videos are generated in electric-optical and infrared modes and tested by machine learning algorithms. The testing results along with examples of synthetic imagery are illustrated in the paper

15. SUBJECT TERMS

human activity, modeling and simulation, synthetic image, synthetic full motion video, machine learning

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT:	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON (Monitor)
a. REPORT	b. ABSTRACT	c. THIS PAGE			John Camp
Unclassified	Unclassified	Unclassified	SAR	15	19b. TELEPHONE NUMBER *(Include Area Code)* N/A

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39-18

Human Activity Synthetic Data Generation

Zhiqing Cheng, Timothy MtCastle
Todd Huster, and Max Grattan
Infoscitex Corporation
Dayton, Ohio
{zcheng, tmtcastle}@infoscitex.com
{thuster, mgrattan}@infoscitex.com

John Camp, Huaining Cheng, and Monique Brisson
711th Human Performance Wing
Air Force Research Laboratory
Dayton, Ohio
{John.camp.1, huaining.cheng}@us.af.mil
Monique.brisson@us.af.mil

ABSTRACT

Data availability often becomes a major hindrance to the development of human centric, computer-vision based technologies, as a large amount of data is usually required for algorithm training and validation, especially when deep learning is used to develop algorithms. Synthetic data which are produced by modeling and simulation could be used to expand and/or supplement real world data which are otherwise not available. While human activity modeling and simulation has achieved success in creating synthetic environments for simulation based training and virtual reality, whether it can be used to generate synthetic data which satisfy requirements for machine learning is yet to be proved.

In this paper, using human activity modeling and simulation to generate synthetic human activity data for machine learning is investigated. The needs for synthetic data are identified from the perspective of human centric, computer-vision based technology development. The basic requirements of synthetic data are defined in light of machine learning. Factors that contribute to the fidelity and applicability of synthetic data are analyzed. In particular, two factors related to human activity modeling and simulation, bio-fidelity and variability are investigated. Several modeling and simulation tools and game engines (e.g., 3dsMAX, Unity, and NVIG) are used for data generation, and their performances are compared and evaluated. Synthetic full motion videos are generated in electric-optical and infrared modes and tested by machine learning algorithms. The testing results along with examples of synthetic imagery are illustrated in the paper.

Keywords: human activity, modeling and simulation, synthetic image, synthetic full motion video, machine learning

ABOUT THE AUTHORS

Dr. Zhiqing Cheng is a principal engineer and a program manager of Infoscitex Corporation, providing support to the Air Force Research Laboratory (AFRL) for the programs of human identification and activity recognition based on human bio-signatures. He has vast research experience in the areas of human modeling and simulation, artificial intelligence and machine learning, computer vision, and optimization. He has assumed many R&D projects as the principal investigator and published over 80 technical papers as the lead author. Currently, as the technical lead and program manager, he is leading the HMASINT program, a multi-directorate endeavor by the AFRL, to develop technologies for human measurement and intelligence information.

Dr. John Camp is a senior computer research scientist employed by the United States Air Force Research Laboratory (AFRL). Dr. Camp is the Program Manager for the Human Detection and Characterization program. As a tech lead and program manager, he has been responsible for the successful development and implementation of a wide variety of products representing virtual humans in modeling and simulation as well as algorithm and tool development for the detection and characterization of humans in imagery data. His research interests include computer vision and graphics. Dr. Camp received a B.S. in mathematics from the University of Florida, a M.S. in computer systems from the Air Force Institute of Technology, and a Ph.D. in Computer Engineering from Wright State University.

Human Activity Synthetic Data Generation

Zhiqing Cheng, Timothy MtCastle
Todd Huster, and Max Grattan
Infoscitex Corporation
Dayton, Ohio
{zcheng, tmtcastle}@infoscitex.com
{thuster, mgrattan}@infoscitex.com

John Camp, Huaining Cheng, and Monique Brisson
711th Human Performance Wing
Air Force Research Laboratory
Dayton, Ohio
{John.camp.1, huaining.cheng}@us.af.mil
Monique.brisson@us.af.mil

INTRODUCTION

Data availability often becomes a major hindrance to the development of human centric, computer-vision based technologies, as a large amount of data is usually required for algorithm training and validation, especially when deep learning is used to develop algorithms. Synthetic data produced by modeling and simulation (M&S) could be used to expand and/or supplement real world data which would be otherwise not available. While human activity M&S has achieved success in creating synthetic environments for simulation based training and virtual reality, whether it can be used to generate synthetic data which satisfy requirements for machine learning (ML) is yet to be proved.

The usage of synthetic data is not novel and has been explored in various contexts. Nonnemaker and Baird (2009) utilized synthetic data generated via parameter space interpolation in training handwriting classifiers. The classifiers trained with synthetic data performed equally or better than their counterparts trained with real data. Keskin et al. (2011) achieved a high level of accuracy when recognizing hand gestures backed by a random decision forest (RDF) trained using synthetic data designed to account for unseen gestures within the target application space. Synthetic data was generated through extrapolating, interpolating, and randomizing components of a set of manually positioned 3-D hand model base gestures. Shotton et al. (2011) developed high levels of accuracy when estimating 3-D positions of body joints using an RDF with synthetic and real data. Motion capture data was recorded using humans in varying poses and with varying attributes and synthetic depth imagery is generated based upon those initial positions combined with various parameter perturbations. Danielsson and Aghazadeh (2014) and Ullah and Laptev (2012) used synthetic training data in other human related recognition tasks such as improvements to human pose estimation from images and human action estimation from videos.

Fanelli et al. (2011) used synthetic, neutral expression, 3-D face patches to train a RDF for effective human head pose estimation. Similarly but with a deep convolutional neural network (DCNN), Peng et al. (2014) determined that when trained using public 3-D models in addition to real data, a DCNN was mostly invariant to missing minute details such as texture or pose when used for a specific task. The authors also determined that when trained for generic classification using real images the DCNN performed better when the added synthetic data contained such details. Validated on two different datasets, Zhang et al. (2015) described a method of modifying synthetic data to more closely resemble observed data, decreasing the drift between both. The method established a classifier learning process that benefited from both observed and synthesized data. Ros et al. (2016) generated a synthetic collection of diverse urban images, named SYNTHIA with automatically generated class annotations. They used SYNTHIA in combination with publicly available real-world urban images with manually provided annotations. They conducted experiments with DCNNs that show how the inclusion of SYNTHIA in the training stage significantly improves performance on the semantic segmentation task.

These investigations indicate that synthetic data is viable for training despite detail differences in some applications and the details may play a critical role in some instances. Synthetic data is also shown to have utility as a sole training input as well as an additive data source. Synthetic training data can be useful and applicable and provide many benefits when modified to appropriately account for real data characteristics and application specifics. Experimentation is necessary to determine applicability and to identify methods of modifying artificial input data or the underlying algorithms to maintain target performance when using natural data. A single, generic, nominal methodology for evaluating applicability and optimizing synthetic training data does not currently exist.

In this paper, using human activity M&S to generate synthetic human activity data for machine learning is investigated. The needs for synthetic data are identified from the perspective of human centric, computer-vision based technology development. The basic requirements of synthetic data are defined in light of ML. Factors that contribute to the fidelity and applicability of synthetic data are analyzed. In particular, two factors related to human activity M&S, bio-fidelity and variability are investigated. Several M&S tools and game engines (e.g., 3dsMAX, Unity, and NVIG) are used for data generation, and their performances are compared and evaluated. Synthetic full motion videos (FMVs) are generated in electric-optical (EO) and infrared (IR) modes and tested by ML algorithms. The testing results along with examples of synthetic imagery are illustrated in the paper.

NEEDS AND REQUIREMENTS FROM MACHINE LEARNING

From the machine learning perspective, basic requirements for synthetic images are as follows.

- Fidelity: The models used to generate synthetic images should provide representation of the real world with sufficient fidelity. Synthetic images should contain image features that are the same as or similar to those embedded in real world images so that these features can be learnt or utilized by ML.
- Compatibility: Synthetic images should be beneficial by themselves and when combined with limited amounts of real world examples.
- Variability: Data variability and volume are key in order to reap the benefits of the discriminating power of ML (deep learning in particular).
- Computational efficiency: In order to effectively integrate into the streamlined ML workflow, a synthetic image generation (SIG) tool should be able to generate synthetic images at sufficiently high computational efficiency.

For supervised learning, which is the main format of machine learning, annotating or truthing needs to be performed on the data to obtain the ground truth information. Data annotating or truthing by human is labor extensive, especially for images and full motion videos (FMV) and is prone to human errors. When computer algorithms are required to utilize multi-modal sensor data, the problem become more pronounced. This is because (a) The availability of real world multi-modal sensor data is much more restricted; (b) The synchronization of different modalities is often hard to achieve; (c) The annotation becomes more onerous; and (d) The fusion of multiple modality real world data is still a challenging task.

APPROACHES

Technology Exploration

From the sources of commercial off-the-shelf (COTS) and government off-the-shelf (GOTS) and open source, we have identified a list of the state-of-art software tools and systems that are commonly used for M&S and games.

- 3D modeling tools: Blender (open source), 3dsMax (COTS), Maya (COTS), VIPRS (GOTS);
- Game Engines: Unity (open source), VBS 3 (COTS), CryEngine (COTS), Unreal Engine (COTS), MetaVR (COTS), NVIG (IR image generator, GOTS);
- Multi-modal sensor modeling tool: DIRSIG (GOTS, http://dirsig.org/index.html)
- Atmospheric radiation modeling tool: MODTRAN (GOTS, http://modtran.spectral.com/).

These existing tools and packages could potentially be leveraged to meet many of these requirements; however, none of them meets all requirements stated above and can provide full capabilities for SIG oriented to ML. In this paper, we focus on the SIG of two most common sensor modalities, EO and IR. We chose to use 3dsMax and Unity for EO image generation and NVIG for IR image generation. We realize that while the tools being selected can be used to generate synthetic images with certain level of fidelity, none of them are essentially based on physics or first principles. Therefore, we are performing another study to utilize physics or first principle based M&S tools, such as Digital Imaging and Remote Sensing Image Generation (DIRSIG). The details of this study will be reported separately.

SIG Procedure

In this paper, we focus on the SIG of EO and IR. We developed a procedure for SIG as follows.

1. **Scenario Storyboarding:** Based on the requirements for a given project, a storyboard is created to outline what a scenario will look like and identify the assets required to meet those requirements. Assets include but are not

limited to (a) Environment (e.g., terrain, street, buildings, trees, and vegetation); (b) Objects (e.g., vehicles and weapons); (c) Human subjects (e.g., human shape models, cloth and other body wear, and carrying objects); (d) Clutters (e.g., discarded tires and trash cans).

2. **Asset Aggregation and Composition:** Pull together assets that meet the requirements for a given scenario storyboard. During asset aggregation moving objects are also linked to the required animation files and path information. All required assets can then be added together to create the synthetic scenario. Sensor metadata (position, resolution, orientation, etc.) is used to define the imaging scenario environment.
3. **Raw Data Generation:** The user starts the animation and rendering begins. As the animation progresses, images are generated based on the imaging sensor parameters and the predetermined motions of the assets in the scene. Ground truth information on the location and attributes of the assets are automatically generated and output for each frame in the form of JSON files.
4. **Post Processing:** In order to better simulate real-world atmospheric effects, post processing of the imagery can be performed. Image filters can be designed with the help of atmospheric modelling tools, such as MODTRAN, to better mimic these effects.

Factors Considered

Many factors contribute to the fidelity and applicability of synthetic data, including material properties, sensor physics, environmental and atmospheric effects when SIG is based on first principles. For human activity synthetic image generation in particular, two factors related to human activity M&S are bio-fidelity and variability. Besides, for synthetic images to be used for ML, ground truth data need to be provided. In the flowing, we will describe our efforts made to improve the bio-fidelity and variability of human activity M&S, to address atmospheric effects, and to develop plug-ins for automatic ground truthing.

1. Biofidelity

The biofidelity of human activity modeling relies on true body shape and true body motion. High biofidelity can be attained through activity replication (Cheng et al., 2012). Activity replication is replicating a human activity that was recorded from a human subject in a laboratory using 3-D modeling. Technologies that are capable of capturing human motion and 3-D dynamic shapes of a subject during motion are not yet ready for practical use. Data that can be readily used for 3-D activity replication are not currently available. Alternatively, a motion capture system can be used to capture markers on the body during motion and a 3-D body scanner can be used to capture the body shape in a pose. Based on the body scan data and motion capture data, M&S techniques can be used to build a digital model to replicate a human activity in 3-D space. In the paper by Cheng et al. (2012), open-source software was used for activity replication. MeshLab (http://meshlab.sourceforge.net/) was used to process 3-D scan data, OpenSim (http://opensim.stanford.edu/) was used to derive skeleton models and the associated joint angles from motion capture data, and Blender (https://www.blender.org/) was used to create an animation model that integrates body shape and motion. Figure 1 shows the models created for four activities (jogging, limping, shooting, and walking) at a particular frame. Note that activity replication can be done using commercial modeling tools (e.g., 3dsMax and Unity which are used in this paper).

Figure 1. Replication of a subject in four activities: limping, jogging, shooting, and walking

2. Variability

The data sets collected at the Air Force Research Laboratory (AFRL) 3-D Human signatures Laboratory (3DHSL) have been used to create a human activity model library that provides a larger variability of human shape and motion (Camp et al., 2013). The variability can be further expanded by human shape morphing and human motion mapping.

• *Morphing* As soon as the point-to-point correspondence is established among shape models, one shape can be gradually morphed to another by interpolating between their vertices or other graphic entities. In order to create a faithful intermediate shape between two individuals, it is critical that all features are well-aligned; otherwise, features will cross-fade instead of move. Figure 2 illustrates an example if shape morphing from one male subject to a female subject (Cheng et al., 2009). Using morphing, new models can be created that resemble to the models being morphed and still provide high biofidelity.

• *Motion Mapping* It is desirable to map the motion from one subject to another, because it is not feasible to do motion capture for every subject and for every motion or activity. By assuming that different subjects will take the same key poses in an action or motion, one approach is mapping joint angles from one to another, as shown in Figure 3 where motion is mapped onto 3dsMax biped models (Cheng et al., 2012). Note that since the pelvis is usually treated as the reference segment, the hip joint center vertical location needs to be adjusted to reflect the variation of subject size in order to ensure appropriate contact between the feet and the ground. While motion mapping may be fairly natural and realistic, it may not be able to provide sufficiently high biofidelity, because the differences between human bodies and the interaction between human body and boundaries are ignored.

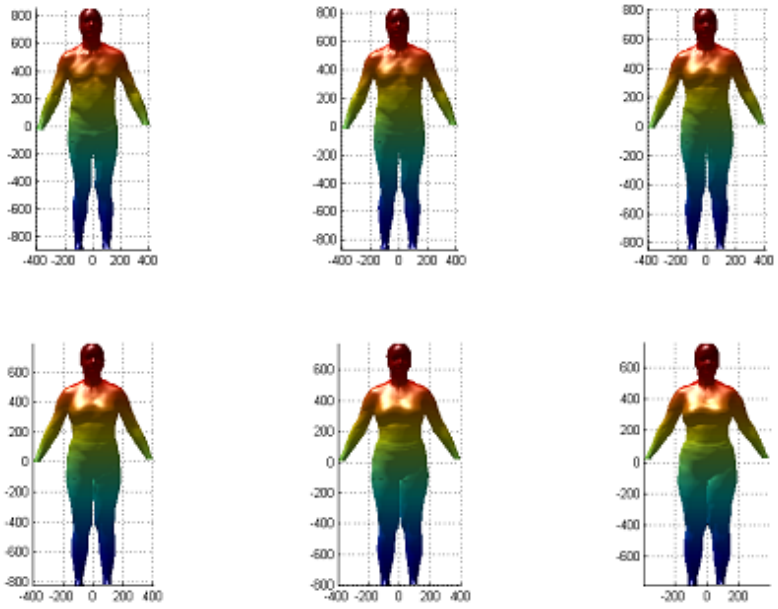


Figure 2. Morphing from one subject to another

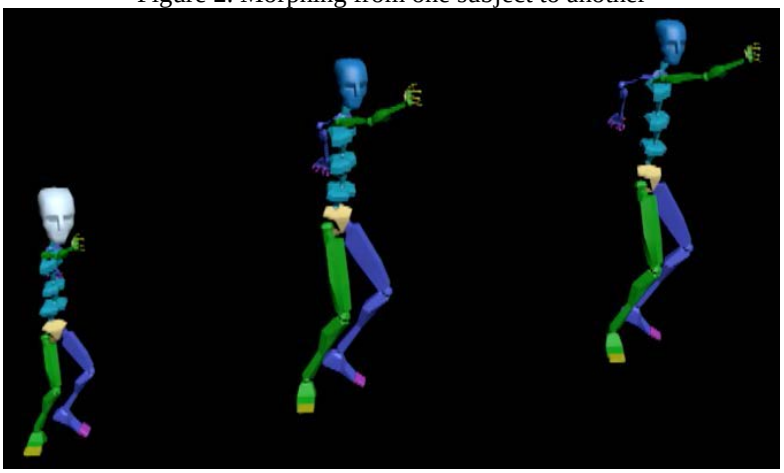


Figure 3. Mapping the captured motion into a group

3. Atmospheric Effects

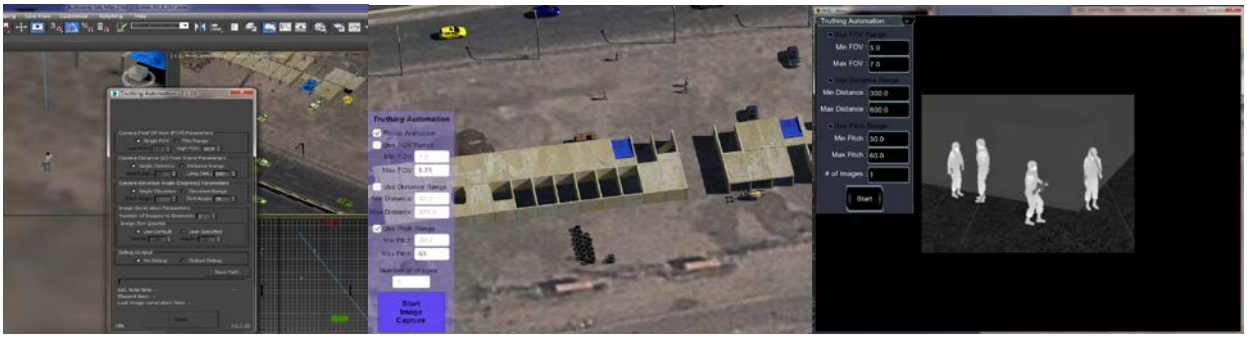
According to radiometry, the spectral content of a target image is a function of the illumination spectrum and the reflectivity spectrum, which must pass through the atmosphere to get to the sensor. Depending on the path length, atmospheric effects can be a negligible or dominant factor. In cases where atmospheric effects are dominant, the image quality can be deteriorated in many ways, including loss of signal level and loss of resolution.

MODTRAN is an atmospheric radiative transfer model developed by Spectral Sciences Inc. and AFRL. MODTRAN has been extensively validated and it serves as a standard atmospheric band model for the remote sensing community. In MODTRAN, the atmosphere is modeled as stratified (horizontally homogeneous), and its constituent profiles, both molecular and particulate, are defined either using built-in models or by user-specified vertical profiles. The spectral range extends from the UV into the far-infrared (0 – 50,000 cm-1), providing resolution as fine as 0.2 cm-1. MODTRAN solves the radiative transfer equation including the effects of molecular and particulate absorption/emission and scattering, surface reflections and emission, solar/lunar illumination, and spherical refraction. Additionally, the MODTRAN model allows a sensor to be placed on the ground looking up to space in any direction. Integrating the observed radiance from several angles effectively computes the down-welled radiance under the given weather conditions.

While MODTRAN can be utilized to simulate atmospheric effects, in this paper we chose to apply two simple methods on the raw synthetic images in the post-processing process. One is applying an image filter from Adobe Premier, the other is using a random "Rayleigh" atmospheric model.

4. Ground Truth Data

Ground truth data is composed of labeled features such as foreground, background, and objects or features to recognize. The labels define exactly what features are present in the images, and these labels may be a combination of on-screen labels, associated label files, or databases. In order to generate ground truth data of synthetic images automatically, which is one of benefits of SIG, we have developed independent software modules which are used as the plug-ins to respective 3D modeling tools and game engines. These include the plug-ins for 3dsMax, Unity, and NVIG, as shown in Figure 4. With the plug-ins, when images are generated from these tools, ground truth data are generated automatically and provided along with synthetic images. The ground truth data are presented in a JSON-based format that they can be readily utilized by ML.



(a) 3dsMax (b) Unity (c) NVIG

Figure 4. Plug-ins for automatic ground truth data generation

CASE STUDIES AND EVALUATION

EO Image Generation

In developing computer vision based technologies for human detection and characterization, we performed SIG to provide data required for ML. A real world image, as shown in Figure 6(a), was selected as the reference. We used 3dsMax to create a 3D model to replicate the scenario and to generate synthetic images, as shown in Figure 5. Using the 3D model created in 3dsMax, synthetic images are also generated using Unity, as shown in Figure 6.

(a) Unfiltered

(b) Filtered

Figure 5. Images generated from 3dsMax

(a) Real imagery (b) Unity output

(c) "Hand-crafted" atmosphere (d) Random "Rayleigh" atmosphere

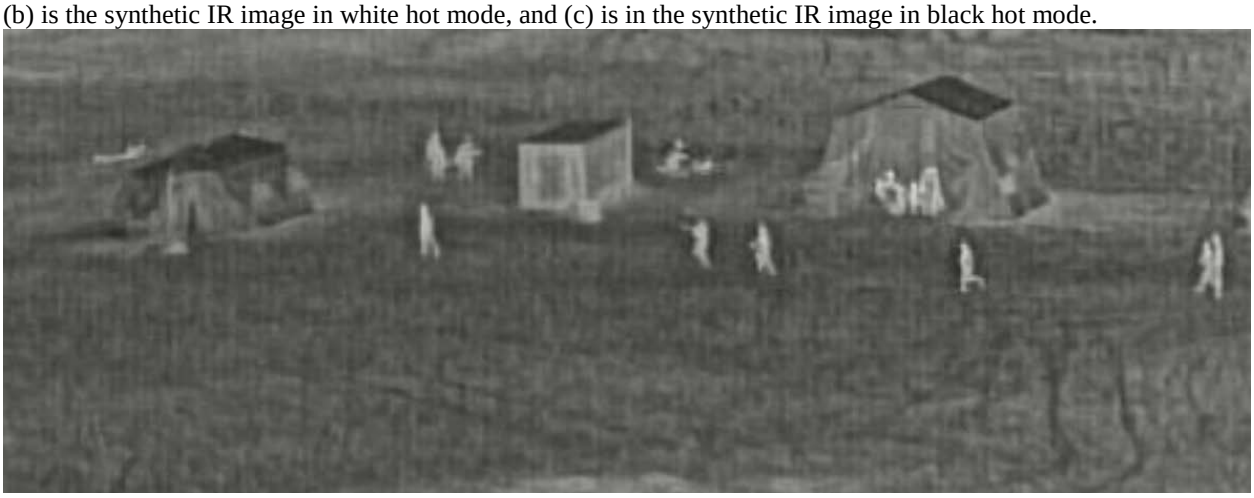
Figure 6. Real world image and the images generated from Unity

Comparing Figure 5(a) and Figure 6(b) with Figure 6(a), we can find that the raw or original images generated from image generators (3dsMax and Unity) are quite different from the real image in terms of color appearance. This can be due to the lack of atmospheric effects on the original synthetic images. Therefore, we applied image filters on the

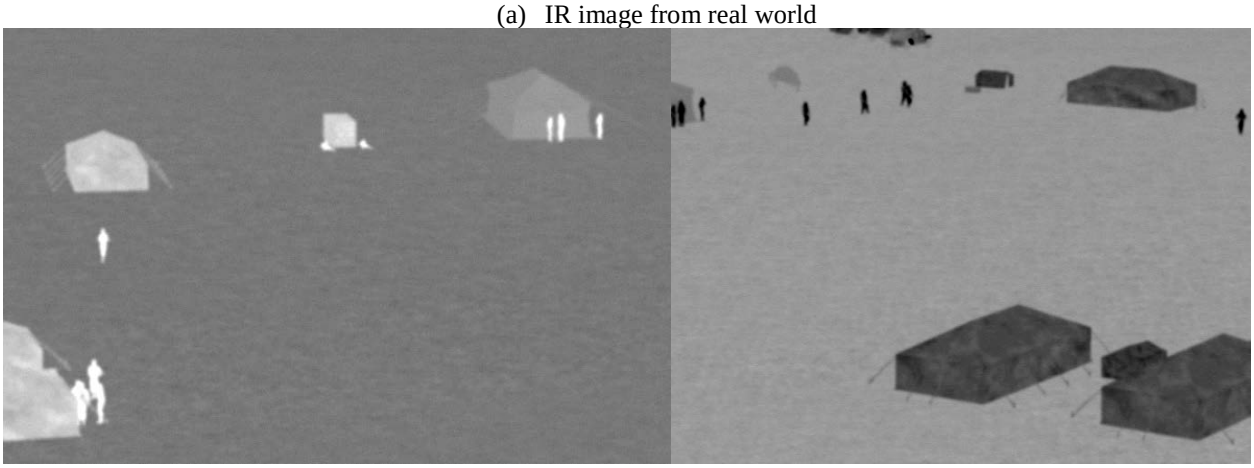
raw images to improve their color appearance, as shown in Figure 5 (b) and Figure 6 (c) and (d). We performed image analysis to determine color spectrum distribution of the real image and synthetic images to evaluate the effects of different type of filters on the image color distributions.

IR Image Generation

We chose to use NVIG to generate IR images. An example is show in Figure 7 where (a) is a real world IR image, (b) is the synthetic IR image in white hot mode, and (c) is in the synthetic IR image in black hot mode.



(a) IR image from real world



(b) Synthetic IR image—white hot (c) Synthetic IR image—black hot

Figure 7. Synthetic IR images to mimic real world image

Testing and Evaluation in ML

It can be challenging to quantitatively assess the quality of synthetic imagery. One significant goal of our work is to aid in the training of ML algorithms on real world data. In many cases, an algorithm is utilized in an environment that is different from the training environments. In these cases, synthetic imagery can lower the risk of failure, either by specifically generating data that is very close to the target environment or by generating data from a very large set of environments that most likely includes the target environment. We focus here on the first case.

We selected 11 short clips from a particular aerial data collection and divided them into training and testing sets of 8 and 3 clips, respectively. All the clips were taken over the same region, but from different times, angles, and zoom levels. We used 3ds Max, Unity, and Adobe Premier to model the scene/environment and atmosphere, as shown above. We trained a region proposal network (RPN) to detect pedestrians with the real and synthetic training data sets and compared the results. We tuned the hyper-parameters of the training process and network configuration on a different set of real aerial data and used the same hyper-parameters for both the real and synthetic training. Results, shown in

Figure 8, show a drop when moving from real to synthetic training, but the method shows promise, as this outperforms the same algorithm when trained on dramatically different real data.

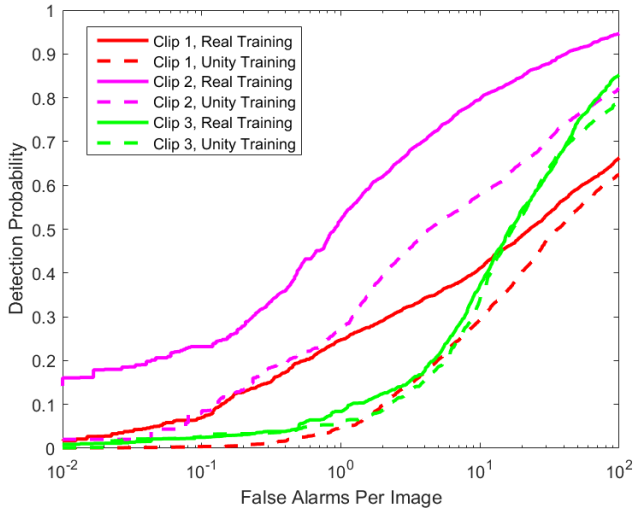
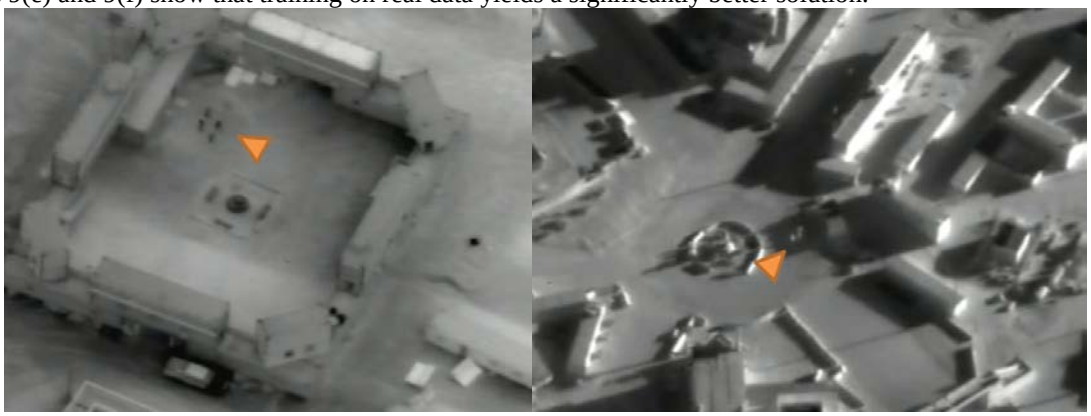
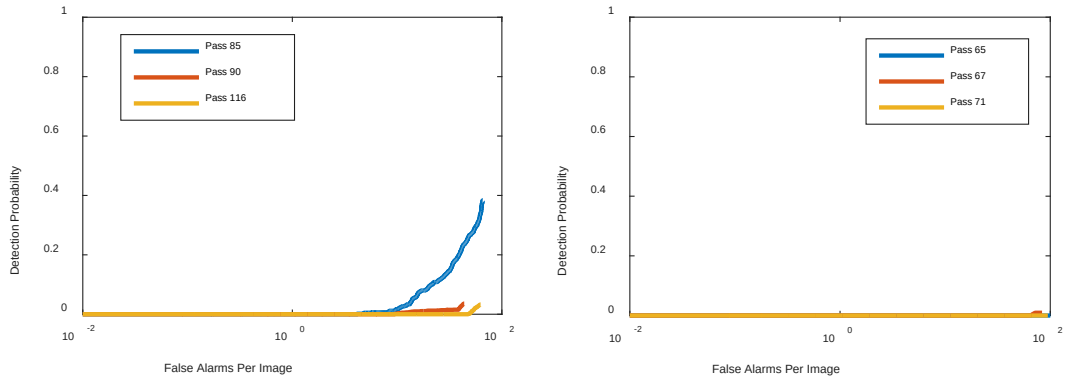


Figure 8. Performance evaluation of synthetic EO images

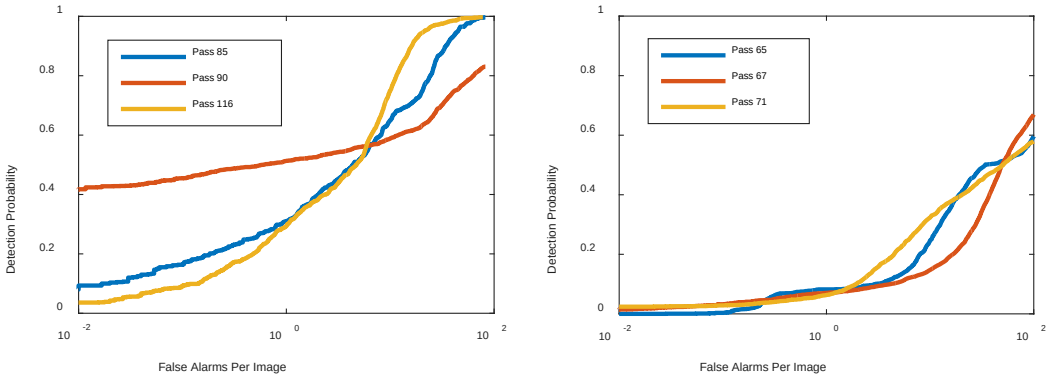
We also performed the same experiment on IR data, using NVIG as the engine. Here, we have both white hot and black hot collections, divided into training and test partitions in similar fashions to the EO clips. As can be seen above in Figure 7, the synthetic IR data was not as convincing as the synthetic EO data. In the NVIG data, the people are always the "hottest" objects in the scene and saturate the sensor. That does not hold in the real data, shown in Figures 9(a) and 9(b). Figures 9(c) and 9(d) show that performance is essentially zero when trained on synthetic data, while Figures 9(e) and 9(f) show that training on real data yields a significantly better solution.



(a) Example black hot data (b) Example white hot data



(c) Black hot data: synthetic training data (d) White hot data: synthetic training data



(e) Black hot data: real training data (f) White hot data: real training data

Figure 9. Performance evaluation on synthetic IR images

DISCUSSION

Based on the investigations we performed on SIG and results presented in this paper, the following observations are in order.

1. Regarding synthetic EO image generation, either 3dsMAX or Unity can be used to meet basic requirements. Using ray tracing, 3dsMAX produces images with high quality; however, it requires substantial time for image generation. On contrary, Unity can produce images at much higher rate; however, its image quality is lower. Table 1 shows the comparison of computational efficiency between 3dsMax and unity. The simulations were performed on a high-end personal GPU workstation.
2. The original images produced by 3dsMAX and Unity lack the color variation induced by atmospheric effects. Applying filters on the original images can reduce the differences of color appearance between synthetic images and real images. However, there are many unknowns about image filtering, which leads the post-processing of raw images remains as the process of trial and error. This can lead to images that look correct to the human eye but are totally different in computed color space.
3. Atmospheric effects need to be and can be addressed more rigorously. For instance, MODTRAN, a widely used atmospheric radiation propagation model, can be plugged into Unity to simulate the effects caused by atmosphere, including weather, time of day, and geo-location.
4. Regarding synthetic IR image generation, whereas NVIG can meet minimum requirements, it is not based on physics or first principles. Therefore, the fidelity and level of details of human IR images generated from NIVG are limited, which reduces the performance of algorithms trained with synthetic IR images.

The bio-fidelity of human models contributes to the fidelity of synthetic images. For the images from aerial platforms with low resolution, the importance of human body features diminishes. However, for the images from ground stations with high resolution, human body features matter to ML algorithms. Since an integrated 3-D model is to be used to generate synthetic images from any view angle at any distance with varying resolutions, it is important for human models used in the integrated 3-D model to have high bio-fidelity. When human motion information is utilized by ML for human detection and identification, human activity recognition, and patter of life analysis, synthetic images based on true human motion will help improve the performance of ML algorithms. The contents of real world images often lack the desired human variability in gender, age, ethnicity, body shape, and clothing for ML. This shortage can be remedied by utilizing a large variety of human models in SIG. One example is the identification of adult versus child in our development of human detection and characterization technology. Since real world data is not available for ML, we have used synthetic images with adult and child models populated in the scene.

Synthetic image generation is often limited to single sensor modality and not human-focused. In order to develop ML algorithms for image related problems (e.g., human threat detection and recognition), a high-fidelity M&S tool is needed to generate synthetic, multi-modal sensor datasets. One of such tool is DIRSIG, a first-principle based sensor modeling and simulation tool that has been actively developed at the Digital Imaging and Remote Sensing (DIRS) Laboratory at Rochester Institute of Technology (RIT) for two decades. It is designed to generate passive broadband, multi-spectral, hyper-spectral low light, polarized, active laser radar, and synthetic aperture radar datasets through the

integration of a suite of first-principles based radiation propagation modules. However, DIRSIG lacks sufficient variability, usability, and speed which are required for SIG oriented to ML.

Table 1. Comparison of computational efficiency between 3dsMax and Unity

Tool	Per frame—w/o truthing	Per frame— w/ truthing
3dsMax	~40 (s)	~55 (s)
Unity	~30 (ms)	~1.5 (s)

CONCLUSIONS

The study of this paper shows that synthetic data which are produced by M&S could be used to expand or supplement real world data which are otherwise not available. The existing M&S tools or game engines, such as 3dsMax and Unity, can be used to generate synthetic EO images that meet basic requirements of machine learning. Raw synthetic images need to be further processed with filters to include the color variation induced by atmospheric effects. Synthetic IR image generation presents more challenges and requires more advanced IR image generator. Investigations are needed to test and evaluate the bio-fidelity and variability of human activity M&S from ML perspective.

An integrated, effective software system for SIG is desired by the ML community. Such a SIG system can be integrated into the ML pipeline and become a handy tool at a ML researcher's disposal. It would (a) expedite the ML process which may otherwise be hindered by data availability; (b) simplify the ML process by reducing or eliminating the time required for data annotation; and (c) enable ML to be usable for those problems that lack real world data. Synthetic images and videos can be used by ML for a wide range of problem domains that include but are not limited to (a) image, object, and scene classification; (b) object detection and localization; (c) image captioning and visual questioning and answering; and (d) video captioning and activity classification.

REFERENCES

Camp, J., Lochtefeld, D., Cheng, Z., Davenport, I., MtCastle, T., Mosher, S., Smith, J., and Grattan, M. (2013). Bio-fidelic Human Activity Modeling and Simulation with Large Variability. *Proceedings of I/ITSEC 2013*, Orlando, Fl.

Cheng, Z., Mosher, S., Camp, J., & Lochtefeld, D. (2012). Human Activity Modeling and Simulation with High Bio-fidelity, *Proceedings of I/ITSEC 2012*, Orlando, Fl.

Cheng, Z., & Robinette, K. (2009). Static and Dynamic Human Shape Modeling. *Lecture Notes in Computer Science 5620*. Springer 2009, ISBN 978-3-642-02808-3.

Danielsson, O. and Aghazadeh, O. (2014). Human Pose Estimation from RGB Input Using Synthetic Training Data. *arXiv:1405.1213v2 [cs.CV]*, 27 May, 2014.

Fanelli, G., Gall, J., and Gool L. (2011). Real Time Head Pose Estimation with Random Regression Forests. *Computer Vision and Pattern Recognition*, pages 617–624, June 2011.

Keskin, C., Kıraç, F., Kara, Y., and Akarun, L. (2011). Real Time Hand Pose Estimation Using Depth Sensors. *Consumer Depth Cameras for Computer Vision, Advances in Computer Vision and Pattern Recognition*, pages 119–137. Springer London, June 2011.

Nonnemaker, J. and Baird, H (2009). Using synthetic data safely in classification. *Proc. SPIE 7247, Document Recognition and Retrieval XVI*, 2009.

Peng, X., Sun, B., Ali, K., and Saenko K. (2014). Exploring Invariances in Deep Convolutional Neural Networks Using Synthetic Images. arXiv:1412.7122v4 [cs.CV], 2014.

Ros, G., Sellart, L., Materzynska, J., Vazquez, D., Lopez, A. (2016). The SYNTHIA Dataset: A Large Collection of Synthetic Images for Semantic Segmentation of Urban Scenes. Proc. CVPR 2016, pages 4321-4330.

Shotton, J., Fitzgibbon, A., Cook, M., Sharp, T., Finocchio, M., Moore, R., Kipman, A., and Blake, A. (2011). Real-Time Human Pose Recognition in Parts from a Single Depth Image. *Computer Vision and Pattern Recognition*, June 2011.

Ullah, M. and Laptev, I. (2012). Actlets: A novel local representation for human action recognition in video. *Proc. 19th IEEE International Conference on Image Processing*, pages 777–780, Sept 2012.

Zhang, X., Fu, Y., Zang, A., Sigal, L., and Agam, G. (2015). Learning Classifiers from Synthetic Data Using a Multichannel Autoencoder. *arXiv:1503.03163 [cs.CV]*, 2015.