Physics-Based, Multi-Modal Synthetic Human Image Generator

Devendra Tolani, Mun Wai Lee, Eric van Doorn, Daniel Ashley, Gang Mei, Roger Xu
Intelligent Automation Inc.
Rockville, Maryland
{dtolani, mlee, evandoorn, dashley, mgang, hgxu}@i-a-i.com

Zhiqing Cheng	John Kerekes, Nina Raqueno	Huaining Cheng
Innovision, LLC	Rochester Institute of Technology	Air Force Research Laboratory
Dayton, Ohio	Rochester, New York	Wright-Patterson AFB, Ohio
zcheng@innovisionllc.com	{kerekes, nina}@cis.rit.edu	huaining.cheng@us.af.mil

ABSTRACT

There is a need to develop a physics-based 3D modeling and simulation (M&S) software to generate multi-modal datasets for machine learning of human activity detection and recognition, due to the high cost and difficulty in collecting synchronized, multi-view human sensing data. This paper presents the effort in developing a novel, integrated, high-fidelity M&S tool: HumanView (HumV) of human signatures. Its key elements include: (a) HumV editor module, which allows users to view and manage available models and associated configurations using an intuitive graphical user interface; (b) HumV models module, which is a data store containing human models, scene models of the environment, and relevant electro-optical/infrared (EO/IR) sensor models; and (c) HumV simulator module, which allows users to simulate multiple scenarios for generation of synthetic sensor data and ground truth labels for analytics. HumV has established a pipeline that seamlessly integrates off-the-shelf free and open-source multi-physics M&S tools and material properties databases with newly developed models and algorithms to address the multi-disciplinary M&S requirements. Specifically, we have developed the Human Activity Replication Tool (HART) - a Blender 3D add-on to provide bio-fidelic M&S of clothed avatars that realistically represent the diversity of human shape, motion, and clothing characteristics. This is followed up by an innovative human thermal model that takes the scene and HART activity models to produce an output of temperature estimates for all the mesh facets of skin and clothing of the human avatars. The thermal dynamics considers the activity/heart rate, environmental radiance, and body/clothing interaction. Finally, various models of human activity, thermal dynamics, scene, materials, environment, sensor, and atmosphere are assembled into the Digital Imaging and Remote Sensing Image Generation (DIRSIG) tool to generate synthetic images or videos. Model validation has been conducted against the experimental data collected using commercial cameras in an outdoor setting.

ABOUT THE AUTHORS

Dr. Devendra Tolani is the Senior Director of the Signals, Analysis and Controls Division at Intelligent Automation, Inc. (IAI). Dr. Tolani received his Masters and Ph.D. in Mechanical Engineering from Pennsylvania State University. He has been the Principal Investigator (PI) on over forty research projects funded by various government agencies. His research interests include Modeling and Simulation, Machine Learning, Sensor Technology, Signal Processing/Analysis, and Systems Engineering.

Dr. Mun Wai Lee is a Technical Director at Intelligent Automation, Inc. and leads the machine vision group in the company. Dr. Lee received the B.Eng and M.Eng degrees from the Department of Electrical Engineering, the National University of Singapore in 1998, and 1999 respectively. He received his Ph.D. degree from the Department of Computer Science at the University of Southern California in 2006. He became an Associate Director in 2016 and now serves as a Principal Scientist. Dr. Lee has been the Principal Investigator for projects sponsored by the DoD, DHS, and NASA. His research interests include computer vision, video-based detection and tracking, human motion capture and behavior analysis, image scene understanding, machine learning, and natural language processing.

Dr. Eric van Doorn is the Chief Technology Officer (CTO) at IAI. Dr. van Doorn received his B.Sc. and M.Sc. from the Free University in Amsterdam, The Netherlands in 1987 and 1991 respectively, and a Ph.D. in Physics from Duke University, Durham, NC in 1997 majoring in physics of fluids. Since 2001, Dr. van Doorn has focused on building a research program in Radio Frequency sensing at Intelligent Automation, Inc., including several Ultra Wide Band (UWB) projects with Army, NAVAIR, NASA, and DARPA. Recently, Dr. van Doorn is focusing on maturing and transitioning RF and other company-wide technologies at IAI for commercial applications and insertion in large government programs.

Mr. Daniel Ashley is a Senior Robotics Engineer who works primarily in the field of sensor simulation and robotic prototype design. He received an M.S. in Computer Engineering at the University of South Florida (USF) in 2014. He has also achieved B.S. degrees in both Computer Engineering and Computer Science with a minor in Business Management from USF in 2012. Prior to joining IAI in 2016, Mr. Ashley worked as a Graduate Assistant for CARRT labs at USF focusing on Assistive and Rehabilitative robotic technologies. At IAI he works on applying these technologies to sensor simulation and advanced inspection systems and robotic prototyping.

Mr. Gang Mei is a Soutions Architect at Intelligent Automation, Inc. Mr. Mei received his M.S. in 2000 in Electrical Engineering from the University of Maryland, College Park. His primary research areas include signal processing, machine learning/deep learning, voice-based mental healthiness monitoring, wireless communication system design and implementation, sensor networks, and embedded system design and implementation.

Mr. Roger Xu is a Senior Vice President of the Signals, Analysis and Controls Division at IAI. He has extensive experience in data mining, machine learning, and technology transition. He obtained his BS degree in 1982 from Jiangsu University and MS degree from Xi'an Jiaotong University in 1988, in China. His research interests include the application of machine learning to audio and image processing, big data analytics, control systems, and network security. He has directed most of the technical activities at IAI in machine learning, signal processing, image processing, control system design, and implementation areas.

Dr. Zhiqing Cheng is the founder of Innovision, LLC., a fast-growing startup focusing on developing digital human models and using these models for human-centered products and services. Dr. Cheng has vast research experience in the areas of human modeling and simulation, artificial intelligence and machine learning, computer vision, and optimization. He had worked at Wright-Patterson Air Force Base as an on-site contractor, providing support to the Air Force Research Laboratory on the programs of human modeling and simulation, human identification, and human activity recognition. He has assumed many R&D projects as the principal investigator.

Dr. John Kerekes received the B.S., M.S., and Ph.D. degrees from Purdue University, West Lafayette, IN, USA in electrical engineering. From 1989 to 2004, he was a Technical Staff Member with the Lincoln Laboratory, Massachusetts Institute of Technology, Lexington, MA, USA. Since 2004, he has been with the Chester F. Carlson Center for Imaging Science, Rochester Institute of Technology, Rochester, NY, USA, where he is currently Professor. From September 2016 to May 2019, he served as the Director of RIT's Digital Imaging and Remote Sensing Laboratory. His research interests include the modeling and analysis of remote sensing system performance in pattern recognition and geophysical parameter retrieval applications.

Dr. Nina Raqueno is a research staff member at Rochester Institute of Technology's Digital Imaging and Remote Sensing Laboratory. She received a Bachelor in Imaging Science from RIT in the 1990s. She also attended SUNY College of Environmental Science and Forestry where she learned practical Remote Sensing tools, such as GIS, GPS, surveying, flight planning, and much fieldwork. Nina has over 20 years of experience coordinating remote sensing field campaigns of all sizes and for a variety of platforms, including satellite, aircraft, rooftop, and now from UAS.

Dr. Huaining Cheng is a senior research computer scientist at the 711th Human Performance Wing, United States Air Force Research Laboratory (AFRL). Prior to joining AFRL in 2006, he was a senior mechanical engineer at General Dynamics Corporation. He holds an M.S. degree in Flight Dynamics from Beijing University of Aeronautics and Astronautics, an M.S. degree in Mechanical Engineering from Wright State University (WSU), and a Ph.D. degree in computer science from WSU. His research interests include 3D shape analysis and human action recognition, multi-domain data integration and analysis, and modeling and simulation of human biodynamics.

Physics-Based, Multi-Modal Synthetic Human Image Generator

Devendra Tolani, Mun Wai Lee, Eric van Doorn, Daniel Ashley, Gang Mei, Roger Xu
Intelligent Automation Inc.
Rockville, Maryland
{dtolani, mlee, evandoorn, dashley, mgang, hgxu}@i-a-i.com

Zhiqing Cheng
Innovision, LLC
Dayton, Ohio
zcheng@innovisionllc.com

John Kerekes, Nina Raqueno
Rochester Institute of Technology
Rochester, New York
{kerekes, nina}@cis.rit.edu

Huaining Cheng
Air Force Research Laboratory
Wright-Patterson AFB, Ohio
huaining.cheng@us.af.mil

1. INTRODUCTION

The advancement of deep learning depends on the availability of a large quantity of training data with ground truth labels. Acquiring such datasets is always a challenge due to the high cost of testing and labeling. Moreover, typical data collection efforts are often single modal, single viewing angle, and limited to the environment condition and configuration at the time of data collection. These limitations not only affect the robustness of algorithms on image recognition but also hinders the fusion of multiple modalities for better recognition performance. Therefore, researchers have been looking into creating synthetic imaging data to supplement the need (Schraml, 2019), using 3D graphics simulation technology for the rapid creation of background scenes, buildings, vegetation, machines, and human avatars. However, the majority of current synthetic image generation efforts are neither human-focused nor biofidelic. Studies that do explore synthetic human images (Varol, et al., 2017; Chen, et al., 2016) for machine learning purposes are mostly concentrated in shape-based pose estimation, which does not require high fidelity at the pixel level. They can rely on avatars of graphical rendering without the need to consider human physiology and thermal exchange between the human body, clothes, and environment. On the other hand, the majority of full-motion videos (FMVs) seen in aerial surveillance are from infrared sensors which are more difficult to synthesize due to the complex thermal effects involving a human body, surrounding environment, and weather, etc. This presents an open research and development opportunity to fill in the technical gap.

This paper discusses our effort in developing a novel, integrated, high-fidelity modeling, and simulation (M&S) tool: HumanView (HumV). The architecture of HumV, as shown in Figure 1, is designed for generating multi-modal synthetic human imaging data with real-world environmental effects and sensor optical physics. Its key elements include:

- **HumV editor module**, which allows the user to select an available model, manage the configuration associated with the model, create specific activities for the model. The module includes:
  a) Scenario Model editor**,** which allows the user to define specific scenarios: for example, a human walking in a desert environment on a windy day.
  b) Sensor Model editor**,** which allows the user to define salient sensor properties, such as passive/active nature, modality, bandwidth, and its pose relative to the platform.
  c) Environment editor, which allows the user to annotate the 3D scene model with atmospheric effects and image conditions such as date and time of the day.
  d) Material Properties editor, which allows the user to assign the human model and the 3D scene model with material properties that are salient to the sensor response.
- **HumV models module**, which is a data store that contains 3D human models, 3D geometric models of the environment, and relevant sensor models (EO and IR). The module provides various 3D human models capturing body motion and deformation associated with various human activities. HumV, by using Blender, accepts a wide variety of 3D computer-aided design (CAD) models for simulating the different environments.
- **HumV simulator module**, which allows the user to simulate multiple scenarios to generate the synthetic sensor data for analytics.

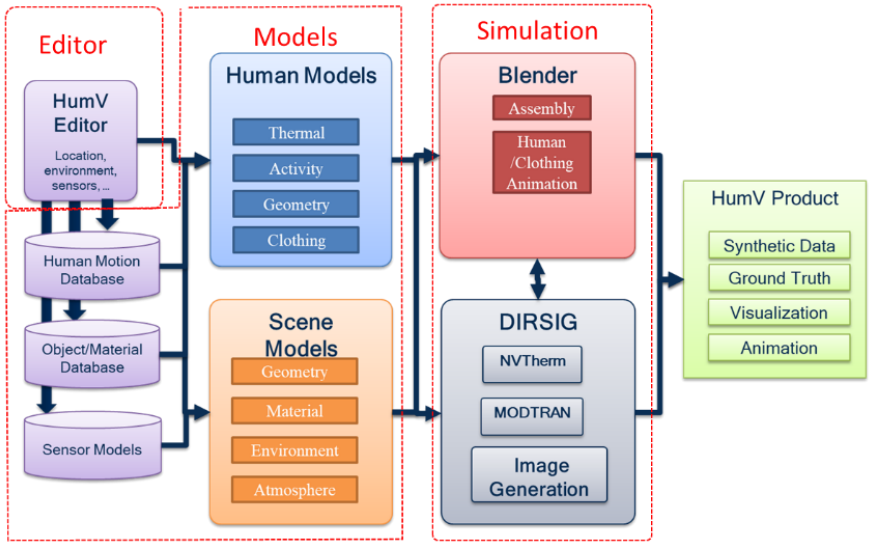


Figure 1. HumV Architecture

Currently, we have developed the core functionality of HumV for evaluating the overall approach and demonstrating basic imaging generation capability. Our main contributions are:

- **Bio-fidelic human and activity models**: HumV can generate bio-fidelic models that realistically represent the diversity of human shape and motion.
- **Physics-based, first principle-based multi-modality simulation**: HumV is designed to perform simulation using physics-based and first principle-based models for radiometry, material reflectance, thermal, sensing and processing. This is to ensure high-fidelity simulation for a diverse set of conditions.
- **Seamless integration of multi-disciplinary M&S tools**: HumV incorporates existing state-of-the-art radiometry M&S tools of Digital Imaging and Remote Sensing Image Generation (DIRSIG) (Goodenough & Brown, 2017) with general M&S tool of Blender 3D (B3D) ("Blender", n.d.) and our new human models.

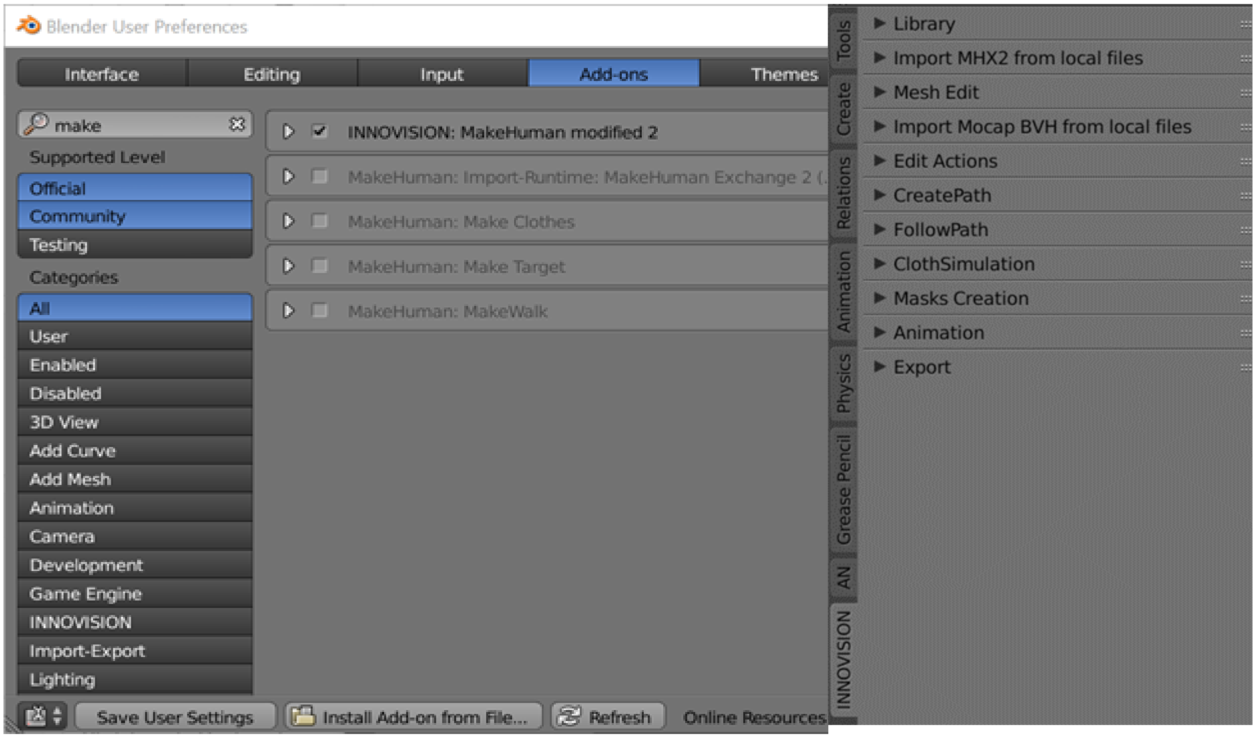
Model validation has been conducted against the experimental data collected using commercial cameras in an outdoor setting. Once fully developed, HumV will be the first-of-its-kind, physics-based, multi-disciplinary solution for M&S of human activities for multi-model sensors. It could support rapid machine learning data generation and algorithm experimentation for FMV analysis, simulating human performance under physical stress and various activities, and mission training and planning.

The rest of this paper is organized as follows. The next section focuses on biomechanical and thermal models of humans, followed by the section discussing the HumV software pipeline and the section presenting validation and simulation results. Conclusions and future directions are given in the last section.

2. MULTI-PHYSICS DIGITAL HUMAN MODELING

2.1 Human Activity Modeling and Simulation

In HumV, human activity modeling and simulation are required to replicate human activities under real-world conditions with high bio-fidelity and to provide human models that meet the requirements of human thermal modeling and multi-modal sensor data generation. To streamline the process of human activity modeling and simulation, we have developed a software – Human Activity Replication Tool (HART) for HumV. As shown in Figure 2 (a), HART is a python-based B3D add-on, which leverages various functions provided by Blender for modeling and animation and utilizes MakeHuman ("MakeHuman", n.d.) to generate human shape models with anthropometric variations of gender, weight, height, etc. The graphical user interface (GUI) of HART shown in Figure 2 (b) has a hierarchical menu-driven structure to help users navigate the human modeling and animation process and utilize various functions to create human activity models. The main functions of HART have also been implemented via python scripts so that the tasks related to human activity modeling can be executed in the backend and seamlessly integrated into the HumV workflow. During HART development, several technical challenges were encountered and overcome, which are discussed as follows.



(a) HART as an add-on to Blender (b) Menus of HART

Figure 2. Human Activity Replication Tool (HART) Overview.

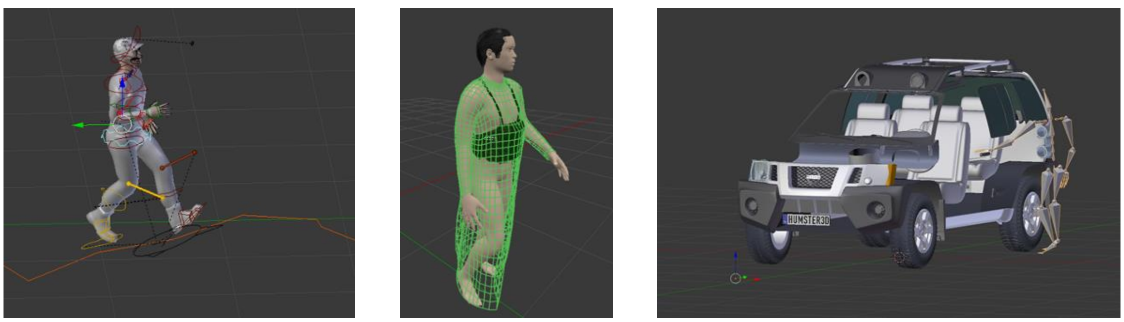
1. *Rigging of a clothed avatar*. For the avatars used in ordinary computer games, cloth and human body surfaces are usually baked together and represented by one layer of mesh. However, for the clothed avatars created by HART, body surface and clothes are represented by two separate layers of mesh, so that the heat emission and transmission from the body surface to clothes can be well accounted for in human thermal modeling. The two layers of mesh need to be rigged well so that a clothed avatar can be readily used for animation. In HART, a rigged and clothed avatar can be generated in three ways:
   - Importing from MakeHuman: If an avatar created by MakeHuman contains clothes, the clothes are automatically rigged.
   - Rigging in Blender: Select the clothes and the avatar to be selected, make sure the skeleton (armature) of the avatar is active, and then using Armature Deform with Automatic Weights to rig the clothes to the skeleton.
   - Using MakeClothes: Create a cloth that fits the template human model provided by MakeClothes, which is an add-on to the Blender. The template human models include baby, teenage, adult male, and female, etc. The cloth created includes three vertex groups: left, mid, and right. After clicking Make Clothes, a user can find the created clothes in the asset of MakeHuman. When an avatar with new clothes is exported from Make Human to Blender, the clothed avatar in Blender has been rigged already. The clothes created using MakeClothes automatically fit different avatars which are generated from the same template human model.

2. *Animating an avatar interacting with the scene*. Two problems are involved. One is the clearance between the feet and the ground. To ensure appropriate contact (no gap and no penetration) between the feet and ground surface, the center of gravity (CG) of an avatar needs to be adjusted according to the change of the altitude of slope or uneven ground. The other problem is the feet sliding on the ground surface. Feet-sliding occurs when an avatar walks on a path with the length of each step not equal to what should have been according to motion capture data. The problem is further complicated when clothing simulation is involved because synchronization is needed between sliding correction and clothing simulation. In HART, both problems are solved by using Animation

Nodes (a Blender add-on) to determine the length of each step and the CG position for each step. An example is shown in Figure 3 (a).

3. *Intersections between the body mesh and cloth mesh*. In HumV, human thermal modeling requires no intersections between the human body surface and clothing. Therefore, HART has built-in per-frame based automatic intersection detection. For minor intersections, they are automatically eliminated in HART based on the ShrinkWrap modifier in Blender which shrinks one surface to another. For Major ones, a user can use Mesh Deform modifier with shape keys in Blender to manually eliminate them or convert them into minor intersections which are then automatically eliminated.

4. *Dismount modeling*. HART has modeled a few dismounts with apparel (magazine vest) and clothes (loss robe or Airman Battle Uniform (ABU). To model a clothed avatar with a magazine vest hidden under the clothes, the vest is wrapped around the human body with B3D's ShrinkWrap modifier to ensure no intersection during the animation, as shown in Figure 3 (b). It is treated as part of the human body with collision physics with respect to the body during clothing simulation, instead of as another layer of clothing under the clothes. Fine mesh is used for regions where clothes are overlaid (e.g., neck and groin) to reduce self-collision. A more challenging case modeled is a person entering and exiting an SUV, as shown in Figure 3 (c), where the hands griping steering wheel and the feet stepping in/out cabin create inconsistencies that are resolved manually.



(a) Walking on a slope (b) A hidden vest (c) Entering into a vehicle

Figure 3. A Few Modeling Cases in HART

We note that HART replicates human motion rather than synthesizing human motion. The change of human body shape and clothing during motion is obtained through rigging, without accounting for non-rigid deformation, thus the fidelity is limited. The evaluation of model rigging error associated with MakeHuman/Blender/MakeClothes was not done in this study, since the data that can be used for the validation were not available. The framework is flexible in ingesting motion data in standard formats (e.g. bvh) to render diverse range of activities and interactions.

2.2 Human Thermal Modeling

The goal of the integrated human thermal simulation component is to take into account the user input of various models including scene, clothing, and activity, and produce an output consisting of temperature estimates of all the facets of skin and cloth of the human in the scene. This integrated thermal model has an operation structure depicted in Figure 4, which involves two main thermal subcomponents: the Skin Temperature model and the Body/Clothing Interaction model. The following subsections summarize various calculations related to them.

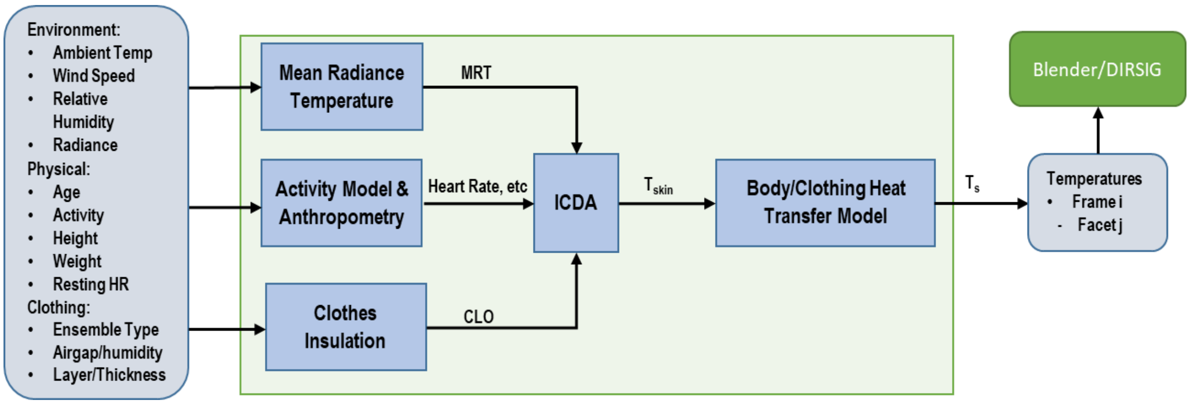


Figure 4. Integrated Human Thermal Modeling

2.2.1 Initial Capability Decision Aid (ICDA) Model for Skin Temperature

ICDA (Yokota & Berglund, 2006) developed by the US Army Research Institute of Environmental Medicine (USARIEM), is the central component for modeling heat production dynamics in the human body. At a high level, ICDA is a heat transfer model that takes environment inputs, human physical properties, and human activity properties and outputs a core body temperature and a skin surface temperature (Figure 5). For the HumV human thermal model, we are primarily interested in the skin surface temperature, T_{skin}.

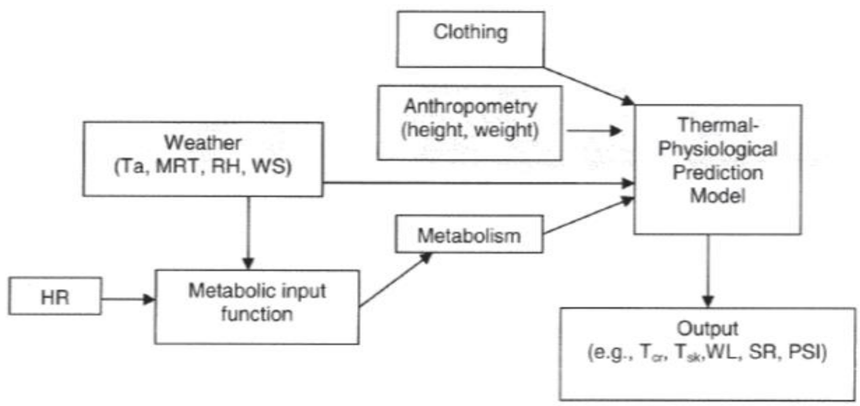


Figure 5. Illustration of ICDA Model

First, the human activity model is used to produce a heart rate required by ICDA. This is accomplished through the use of a lookup table that maps activity to a power output during aerobic exercise and an equation that relates power output to heart rate for a given individual at a certain level of fitness (Arts & Kuipers, 1994). The other important environmental inputs to the ICDA model are the ambient temperature T_a, the mean radiant temperature MRT, the relative humidity RH, and the wind speed WS. All of these environmental conditions are defined or derived from user options with the exception of MRT. MRT is found using the Stefan-Boltzmann law,

$$E = \sigma(MRT)^4 \qquad (1)$$

where the energy E is the total amount of incoming radiation acting on the human model and σ is the Stefan-Boltzmann constant.

2.2.2 Incoming Radiation

Incident radiation is a major contributor to the thermal environment of our human model. The key components of the radiation tracing portion of the thermal model are the output from a spherical collection module that collects incident radiation surrounding an object. It is implemented in DIRSIG 5 as the spherical collector through which one defines a point in 3D DIRSIG scene and DIRSIG then calculates all of the incoming radiation that reaches that point in 3D space. The radiation is calculated over a set of user-defined wavelengths. For each wavelength, the spherical collector returns an orthogonal image (spherical collector image), hence the output of the spherical collector is a group of spherical images. Each pixel of a spherical collector image maps to the direction of the incoming radiation where the horizontal of the image spans the Azimuth, ϕ, and the vertical of the image spans the Zenith, θ, of the incoming radiation. Each pixel subtends a solid angle defined by the integral solid angle definition:

$$\Omega = \int_{\theta_0}^{\theta_n} \int_{\phi_0}^{\phi_n} \sin(\theta)\, d\theta d\phi = (\phi_n - \phi_0)(\cos(\theta_0) - \cos(\theta_n)) \tag{2}$$

For HumV, we configure the spherical collector to compute multiple bands of radiation in the visible and IR bands with a pixel resolution of 360x360. We developed the radiation ray tracing with respect to the facets (polygons) of the human model using a far-field approximation. It assumes all incoming radiation as defined by our spherical collector is originating from a point in space that is effectively infinitely far from our human model.

We are interested in calculating the total amount of radiation, q_{rad_i}, acting on the *i*-th facet of the human model as defined in Equation 3:

$$q_{rad_i} = \sum_{j=1}^{J} \eta_s \cos(A_j)\, I_j \Omega_j 1_{incident_j} \tag{3}$$

The total amount of incident radiation is the summation over the effect of all incoming rays of radiation where J is the total number of incoming rays. The effect of an incoming ray of radiation with an intensity I_j is determined by the value of the intensity, the coefficient of absorption of the material η_s, the cosine of the angle of incidence A_j, the solid angle of the incoming ray Ω_j, and whether the incoming ray reaches the facet $1_{incident_j}$. The intensity, I_j, of the incoming ray of radiation is defined as the sum over the pixel values, p_b, of all of the spherical collector images, B, multiplied by the width of each bandwidth, w_b.

$$I_j = \sum_{b}^{B} w_b p_b \tag{4}$$

The angle of incidence, A_j, of an incoming ray is the angle between the normal vector of the facet and the vector defining the incoming ray. The check function, $1_{j,i}$, for a face i takes value 1 if the ray j is incident on face i. This value is determined by the output of the radiation ray-tracing procedure.

2.2.3 Clothing Heat Transfer

The clothing surface temperature T_s is solved through an equation of energy balance shown in Equation 5.

$$q_{in} + q_{rad_i} = q_{conv} + q_{emission} \tag{5}$$

Besides the aforementioned q_{rad_i}, q_{in} models the conductive heat transfer from the skin surface to the clothing surface in which the air gap between the skin and clothes is outputted from the HART avatar. q_{conv} is the heat transfer from the clothing surface to the environment due to the movement of the surrounding air. $q_{emission}$ is the radiative release of energy from the clothing surface to the environment.

The conduction term, q_{in}, is calculated as follows by modeling the airgap and clothes as thermal resistors with the resistance of R_1 and R_2, respectively,

$$q_{in} = \frac{1}{(R_1 + R_2)} (T_{skin} - T_s) \tag{6}$$

If we define k_a and k_b as the resistivity and l_a and l_b as the thickness of the air gap and clothes, respectively. We have,

$$R_1 = \left(\frac{k_a}{l_a} + 4\sigma\left(\frac{T_a + T_{skin}}{2}\right)^3\right)^{-1} \quad \text{and} \quad R_2 = \frac{l_b}{k_b} \tag{7}$$

where σ is the Boltzmann constant and T_a is the ambient temperature. Using Stefan-Boltzmann law, clothing surface emission can be calculated as,

$$q_{emission} = \varepsilon\sigma T_s^4 \tag{8}$$

where ε is a material property coefficient for radiative heat transfer. Finally, the convection term can be calculated as,

$$q_{conv} = h(T_s - T_a) \tag{9}$$

where the convective heat transfer coefficient h is quantified using a standard engineering approximation of $h = 10.45 \ - V_a + 10V_a^{\frac{1}{2}}$ in which V_a is the wind speed the surrounding environment.

2.3 Environmental and Sensor Models

HumV's environmental and sensor models come directly from DIRSIG, which have been matured and validated through many years of development. HumV utilizes DIRSIG's scene construction, sensor, and material modules extensively. DIRSIG associated material and environmental property databases are incorporated into HumV.

DIRSIG is designed to generate passive broadband, multi-spectral, hyperspectral low light, polarized, active laser radar, and synthetic aperture radar datasets through the integration of a suite of first-principles based radiation propagation modules. These object-oriented modules address tasks ranging from bi-directional reflectance distribution function (BRDF) (Kerekes & Baum, 2003) predictions of a surface to time and material-dependent surface temperature predictions and the dynamic viewing geometry of scanning imaging instruments on the agile ground, airborne, and space-based platforms. DIRSIG also integrates MODTRAN ("MODTRAN", n.d.), the widely used atmospheric radiation propagation model. MODTRAN models the atmosphere as stratified (horizontally homogeneous) and 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, which is needed by the aforementioned spherical collector.

3. INTEGRATED HUMV SOFTWARE DESIGN AND USER INTERFACE

Our goal is to release an end-to-end prototype solution demonstrating the whole simulation pipeline from user input to simulated image output. Due to the multidisciplinary need in generating synthetic images and video, we have a unique challenge of seamless integration of algorithms developed in this effort such as HART, human thermal model, and DIRSIG's newly-developed spherical collector of environmental radiation with third-party M&S tools such as B3D. Therefore, we adapted a modular approach (Figure 1) under the Docker framework ("Docker", n.d.) in which users configure human and human thermal models via the HumV editor. The outcomes from these two models are ingested by HumV automatically into the DIRSIG scene to conduct a physics-based simulation of specific sensors defined by the DIRSIG sensor model. The corresponding synthetic images and pixel-based ground-truth outputs from DIRSIG are the final products. This modular software architecture further facilitates future functionality upgrades and software updates by individual modules in a spiral manner.

Each module within the HumV software is encapsulated by a Docker container and completely transparent to the user, except for the HumV-GUI. The HumV-GUI is the only module in the system where it is necessary for the user to provide input to the system. Using multiple Docker containers, we parallelize HumV by running a Docker container for each frame of the simulation independently. However, because the thermal dynamics module in the DIRSIG hasn't been parallelized, the Docker containers used in the simulation pipeline are managed by a master python script to create and assign tasks to the Docker containers with the limitation of one Docker container per Central Processing

Unit (CPU) core. Therefore, the current HumV could only be speeded up with multi CPUs instead of GPUs. We expect the process will be fully GPU parallelized later when the thermal dynamics of DIRSIG is upgraded.

In HumV, the entire M&S workflow is defined through the graphical user interface (GUI) – HumV Editor shown in Figure 6. It translates the users' parameter inputs into simulation settings via six tabbed pages so as not to overwhelm the user with a single large setup page. The tabs are: (i) General; (ii) Human/Clothing/Motion; (iii) Human Placement; (iv) DIRSIG Scene; (v) Sensor; and (vi) Thermal. Each of the inputs to the HumV Editor corresponds either directly or in some cases indirectly to a field in the configuration file that is used by the modules in the HumV pipeline backend.

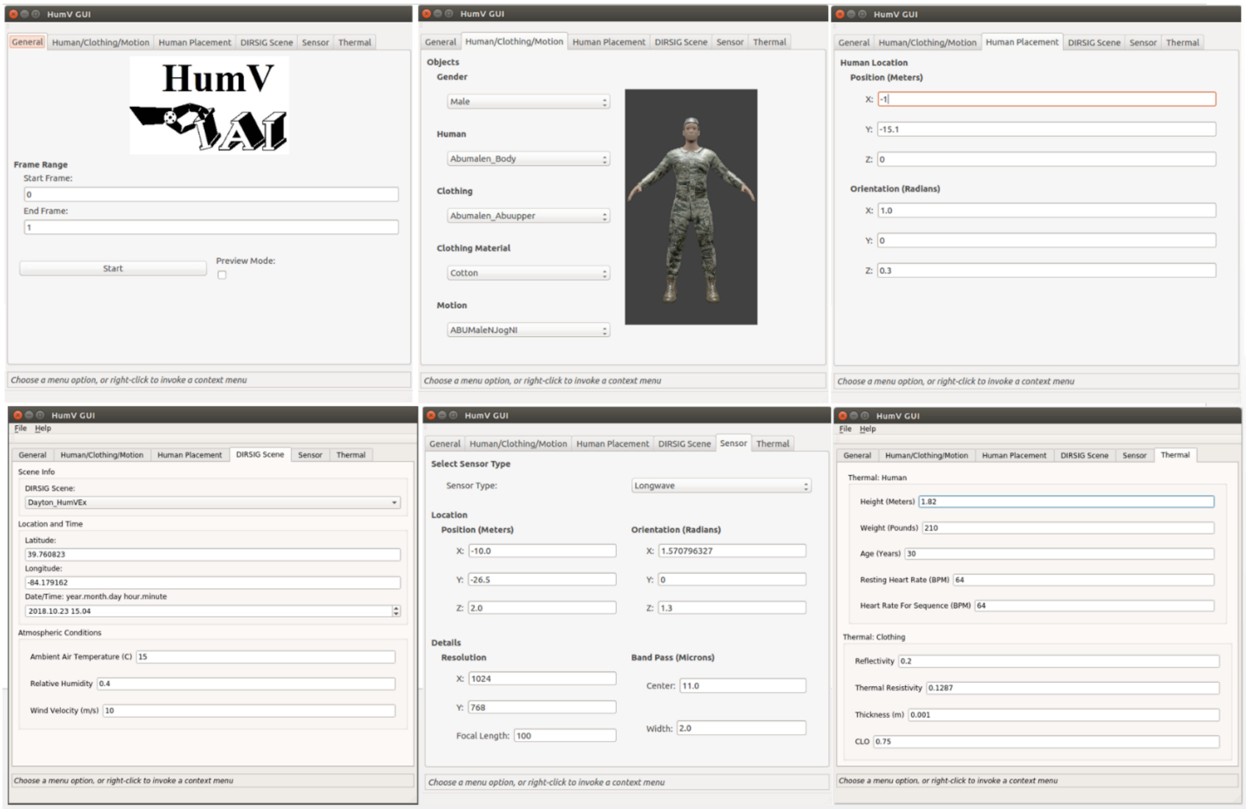


Figure 6. HumV Editor GUI for M&S User Inputs

4. SIMULATION RESULTS AND TEST VALIDATION

4.1 Human view validation experiment (HUMVVEX)

A key aspect of the Human View project was the planning, execution, and data analysis of a validation experiment. The experiment was conducted on October 23-24, 2018, on the property of the National Center for Medical Readiness (NCMR) located at 506 E. Xenia Dr., Fairborn, Ohio. Figure 7 shows a bird's eye view of the site. Excellent on-site support was provided by the TechWarrior Enterprise and Wright State Research Institute. Cameras were placed on the concrete pad with a view toward the warehouse shown in the upper right of the image. The humans and other objects were placed just to the right of the warehouse door.

Figure 8 provides a picture of the particular area used for the experiment along with an example color image acquired by one of the sensors. The left is a context picture showing one of the imaging sensors with the operator, the tables supporting calibration blackbodies, and the area of the warehouse used as background. The right is an example color image acquired by one of the imaging sensors used in the experiment showing a typical field of view. Validation imagery and data were acquired for many combinations of scenario parameters. Table 1 lists the parameters and their options.

Figure 7. NCMR Site Used for the HUMVVEX Data Collection

Figure 8. Left: A Picture of the Area Used for the Experiment Right: An Example Photo

Table 1. Scenario Parameter Options

Parameter	Options
Illumination	Full sun or in shadow
Clothing	Bare upper body with shorts, white or black shirt with shorts or leggings, military uniform, or robe
Activity	Standing/turning, walking, or post-exercise
Weapon	Present or not present
Gender	Male or female

A total of 80 scenarios were collected. Both imagery and ancillary data were acquired during the experiment. Imaging systems used during HUMVVEX included electro-optical (EO) visible color, mid-wave infrared (MWIR), and two longwave infrared (LWIR) systems. Other non-imagery data were also acquired during HUMVVEX. These included field spectral reflectance and emissivity, downwelling irradiance, and temperatures of the calibration blackbodies and other objects in the scene.

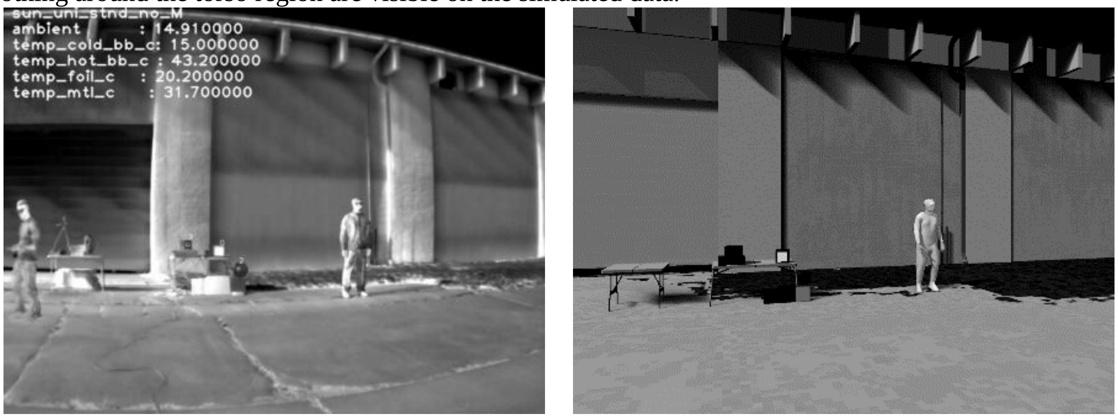
4.2 Validation of Image Intensity on Person

To validate the HumV software, we developed a nominal DIRSIG background scene containing geometry and material properties for the ground surface and warehouse background. The support tables and calibration blackbodies were also included in the scene. Figure 9 shows a longwave infrared rendering of the scene. This image is a rendering of the Atom 1024 LWIR camera for 14:05 on 24 October 2018.

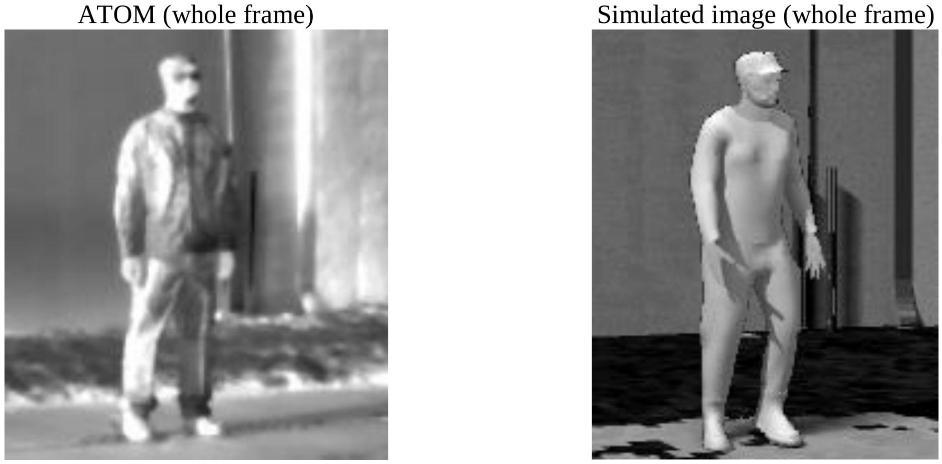
Figure 9. Simulated LWIR Image of the Scene for the HUMVVEX Validation Experiment

Geometry for the scene was constructed using available aerial photos and measurements acquired during HUMVVEX. The warehouse walls and the ground surface were attributed using spectral reflectance and emissivity data collected during HUMVVEX. Other materials were attributed to data from the DIRSIG library based on best estimates. Thermal properties for the objects were estimated from values obtained in the literature and adjusted to provide a reasonable match to the observed characteristics in the data.

We examine the image intensity of the person in the simulated image in detail and compare it with the real image. The example is shown in Figure 10. It shows that the intensity values are generally correct. Effects such as shadows and loose clothing around the torso region are visible on the simulated data.



ATOM (whole frame) Simulated image (whole frame)



ATOM (zoom-in) Simulated image (zoom-in)

Figure 10. Simulated and Real Image Used in the Validation

To obtain a quantitative analysis result, we segment the human image area into different body regions, as shown in Figure 11. This allows us to perform analysis of individual regions.

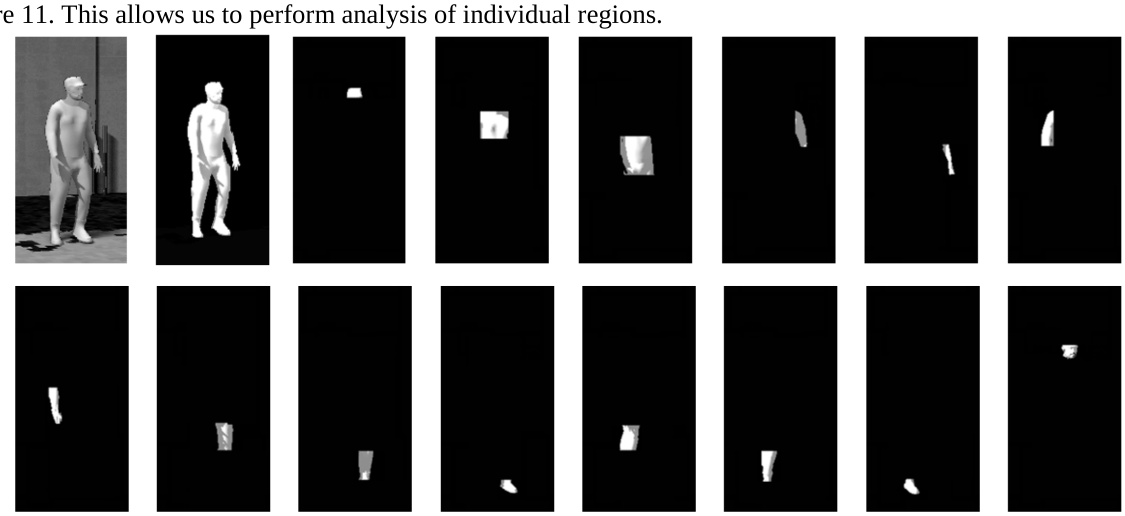


Figure 11. Masks Region for Different Body Areas

Using the segmentation masks for the different body regions, we sample the image and compute the average intensity values on the HumV simulation image and the real Atom image. The result is shown in Table 2.

Table 2. Comparison of Average Image Intensity Per Body Region

#		HumV Simulation	Atom	Difference
1	Full body	132.0	160.3	-12.1
2	Hat	187.3	204.3	-10.8
3	Chest	148.9	163.5	-5.7
4	Torso	111.3	138.7	-4.9
5	Left upper arm	78.2	101.2	-17.3
6	Left lower arm	154.0	116.7	-28.5
7	Right upper arm	173.0	180.9	23.1
8	Right lower arm	177.0	164.0	-11.9
9	Left upper leg	87.8	138.1	-45.2
10	Left lower leg	83.7	120.3	-34.2
11	Left foot	182.0	178.0	-13.1
12	Right upper leg	151.5	189.7	6.7
13	Right lower leg	145.0	183.5	10.5
14	Right foot	179.2	181.1	30.5
15	Face	128.0	147.4	54.8

The result shows that the intensity values of the different body regions and their ordering are comparable. Some differences are due to slight differences in the body pose. For example, the lower arms on the simulation are slightly raised, causing a slight difference in shadow and sun loading on the arms and torso.

Figure 12. Additional Examples for Comparing the Simulated Images with Real Thermal Images.

5. CONCLUSIONS

The ability to automatically generate synchronized, co-registered multi-modal sensor data is a critical need for many DoD applications such as sensor fusion, sensor exploitation, signature analysis, and counter-surveillance. Driven by the requirement of our customer, the AFRL, we designed and developed the HumV software system, providing an end-to-end software processing pipeline to perform human activity simulation. Besides the software development, we also prepared various data sources needed for human activity simulation, completed a validation data collection for Human View Validation Experiment at the National Center for Medical Readiness (NCMR), and developed a DIRSIG scene of the HUMVVEX data collection scenarios. This scene is then used in a full HumV simulation for verification of the simulation results against the collected real data.

In summary, our HumV software successfully demonstrated a new capability of simulating thermal infrared imagery of close-up scenes with humans performing activities. While additional work remains to fully validate and further develop this capability, initial results show the method and integrated M&S framework are promising in generating multi-modal synthetic human imaging data with real-world environmental effects and sensor optical physics.

HumV has broad potential applications. The tool is being studied for use in sensor feasibility analysis for ISR mission planning at the Air Operation Center (AOC). The technology would also be useful in other areas such as analytical technology development for homeland security applications. HumV would be beneficial to developers of first responder support equipment for operations such as search and rescue. Commercial applications include the fast-growing field of robotics. HumV may be used by developers of self-driving cars and any other Artificial Intelligence or robotic systems that interact with humans.

ACKNOWLEDGEMENTS

This work is partially supported by the Small Business Innovation Research (SBIR) contract FA865017C6875. Intelligent Automation Inc. is the lead developer. Innovision LLC is responsible for human anthropometric and activity modeling. Rochester Institute of Technology contributes to the environment and sensor modeling as well as model validation test.

REFERENCES

Schraml D. (2019). Physically Based Synthetic Image Generation for Machine Learning: A Review of Pertinent Literature. *Proceedings of SPIE 11144, Photonics and Education in Measurement Science*.

Varol G., Romero J., Martin X., Mahmood N., Black M.J., Laptev, I., & Schmid C. (2017) Learning from Synthetic Humans. *Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 109-117.

Chen W., Wang H., Li Y., Su H., Wang Z., Tu C., Lischinski D., Cohen-Or D., & Chen B. (2016) Synthesizing Training Images for Boosting Human 3D Pose Estimation. *Proceedings of 2016 Fourth International Conference on 3D Vision (3DV)*, pp. 479-488.

Goodenough A.A. & Brown S.D. (2017) DIRSIG5: Next-Generation Remote Sensing Data and Image Simulation Framework. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 10, no. 11*, pp. 4818-4833.

Blender. https://www.blender.org

MakeHuman. http://www.makehumancommunity.org/

Yokota M. & Berglund L.G. (2006) Initial Capability Decision Aid (ICDA) Thermal Prediction Model and its Validation. *USARIEM Tech Report T06-03, U.S. Army Research Institute of Environmental Medicine, Natick, MA 01760*.

Arts F.J. & Kuipers H. (1994) The Relation between Power Output, Oxygen Uptake and Heart Rate in Male Athletes. *International Journal of Sports Medicine. vol. 15(5)*, pp. 228-31.

Kerekes J.P. & Baum J.E., (2003) Hyperspectral Imaging System Modeling. *MIT Lincoln Laboratory Journal, vol. 14(1)*, pp. 117-130, 2003.

MODTRAN. http://modtran.spectral.com/

Docker. https://www.docker.com