Virtualizing Humans for Game Ready Avatars

Jay Saffold, Tovar Shoaf, Jason Holutiak
Research Network, Inc
Kennesaw, GA
jsaffold@resrchnet.com,
tshoaf@resrchnet.com,
jholutiak@resrchnet.com

Timothy Roberts, Pat Garrity
U.S. Army Research Laboratory-Human Research and Engineering Directorate, Simulation and Training Technology Center (ARL-HRED STTC)
Orlando, FL
timothy.e.roberts50.civ@mail.mil,
patrick.j.garrity4.civ@mail.mil

ABSTRACT

The U.S. Army Research Laboratory-Human Research and Engineering Directorate, Simulation and Training Technology Center (ARL-HRED STTC) performs research and development in the field of creating realistic, individualized virtual avatars from live subjects that retain the physical characteristics and appearance of the subject including height, weight, skeletal dimensions, body morphology and facial/body appearance. While photogrammetric extraction technologies are maturing there are a number of additional steps which must be performed to "virtualize" live humans into game ready avatars. Game ready in this context means the mesh stretches properly with motion, there are sufficient level-of-detail options, and the number of polygons is optimized for computer rendering in real-time on commercial graphics adapters and central processing units. Photogrammetric algorithms which extract mesh information from 3D subjects also do not inherently include the underlying bone structure (rigging) required for avatars to move in virtual environments. A novel integrated system approach developed leverages low-cost data capture systems and targets automation of all the steps necessary to go from live human to a high-fidelity game-ready avatar. This paper discusses the different trade spaces associated with various photogrammetric techniques/algorithms, commercial software packages, data capture approaches, subject lighting, frame occupancy, motion during data collection impacts, and converting what is originally a very dense mesh through "retopologization" into optimized levels-of-detail which are properly weighted to a virtual bone system. Each step in the process is discussed along with approaches for automation and the associated trade spaces which affect the quality of the outcome.

KEYWORDS

Dismounted Soldier, Virtualization, Avatar, Game, Virtual Environment, Immersion

ABOUT THE AUTHORS

Jay Saffold is the Chief Scientist for RNI and has over 32 years engineering experience in both military and industry research in game-based training, immersive systems, RF tags, virtual reality, digital databases, soldier tracking systems, millimeter wavelength (MMW) radar, multimode (MMW and optical) sensor fusion, fire-control radar, electronic warfare, survivability, signal processing, and strategic defense architecture. Mr Saffold is the lead developer of the Game Distributed Interactive System (GDIS) and Man Machine Interfaces for Simulation and Training (MMIST) system. He lectures annually for GTRI on remote sensing and signal processing. He holds a B.S.E.E. degree from Auburn University.

Tovar Shoaf is the Lead Software Engineer for RNI with 4 years of experience in the military simulation industry. He holds a BS in Game Development from Full Sail University (2011) which focuses on efficient, real-time input and processing. Prior to working at RNI he worked on flight reconstruction and analysis from both live and simulated flights. Since joining RNI, his main focus has been implementation and modification of software for man-worn tracking systems that fuse information from a range of sensors into stable position and pose data. He has also developed game engine interfaces, game-based trainers, mobile applications, and multiplayer servers. Mr. Shoaf's current interests include real-time position and pose tracking, immersive multiplayer simulation, simulation data recording and analysis, and software/hardware interface integration.

Jason Holutiak is the Lead 3D Artist for RNI and has over 5 years' experience in both military and industry research in game-based training, immersive systems, and 3D content creation for serious game applications. He holds an A.A.S in Multimedia and a B.S. in Computer Animation which focused on the design and implementation of 3D assets for games and film. During his time with RNI, Mr. Holutiak's main focus has been creating 3D virtual environments and assets based on real locations for Live and Virtual bridging demonstrations. His current focus has been the research and implementation of an automated virtualization process that creates game avatars from live humans.

Timothy Roberts is a Science and Technology Manager of Dismounted Soldier Simulation and Training Technologies at the U.S. Army Research Laboratory-Human Research and Engineering Directorate, Simulation and Training Technology Center (ARL-HRED STTC). He currently works in Dismounted Soldier Training Technologies conducting research and development in the area of Dismounted Soldier training and simulation. His research interests include dismounted soldier training environments, games for training, virtual locomotion, multi-modal interfaces, path planning, control systems and robotics. He received a B.S. in Electrical Engineering and a M.S.E.E. in Control Systems and Robotics from the University of Central Florida.

Pat Garrity is a Chief Engineer at the U.S. Army Research Laboratory-Human Research and Engineering Directorate, Simulation and Training Technology Center (ARL-HRED STTC). He currently works in Dismounted Soldier Simulation Technologies conducting research and development in the area of dismounted soldier training and simulation where he was the Army's Science and Technology Manager for the Embedded Training for Dismounted Soldiers program. His current interests include Human-In-The-Loop (HITL) networked simulators, virtual and augmented reality, and immersive dismounted training applications. He earned his B.S. in Computer Engineering from the University of South Florida in 1985 and his M.S. in Simulation Systems from the University of Central Florida in 1994.

Virtualizing Humans for Game Ready Avatars

Jay Saffold, Tovar Shoaf, Jason Holutiak
Research Network, Inc
Kennesaw, GA
jsaffold@resrchnet.com,
tshoaf@resrchnet.com,
jholutiak@resrchnet.com

Timothy Roberts, Pat Garrity
U.S. Army Research Laboratory-Human Research and Engineering Directorate, Simulation and Training Technology
Center (ARL-HRED STTC)
Orlando, FL
timothy.e.roberts50.civ@mail.mil,
patrick.j.garrity4.civ@mail.mil

INTRODUCTION

The Army spends an extraordinary amount of time and resources in developing virtual environments that geographically and geotypically represent the live area that is being virtualized. Many resources are put into making the virtual environment exactly match the live environment used for training. The exact topology of the environment is carefully virtualized. Building textures, materials and their properties as well as environmental models of trees, landscaping, street signs, etc… are all well matched in the virtual environment for mission training and mission rehearsal. What has not received much attention in the past is the actual virtualization of the live Soldier that will be training in the environment. The typical avatar is a base model, usually pre-made based on a picture of a Soldier and is pre-canned in most cases. The actual Soldier's appearance is an important feature while co-existing in a virtual environment (Banakou & Chorianopoulos 2010). The human avatar should potentially look like the actual Soldier involved in the training exercise and should retain the Soldier's height, weight, gender, race, and physical appearance. Many times while in a mission scenario, trust is built around those you train with and are used to and feel more comfortable when training with your "battle buddies". While in a virtual training exercise, you should retain that trust in your teammates and seeing their avatar that looks just like them may enable this trust and confidence to be established while training in the live, virtual or mixed environments (Bente, Ruggenberg & Kramer 2004).

BACKGROUND

The U.S. Army Research Laboratory has been researching and developing a new concept for virtualizing humans in a cost effective manner. The overall idea is that one day a Department of Motor Vehicles (DMV) like concept system could be used where the subject leaves the virtualization area with a complete avatar on a removable media. A Solider could simply come into an office such as the one that prints the Common Access Cards, be scanned and a few hours later receive his/her avatar on the Soldier's identification card. The avatar would realistically represent the Soldier's height, weight, gender, race and appearance and could be updated on an annual, bi-annual or semi-annual basis to closely keep the live Soldier and his/her avatar as close as possible over time. This will allow fellow Soldiers to train together in the live, virtual and mixed environments. The current system being developed at the U.S. ARL-HRED STTC will also allow any avatar that is created to work in multiple game engines with very little backend. The avatars are currently created in the Unity and VBS2 game environments, but could be easily transitioned to the Unreal game engine or any other "game of the day". This allows a Soldier to only be scanned once per time period and be allowed to use his/her avatar in any game environment among multiple game engines. The virtualizing humans concept described in this paper also allows other Army researchers to start to build intrinsic values that are tied to the avatar. Speed, gait, marksmanship, physical fitness, etc…are all future values that could be added to the newly improved Soldier avatar so that now the avatar not only looks like its live counterpart, but also acts (moves, shoots and communicates) like its human counterpart. Obviously an unfit Soldier will take longer to arrive to the target and have decreased marksmanship skill as he/she is tired once arriving at the target than a fit Soldier. In most current virtual environments and game engines, most avatars all move at the same speed and can shoot at the same level no matter what distance they have run or moved. The current avatars are often referred to as "hero" avatars, meaning that they can run forever, never get tired, and always shoot and communicate with great accuracy. This is not realistic as no live human could possibly perform these feats as live humans get fatigued and

have muscles that get tired. This paper describes the virtualizing of humans so that the avatar appears as their live counterpart (weight, height, race, gender) and can eventually start performing the same as their live counterpart. This will allow avatars in the future to look and act exactly as their live human counterpart.

GAME READY AVATAR AND STANDARDS

Before discussing the techniques and technologies used to virtualize humans, it is important to understand the characteristics of a game ready avatar. In this context, "game ready" means the finalized model has the correct rig (bone definitions), the mesh created from virtualization processes is appropriately vertex weighted to the rig, and the textures are correctly UV mapped to the mesh. It is also important that the avatar have a base set of animations available for testing the mesh vertex weights and allowing for quick identification of possible "poses" which cause too much stretching once vertex weights are applied.

DATA CAPTURE SYSTEM

In the project, very low cost implementations of the data collection system was focused on as an alternative to a number of existing systems ranging from elaborate light stages to do-it-yourself (DIY) systems to single camera's that are manually walked around an object (Debevec, 2012)(Garsthagen, 2013)(PhotoModeler, 2015). Single camera approaches are evolving where the data is collected by simply walking around the object however they have not found their way into detailed avatar modeling yet because they do not produce fine enough detail on the key body components. Single camera approaches also may suffer from non-repeatability which was not conducive to the research goals.

As initial criteria, a target of under $5,000 was set for the hardware and software components and the system needs to be lightweight allowing easy transport and setup at alternate facilities. After review of a number of different approaches a turntable-based system was selected since it would limit the number of camera's required to implement a full hemisphere of photographs, still provide a data collection system with some level of rigidity, and allow for migration to automated techniques with repeatability. The turntable approach moves the subject while the cameras are fixed. The selected cameras were Canon DSLRS (Eos Rebel T3iEF-S). These were selected due to the availability of a software development kit (SDK) and the ability to capture high resolution photographs which were suitable for implementing trade studies related to image quality (up to 18 megapixel). The final implementation in this phase of the effort consisted of three Canon DSLR cameras successively mounted on a single heavy duty stand and positioned with vertical diversity to allow (a) capture of the full body height and width over all rotations, (b) provide visibility to "under and over" areas like the top of the head and under the nose, and (c) repositioning to focus solely on the head area – the dominant avatar feature - for high frame occupancy. The key components of the low cost data capture system were:

- **Turntable (ARQ-24-110)** – used to slowly rotate the subject in front of the fixed camera system. The rotation rate was about 12 deg/sec
- **Canon DSLR Digital Cameras (Eos Rebel T3iEF-S)** – the primary photo collection apparatus
- **Heavy Duty Tripod (12' - 806D)** – rugged fixed mount with vertical and orientation diversity to mount the digital cameras
- **Camera Adapters (CX-3000)** – rugged adapter clamps to mount the cameras to the main trunk of the 12' stand with each lock / unlock mechanisms
- **Diffuse Lighting w/Stands (FAN102kit)** – repositionable light sources to remove dominant shadows on the subject at camera system orientation
- **Cloth Background (AB50310X20)** – a uniform background with a non-creasing fabric used to provide high contrast between the subject and the background in the direction of the fixed camera rig. An "orange" and "white" color depending on the subject's clothing and skin color was chosen.
- **Backdrop Support Stand System (5400)** – a 12' high by 15' wide adjustable mount to hang the cloth background from behind the subject

- **Heavy Duty Muslin Clamps (45HDC6)** – rugged clamps to attach the cloth background to the horizontal and vertical supports of the Backdrop Support Stand System

The total cost for the data capture system (including software) was $3,084.36 which was well below the target.

PHOTOGRAMMETRY AND TRADES

Photogrammetry is the science of obtaining reliable information about the properties of surfaces and objects without physical contact with the objects, and of measuring and interpreting this information (Schenk, 2005). Photogrammetry uses methods from many disciplines, including optics and projective geometry.

The 3D coordinate system defines the locations of object points in the 3D space. The image coordinates define the locations of the object points' images on the film or an electronic imaging device. The exterior orientation of a camera defines its location in space and its view direction. The inner orientation defines the geometric parameters of the imaging process. This is primarily the focal length of the lens, but can also include the description of lens distortions. Further additional observations play an important role: With scale bars, basically a known distance of two points in space, or known fix points, the connection to the basic measuring units is created. In the virtualization application, *Stereophotogrammetry* techniques are used. *Stereophotogrammetry* involves estimating the three-dimensional coordinates of points on an object. These 3D points are determined by measurements made in two or more photographic images taken from different positions (similar to stereoscopy). Common points are identified on each image. A line of sight (or ray) can be constructed from the camera location to the point on the object. It is the intersection of these rays (triangulation) that determines the three-dimensional location of the point. The point cloud is processed (outlier's removed) and triangles are formed from adjacent points. These triangles constitute a "mesh" that represents a 3D version of the subject object. *Stereophotogrammetry* is emerging as a robust non-contacting measurement technique to determine dynamic characteristics and model shapes of non-rotating (Baqersad, 2012) and rotating structures (Lundstrom, 2012).

With modern computing and digital imaging resources, there has been an explosion of Photogrammetric software applicable to human virtualization. While most photogrammetric software uses similar algorithms to track image pixels and produce 3D point clouds, there is a large variation of approaches to triangulate and optimize these clouds to mesh surfaces (Remondin, 2010). For the virtualization effort, three different applications were reviewed for the extraction of the 3D mesh (and textures) using photogrammetric processing. Based on this review, the Agisoft PhotoScan (Agisoft, 2013) software was chosen as the best compromise between costs and localized computing requirements.

Trade Studies

The quality, characteristics, and geometry of the photo's collected for photogrammetric extraction is a large part of the virtualization process and will impact the mesh and texture sets produced. The photogrammetric software implementation was evaluated based on seven (7) potential error sources associated with data collection activity. These trade studies related to:

- Subject Motion During Data Collection
- Number of Photos / Angular Intervals
- Camera Horizontal and Vertical Diversity
- Background / Type
- Lighting Quality / Shadow Removal and Motion
- Number of Pixels (Image Size)
- Subject Frame Occupancy

The following sections discuss these trade spaces and the results using the photogrammetric software and the low cost data capture system.

Subject Motion During Data Collection - Due to the nature of the data capture system implementation (rotating subject), the potential for the subject to move during a scan is high. Subject motion can cause distortion in the extracted point clouds (and subsequent meshes) due to the misregistration of tracked pixels in sequential photographs. Subject motion also manifests in the photogrammetric software during the alignment processes as a camera location anomaly. For this phase of the program, an alignment threshold of 10% was implemented in the vertical dimension for the projected camera / image location in Agisoft's display. In these processes the software estimates the camera location over the data collection interval (photo sets). When motion is present, the pixel tracking algorithms also have difficulty locating the same pixel from photograph to photograph which produced noise (and other artifacts) in the resultant point cloud and extracted mesh. Further, as the subject rotates, humans naturally try to focus their eyes on stationary locations which cause the pupils to appear to move over the scan duration; producing blurred images in this critical area of the face (Figure 1).

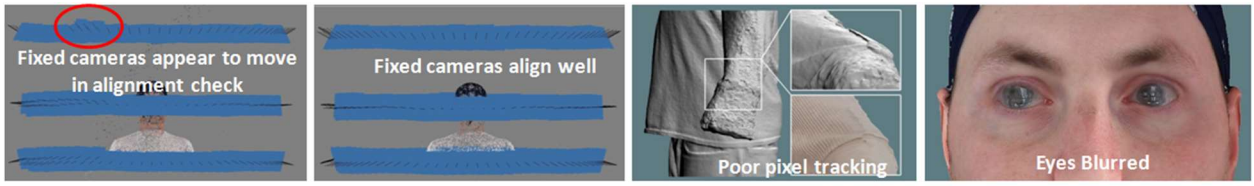

Figure 1. Manifestation of Subject Motion and Alignment Process Indicators

Since it was almost impossible to get subjects to look "straight ahead" during a rotation, a separate close up photo was taken of the face with the subject eye lids wide open in order to replace the extracted eye texture with the close up photo. This additional step was required on all the test subjects used during the project.

Number of Photos / Angular Intervals - Due to the nature of point cloud generation and pixel tracking in stereophotogrammetry, photographs must be collected in a manner that will resolve the fine details associated with the subject; particularly the head area where the topology changes rapidly over angular distance (Saffold, 2005). With the current system, around 99 photos were typically used for each camera during one turntable rotation (about a 3.63 degree angular interval). This allowed creation of a quality mesh without excessive processing time. As part of this trade space, we tested other photo intervals by cutting the amount of photos in half till we reached a point where the mesh was completely unusable. At about a 7.5 deg interval, the extracted mesh quality visibly deteriorated due to the point cloud result being too coarse in key facial areas such as the nose which has the highest topological variation over angular interval. Coarser intervals were looked at of 15 deg (24 photos) and 30 deg (12 photos) which produced further deterioration in the mesh quality.

Camera Horizontal and Vertical Diversity - Implementations using a single camera did not do well at closing the mesh in areas not directly facing the camera lens. As a result, 3 cameras were implemented to scan the head (and body) positioned so that the lower camera looked up at the focus point, the middle camera looked directly at the focus point, and the upper camera looked down on the focus point. The center camera was set up to cover the bulk of the subject focus area. The upper camera was angled down to cover areas like the top of the head and ears. The lower camera was angled upward to cover areas like under the chin and nose. Each of the cameras was placed (zoom, position, orientation) to provide some overlap between images that helped the software register the camera projection points and improved the overall accuracy of the scan. Figure 2 illustrates the scan results for a single camera centered on the focus point, and the three camera approach implemented with vertical diversity.

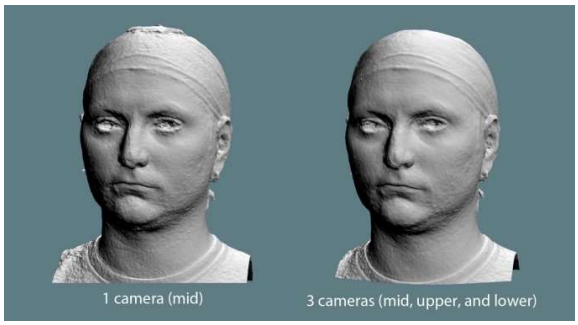

Figure 2. Camera Location Diversity Impacts

Background / Type - With the low cost capture process the background needs to be easily masked out. The easiest way to mask the background is to use a consistent background color that provided high contrast to all the foreground objects and is completely stationary over the entire scan. To achieve this, a cloth backdrop was used that is a solid

color made from a fabric that resists creasing. The photogrammetric software provided an ability to mask foreground objects from static backgrounds as part of the calibration process. This process involved providing the software a dedicated photograph (from each camera) of just the background itself. This particular method of masking in uses a tolerance value to find the difference in pixels between the background and foreground color values. Having a foreground subject that does not contrast well with the background can end up getting masked over. After creating the mask you can visually see it as darker areas with a white outline (Figure 3) which result in noisy mesh creation and significant "gaps" in these areas. As a result, either solid orange or while background colors were used for the test subjects depending on clothing and skin color.

Figure 3. Poor Contrast Between Back and Fore Grounds

Lighting Quality / Shadow Removal & Motion - High quality lighting is important for a number of reasons. From a photography stand point, quality lighting will allow for a small aperture that reduces the images depth of field which allows for high-detail, sharp imagery. It also allows you to use a lower sensor sensitivity setting (Camera ISO) that can reduce grain and can give you faster shutter speeds which can reduce motion blur. Having uneven lighting can create shadowed areas that will have less detail in your scans and those shadows can be "baked in" your texture maps. Further, shadows cast on the subject areas will cause misregistration of pixel locations in 3D space due to many pixels in the shadow area having similar "color" values. Thus, lighting impacts both the quality of the extracted mesh and the texture map created for the avatar. In order to provide uniform lighting, 4 diffuse indirect lighting systems were placed (and angled) for the subject illumination to (a) wash out the native fluorescent lighting, (b) provide a uniform light field across the focus area for more natural textures, and (c) eliminate shadows which cause registration errors in the extracted point cloud and compete with in-game lighting shadow algorithms.

Number of Pixels (Image Size) - A variety of images sizes were tested by lowering the resolution from 18 megapixels all the way down to 3 megapixels. Each of these photo sets were processed using a common set of program parameters. The amount of total points created while building the dense point cloud decreased from 2,289,954 in the 18 megapixel photoset to 324,339 in the 3 megapixel photoset. The extracted mesh showed a drop in poly count from 459,025 in the 18 megapixel photoset to 179,999 in the 6, 4, and 3 megapixels photosets. Extracted mesh detail began to visibly fade around the 8 megapixel resolution mark while resulting texture detail was preserved all the way down to the 4 megapixel mark. At 18 megapixel, the mesh clearly showed "bumps" at facial hair locations and the mouth, nose, and eye areas are sharply defined. The result of this trade indicated that mesh quality preservation was the driver in determining a requirement for number of pixels input to the program in the original photographs. It was noted that even small amounts of subject motion combined with UV mapping reduced the texture quality below the original photo sets regardless of output texture image size.

Subject Frame Occupancy – it was determined that focused objects on the subject need to occupy a significant portion of the photograph frame for mesh extraction quality. For a fixed pixel density (image size), higher subject frame occupancy provides both more pixels across the foreground object and finer subject information in a single pixel which improves the algorithms ability to track pixels uniquely across multiple photos. The trials performed in the trade space indicated that about 80% frame occupancy was required to produce good results.

When performing the body area scans, frame occupancy was difficult to achieve on the hands which have a lot of topographical variation in small space (fingers). As a result, "stock" hand models were used for all the virtualized subjects which were scaled to the morphology of the subject's arms. The stock hand models were attached to the extracted mesh in post processing. The ideal subject pose during data capture would be to have the arms away from the body, palms forward and the fingers open. This pose provides a visible space between arms, legs and fingers to give the photogrammetry software a best case to gain visibility in these areas and create the mesh correctly. The subject pose was also set to closely match the default pose associated with the rig. This pose is typically called a "Jesus Pose" where the arms are extended at a 45 deg down angle from the shoulder and the palms of the hand face the primary camera. The fingers on the hand are spread apart as much as possible in this pose.

THE VIRTUALIZATION PROCESS

The virtualization process steps used in this effort are illustrated in Figure 4 and clearly include many additional steps over and above mesh and texture extraction. The process for "virtualizing" live humans begins with a small amount of training to prepare the subject for positioning on the turntable, instruction on keeping still (facing one direction as the unit turns), and posing the arms and hands. The training includes rotating the subject a few times to gain a familiarity with the motion and verify the posing is correct. The remaining steps are discussed in this section.

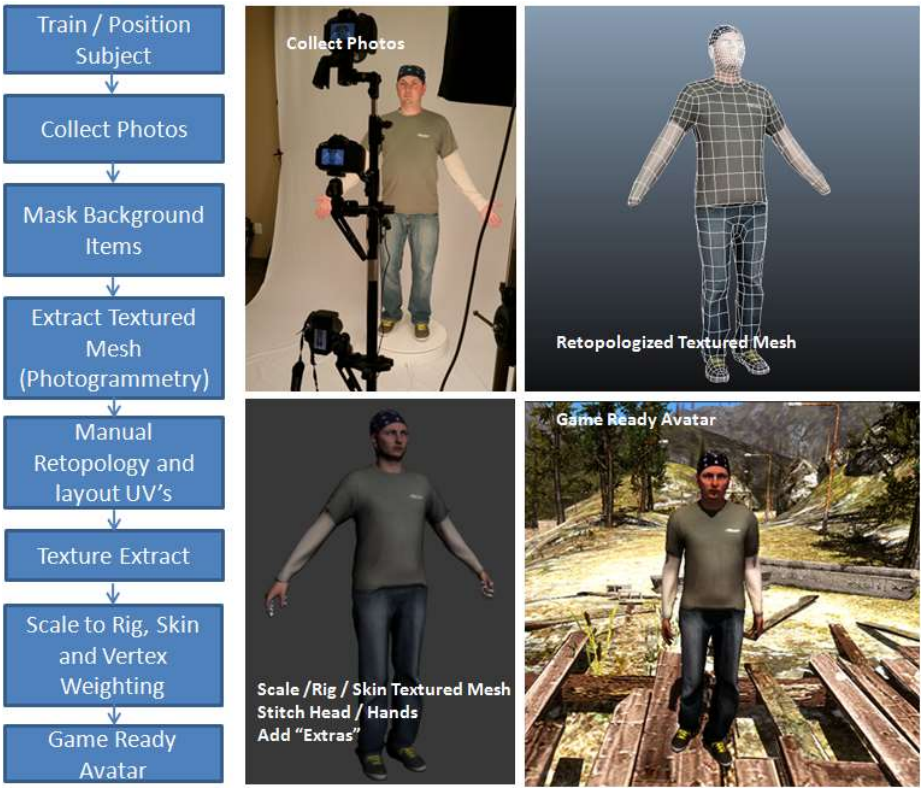

Figure 4. Virtualization Process

Collect Photos and Quality Check - The next step in the process is to document subject dimensions (height, arm span, ratio of arm span to total height) and collect the photo scans of the rotating subject. A scan involved a 360 degree rotation of the subject in front of the three camera rig.

After the "body scan" was complete, the three cameras were repositioned to focus on the head area leaving a little overlap with the subject's chest and a second scan was performed repeating the steps above. This overlap was used to help align the body and head meshes in a later step. Once the body and head scans were complete, the subject was moved closer to a single camera where a close-up of the face (eyes wide open) was taken to allow for eye textures to be used in the final model. Figure 5 illustrates the three key photo collection activities for the data collection step.

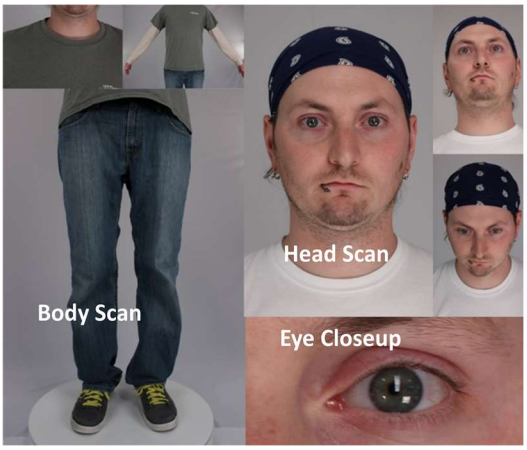

Figure 5. Body Scan, Head Scan, and Eye Close-up Data

The scan data were then imported into the photogrammetric software for a quick verification of photo alignments (subject motion). If the photo sets from the three cameras in the body or head scans did not vertically align well, the scan was repeated.

Mask Background Items - When masking the background in the photogrammetric software a photo was taken without the subject in the frame showing only the background elements. The "from background" option takes the differences in the background photo and the subject photos to mask the background leaving only the subject unmasked. Using the solid color background selected for maximum contrast between foreground objects worked well on all subjects.

Extract Textured Mesh – Within the photogrammetric software, a 4 step process was used to produce a textured high poly triangulated mesh. The process begins with photo alignment which finds the camera position and orientation while building a sparse point cloud model. Based on the camera positions we can then build a dense point cloud model that also calculates depth information for each camera. After the dense cloud model is created that data is then used to create a high poly triangulated mesh. From there the mesh is UV mapped and a high resolution texture is extracted from the photos to complete the process.

Retopologize and Layout UVs - The 3D model extracted from the photogrammetry software is a high poly triangulated mesh with over 400,000 polygons that must be reduced and retopologized to work in a game engine. The high poly extracted mesh from the photogrammetric software was extracted into Autodesk Maya LT (Autodesk – Maya LT, 2015) in order to perform the polygon reduction, retopologization, and UV layout. This process, manually implemented, took the longest time to complete because it requires planning that is unique to each mesh. Once retopologization is complete, a UV layout for texture wrapping was created. Performing a UV layout for the texture(s) consisted of creating a flattened 2D map of the 3D model. This was performed for both the body and head meshes.

Extract Textures – Once the mesh is retopologized and UV mapped, textures must be converted from the high poly triangulated mesh to the lower poly version and the new UV maps. In order to accomplish this, the texture extract tool available in Autodesk's Mudbox application (Autodesk Mudbox, 2015) was used. From the original high-poly mesh, an 8192 x 8192 uncompressed tiff image was extracted preserving the resolution of the original source texture. This texture was then imported into Adobe Photoshop (Adobe Photoshop, 2014) and decimated to 4096 x 4096; the maximum image size supported by the Unity3D engine. Of course, any resolution below the original source image is available when down-sampling in Adobe Photoshop. The final texture image was then saved in TGA format (uncompressed) and used with the model file when imported into Unity3D.

Scale Mesh to Rig / Vertex Weight – In the GDIS-Unity implementation, a common bone rig is used for all characters. This rig can be scaled and stretched as needed to match human morphological variations. In order for the new mesh to work with this rig, the mesh must be scaled and adjusted to account for morphology variation from the default. In this phase, the rigging was done manually but lends itself to automation in the next phase. The upper arm bones were rotated in the rig from the default "Jesus Pose" to match the mesh arm positions. The newly-aligned rig was then skinned (vertex weighted) using the new mesh. After skinning, the rig's pose was reset back to default. This method was also used on the legs in order to match the mesh position with the rig position. After the vertex weighting, scaling, and repositioning was complete, the weighted mesh was verified by playing the rig test animation to check that everything deformed correctly. Once deformation tests meet satisfactory requirements, the completed character (mesh and texture) were exported to the desired file format (FBX) for use in the target application. The additional post processing required to complete the full body model included (a) stitching the head and stock hands and (b) replacing eye textures. Extra items (glasses, earrings, etc.) are also attached in this step. These steps were performed prior to export of the textured mesh.

RESULTS

With this system, five fully rigged game ready avatars were created each with unique human morphologies. The subjects also included a wide range of human characteristics including one subject (Subject 5) with a number of "stressors". Figure 6 illustrates the five subjects submitted to virtualization and the virtualization results rendered as game-ready avatars in the GDIS-Unity system using the described processes. The rendered version with the game-ready avatars is being lit by in-game lighting which includes some color mapping causing the slight change in color appearances. The avatars are in an "idle" animation pose.

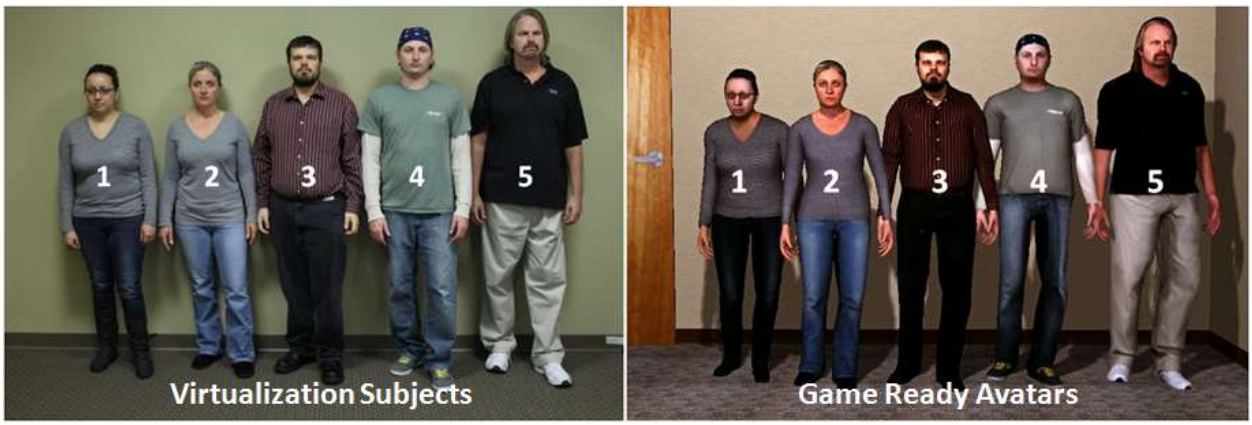

Figure 6. Subjects for Virtualization and Resulting Game Ready Avatars

The subject's ranged in height from 64" to 74" and the females had their hair tied back into tight "buns". With little or no automation techniques applied, Table 1 illustrates the time associated with each process for each subject.

Table 1. Virtualization Timeline (without Automation)

	Subject Number				
Process Step	**1**	**2**	**3**	**4**	**5**
Train / Position Subject	5	5	5	5	5
Collect Photos (background, body, head, eye)	10	10	10	10	10
Mask Background	60	50	50	60	120
Extract Textured Mesh	166	222	222	212	232
Repotologize / UV Layout	602	762	782	842	1262
Texture Extract	44	24	44	54	44
Rigging, Skinning, Vertex Weighting (Extras Included)	484	484	484	484	874
Total (Minutes)	**1371**	**1557**	**1597**	**1667**	**2547**
Total (Hours)	**22.85**	**25.95**	**26.62**	**27.78**	**42.45**

As illustrated in Table 1, Subject 5 proved the most difficult to virtualize using mostly manual processes due to the high number of stressors associated with this subject.

Human Subject Stressors [Subject 5]

Current game physics technologies have difficultly doing hair, flowing garments, and discontinuous clothing (shirt sleeves). The photogrammetric extraction algorithms have difficult identifying and tracking pixels when there is no color contrast between foreground and background. These algorithms also have difficulty with solid colors in the foreground object (like the shadow-less shirt on Subject 5). Based on these stressors, a few guidelines were established to keep the avatars simple enough for any game engine to handle and provide good scan data for mesh creation. The clothing preferences were long sleeves, pants and shoes to keep from having anything flowing away from body. The long haired subjects were asked to keep their hair in a "bun" or hidden under something like a bandana. Subject 5 violated all of these guidelines in order to illustrate their impacts on the virtualization timeline and resultant game ready avatar.

The short sleeves also caused difficultly when attempting to attach stock hands to the subjects "bare" arm since there was no area for placement of the hand discontinuity under a long sleeve. As such additional time had to be spent deforming the stock hand mesh vertices to match the arm vertices at the joint and texture matching between the stock hand and arm (including skin discoloration) had to be performed.

AUTOMATION OF PROCESSES

Automation approaches have been researched for the key virtualization steps. The current thinking to reduce the total virtualization times is summarized below:

- **Collect Photos** – a data capture system based on simultaneous low cost cameras and integrated lighting which completely cover the upper hemisphere around the subject offers the best approach combined with photo and camera management software
- **Mask Background** – a developed image processing script to auto detect background items and mask these pixels in production photographs which could be added to the photo and camera management software allowing the photos submitted for point cloud generation to already include a mask.
- **Extract Textured Mesh** – additional CPU's and the SDK for AgiSoft could be used to parallelize point cloud to mesh creation
- **Repotologize / UV Layout & Texture Extract** - R3DSWrap (R3DSWrap, 2015) is a topology transfer tool that allows the ability wrap existing topology around a 3D scan or other high poly model. It also includes pre-alignment, sculpting and texture baking features in addition to an SDK.
- **Rigging, Skinning, Vertex Weighting (Extras Included)** - 3DS Max Skin Wrap modifier (Autodesk 3DS Max, 2012) creates a skinned mesh based on another mesh (like a baseline male or female). Automatic rigging (and skinning) can be accomplished using Mixamo (Adobe/Mixamo, 2015)

Using the proposed approaches / tools for automation, the time to completely virtualize a "compliant" human with little or no stressors could be reduced to less than 2 hours.

SUMMARY AND CONCLUSIONS

A low cost virtualization system is described in this paper which provided good results on a wide variety of human subjects in a manageable timeframe. A number of trade spaces were evaluated related to data quality, camera setting and placement, and subject characteristics to illustrate all the steps needed to ultimately move from an original mesh (shell) to a game ready avatar. In the development of this paper a set of avatar guidelines are proposed which will provide good "game readiness" for most modern game engines. The virtualization system is still considered a research system as demonstrated by large amount of "manual" processes; however the potential to automate these processes exist today in most areas. This activity represents only the first phase of this research and additional experimentation and higher levels of system automation with a wider variety of subject morphologies and game engines will be accomplished in later phases. Based on the trade space results and lessons learned from virtualizing five humans, Table 2 summarizes a set of guidelines that should generally apply to any apparatus used for game ready avatar creating using photogrammetry.

Table 2. Data Capture and Subject Guidelines for Virtualization

Description	Guideline
Avatar Mesh Poly Count (1st LOD)	10,000 to 30,000 Polygons
Levels of Detail (LODs) Transitions	100% to 60% Screen Height (1st LOD) 60% to 30% Screen Height (2nd LOD) 30% to Cull Screen Height (3rd LOD)
Levels of Detail (LODs) Poly Reduction	2:1 (2nd LOD), 4:1 or greater (3rd LOD)
Texture Size	4096 x 4096 Pixels (uncompressed) Single Texture Sheet is optimum
Bone Rig	57 Bones with 2 "Twist Bones" per Arm
Photograph Angular Interval	4 degrees or better (horizontal scan plane)
Vertical Diversity	Angled up (45 deg), down (45 deg), and straight at focus area
Background	Solid color with high contrast to foreground optimum
Lighting	Diffuse Indirect with Shadows Eliminated Everywhere Possible
Photo Image Size for Mesh Extraction	8 MP or higher
Subject Focus Point Photo Frame Occupancy	80% or higher
Camera Coverage	Full 360 deg hemisphere over subject
Camera Image Quality	8 MP or better
Subject Characteristics	Pose: "Jesus Pose", Eliminate or minimize long flowing hair, flowing garments, solid color clothing, clothing / skin discontinuities, and reflective / translucent items (glasses, etc.) whenever possible

While the trial set presented is too low to draw any broad conclusions, the data indicated a number of trends that should be verified over a statistically valid set of trials, conditions, subject, and avatar definitions. Of course, the "quality' of a game ready avatar must also be weighed against the specific requirements associated with an application (like first vs. third person views, and avatar counts) or training exercise. Based on the limited data set and trials, the following key trends were noted:

- The turntable approach (while low cost) introduces a number of issues that must be addressed after data collection like (motion effects, moving shadows, stitching of subsequent scan data, and eye re-texturing) which would not be present using an "instantaneous" data capture system.
- Camera quality seems to have competing requirements related to mesh accuracy and texture quality. This paper used mesh accuracy as the driver for virtualization.

REFERENCES

Adobe, Photoshop CC Help Reference Guide, (2014). https://helpx.adobe.com/pdf/photoshop_reference.pdf

Adobe/Mixamo, Mixamo Documentation, (2015). https://community.mixamo.com/hc/en-us/categories/200186373

Agisoft, Agisoft PhotoScan User Manual Professional Edition, Version 1.0.0, 2013

Autodesk, Mudbox (2015). Mudbox 2015 Reference, 2014, http://docs.autodesk.com/MUD/2015/ENU/#!/url=./files/mePortal.htm

Autodesk, 3DS Max (2012). 3DS Max Reference, 2012, http://download.autodesk.com/us/3dsmax/2012help/index.html

Autodesk, Maya LT 2015, Maya 2015 Reference, (2014). http://help.autodesk.com/view/MAYAUL/2015/ENU/

Banakou, D., & Chorianopoulos, K. (2010). The effects of avatars' gender and appearance on social behavior in online 3D virtual worlds. *Journal for Virtual Worlds Research*, *2*(5).

Baqersad, J., et al, (April 26 2012). Dynamic Characteristics of a Wind Turbine Blade using 3D Digital Image Correlation, Proc. SPIE 8348, *Health Monitoring of Structural and Biological Systems 2012*

Bente, G., Rüggenberg, S., & Krämer, N. C. (2004). Social presence and interpersonal trust in avatar-based, collaborative net-communications. In *Proceedings of the Seventh Annual International Workshop on Presence* (pp. 54-61).

Debevec, Paul, University of Southern California (USC) Institute for Creative Technologies (ICT), (2012). The Light Stages and Their Applications to Photoreal Digital Actors, *SIGGRAPH Asia*

Garsthagen, Richard, Pi 3D Scanner: a DIY Body Scanner, (2013). https://www.raspberrypi.org/pi-3d-scanner-a-diy-body-scanner/

Kocon, Maja, (23–26 October 2010). Rigid Bones Grouping Scheme for Facial Expressions Synthesis Utilizing Three-Dimensional Head Model, West Pomeranian University of Technology*, XII International PhD Workshop OWD 2010*

Lundstrom, T., et al, (2012). Using High-Speed Stereophotogrammetry Techniques to Extract Shape Information from Wind Turbine/Rotor Operating Data, Topics in Modal Analysis II, Volume 6, *Conference Proceedings of the Society for Experimental Mechanics Series*

PhotoModeler Close-Range Photogrammetry and Image Based Modeling, (2015). Extracts 3D Measurements and Models from Photographs Taken with an Ordinary Camera., http://www.photomodeler.com/index.html.

Remondino, Fabio, (2010). From Point Cloud to Surface: The Modeling and Visualization Problem, Institute of Geodesy and Photogrammetry Swiss Federal Institute of Technology, *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* Vol. XXXIV-5/W10

Saffold, Jay, Phase II Photogrammetric Camera 3D Digitization System, Final Technical Report, (July 2005). RNI-RPT-05-0730-01, US Army CECOM C2 Directorate, Ft Monmouth NJ, SBIR Topic # ARMY 01-099, Contract # DAAB07-03-C-L006

Schenk, T., 2005, Introduction to Photogrammetry, GS400.02, Department of Civil and Environmental Engineering and Geodetic Science, the Ohio State University, 2005.

Tan, Joai Thong, (2007). Autodesk, Facial Expression Using Morpher in 3D Studio MAX, http://www.tootish.com/bakuteh/tutorials/MAX/Morph_Targets.pdf

Unity3D, (2014). AI System for Mecanim Documentation, http://zerano-unity3d.com/AI%20System%20Doc.pdf