Biofidelic Virtual Terrorist —A Modeling and Simulation Tool for Human Threat Recognition Training

Zhiqing Cheng, Stephen Mosher
Tim MtCastle, and Jeanne Smith
Infoscitex Corporation
Dayton, Ohio, USA
{Zcheng, Smosher}@infoscitex.com
{Tmtcastle, Jsmith}@infoscitex.com

Julia Parakkat and Kathleen Robinette
711th Human Performance Wing
Air Force Research Laboratory
Dayton, Ohio, USA
Julia.parakkat@wpafb.af.mil
Kathleen.robinette@wpafb.af.mil

ABSTRACT

Effectively detecting suspicious, hostile, and terrorist activities and human threats in both populated urban environments and in remote terrains has become increasingly important to US homeland security and military operations in asymmetric wars. A virtual terrorist recognition training system based on modeling and simulation can be used to train operators to recognize human threats from various ranges, viewing angles, and resolutions. However, human modeling and simulation technology currently used in these training tools lacks sufficient biofidelity and thus is not able to describe and demonstrate the nuances of human activities and human signatures that are indicative of threat. Therefore, the Air Force Research Laboratory is developing a software tool called the Biofidelic Virtual Terrorist (BVT). By using advanced dynamic 3-D human modeling technology, the BVT replicates and creates human threat activities in 3-D space. The development emphases of the BVT were placed on the biofidelity of human body shape and motion, representation and immersion of biosignatures of human threats, fidelity and reality of virtual environments or scenarios, and the interaction between humans and environments. In addition to being used as a training tool, the BVT can be used as a test bed for evaluating the performance of tools developed for automatic human threat detection. The BVT can also be used to generate data for rare activities and scenarios which otherwise would be impossible or hard to acquire in the real world.

ABOUT THE AUTHORS

Dr. Zhiqing Cheng is a principal engineer 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 or a major investigator and published over 70 technical papers as the lead author. Currently, as a technical lead, he is working on the HMASINT program, a multi-directorate endeavor by the AFRL, to develop technologies for human measurement and intelligence information.

Julia Parakkat earned the degree of B.S. in Biosystems Engineering (2002) from the University of Kentucky and the M.S. in Industrial Engineering with an emphasis in Biomechanics (2004) from the Ohio State University. She currently serves as the Chief of the Human Biosignatures Section in the Human Effectiveness Directorate of the Air Force Research Laboratory. Her research interests are in biomechanics and biosignatures as a means to characterize human activities.

Dr. Kathleen Robinette earned the degree of B.A. in Anthropology (1978) and the M.S. in Mathematics/Statistics (1983) from Wright State University, and the Ph.D. in Biostatistics (2003) from the University of Cincinnati. She is a Fellow of the Air Force Research Laboratory (AFRL) as well as the Human Factors and Ergonomics Society. She is currently leading AFRL's Strategic Technology Thrust on Developing and Exploiting Behavior Understanding with Layered Sensing (DEBU) and the Biosciences and Performance Division's ISR Biosignatures Team. She is President of the World Engineering Anthropometry Resource (WEAR) Association, Co-Chair of the CODATA technical committee on anthropometry, serves on the National Academies Board on Research Data and Information, and as an adjunct faculty member at Wright State University. She is perhaps best known for directing the first successful 3-D whole body human measurement survey (CAESAR), through partnerships that included 35 government, industry, and academic organizations.

Biofidelic Virtual Terrorist —A Modeling and Simulation Tool for Human Threat Recognition Training

Zhiqing Cheng, Stephen Mosher
Tim MtCastle, and Jeanne Smith
Infoscitex Corporation
Dayton, Ohio
{Zcheng, Smosher}@infoscitex.com
{Tmtcastle, Jsmith}@infoscitex.com

Julia Parakkat and Kathleen Robinette
711th Human Performance Wing
Air Force Research Laboratory
Dayton, Ohio
Julia.parakkat@wpafb.af.mil
Kathleen.robinette@wpafb.af.mil

INTRODUCTION

Effectively detecting suspicious, hostile, and terrorist activities and human threats in both populated urban environments and remote terrains has become increasingly important to US homeland security and military operations in asymmetric warfare. A virtual terrorist recognition training system based on modeling and simulation (M&S) or serious games can be used to train operators to recognize human threats from various ranges, viewing angles, and resolutions. Therefore, the Air Force Research Laboratory is developing a software tool called the Biofidelic Virtual Terrorist (BVT). By using advanced dynamic 3-D human modeling technology, the BVT replicates, creates, and reconstructs human threat activities in 3-D space. Since human modeling and simulation technology currently used in M&S based training tools lacks sufficient biofidelity and subsequently is not capable of describing and demonstrating the nuances of human activities and human signatures that are indicative of human threats, the development emphases of the BVT were placed on the biofidelity of human body shape and motion, representation and immersion of biosignatures of human threats, fidelity and reality of virtual environments or scenarios, and the interaction between human models and their virtual environments.

REQUIREMENT ANALYSIS

As a tool to be primarily used for human threat recognition training, basic requirements of the BVT are described below.

Bio-fidelity

Disguises are a common tactic used by terrorist groups. Therefore, the cues or indications of human threats are usually very elusive and vague. Therefore, human M&S technology must provide sufficient bio-fidelity to describe and demonstrate the nuances of human activities and human signatures. Due to the human-centricity, this is particularly critical for human threat recognition training and dismount detection training. Therefore, in the BVT development, emphasis was placed on the bio-fidelity of human modeling to provide realistic description and representation of human features, so that human signatures associated with a particular activity or threat can be faithfully exhibited.

There are various types of biological features or bio-signatures inherent in humans and their activities. The features that can be observed from long distances and can be used for activity recognition and threat detection include body shape and motion. Therefore, in the BVT, efforts were made on representing, modeling, and displaying these features.

Virtual reality fidelity

A human activity or threat always occurs in a space or environment. Spatial or environmental features provide important information for activity recognition and threat detection. For instance, a digging activity could occur in a backyard or by a road side; however, the implication of this activity in these two environments may be quite different. Spatial or environmental features inevitably come into human vision and thus influence human cognition. Therefore, creating a virtual reality that faithfully represents a geographic region of interest and actual scenarios becomes important and necessary.

Immersion

A virtual reality needs to provide the users or trainees real cognitive exposures and experience as much as possible. Immersion can be achieved and enhanced by providing (a) multiple features of space/environment and human bio-features (shape, motion, facial expression, voice); (b) fidelity of these features; (c) 3-D representation and display of these features; and (d) freedom of view manipulation.

Interaction
This allows (a) Human subjects (digital avatars) to interact with a 3-D environment; and (b) Human subjects to interact with each other. This also implies that users have various ways to control or manipulate the display of contents.

User selection
A wealth of content needs to be provided to allow users to select from a list of typical activities with typical scenarios.

User creation
It is neither possible nor feasible to create an exhaustive activity library. This means that it will be beneficial for the users to create new activities by prescribing parameters for human (e.g., gender, height, weight), activity (e.g., running, shooting), setting (e.g., building, street), and geo-locations (e.g., a particular place in the world).

Analytical Functions
Certain analytical functions may help users to enhance or expand their cognitive capabilities for threat detection. These include (a) shape variability analysis; (b) gait and motion analysis; and (c) spatial-temporal pattern recognition.

Cultural features and other factors
Humans have large variations from one subject to another in terms of shape and motion. These variations depend on gender, age, ethnicity, cultural background, and other factors. On one hand, we need to take these features and variations into account during modeling so the models can faithfully describe the body shape and motion of subjects with specific attributes; on the other hand, when a model represents these features faithfully, it will help a user or trainee to recognize the subject.

TECHNOLOGY DEVELOPMENT

System Layout
Figure 1 illustrates the system layout of the BVT as a software tool. The development process consists of three phases: Phase I: activity replication; Phase II activity creation; and Phase III: activity reconstruction. Modeling and simulation technology has been used to build the BVT. Major technical issues in each phase are addressed in the following sections.

Phase I—Activity Replication
Activity replication is replication of a human activity that has been recorded in a laboratory via 3-D modeling. The 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. The data that can be readily used for activity replication are not available. Alternatively, a motion capture system can be used to capture markers on the body during motion and a 3-D whole body scanner can be used to capture the body shape in a pose. Based on the body scan data and motion capture data, some techniques can be used to build a digital model to replicate a human activity in 3-D space or to animate the model with the prescribed motion.

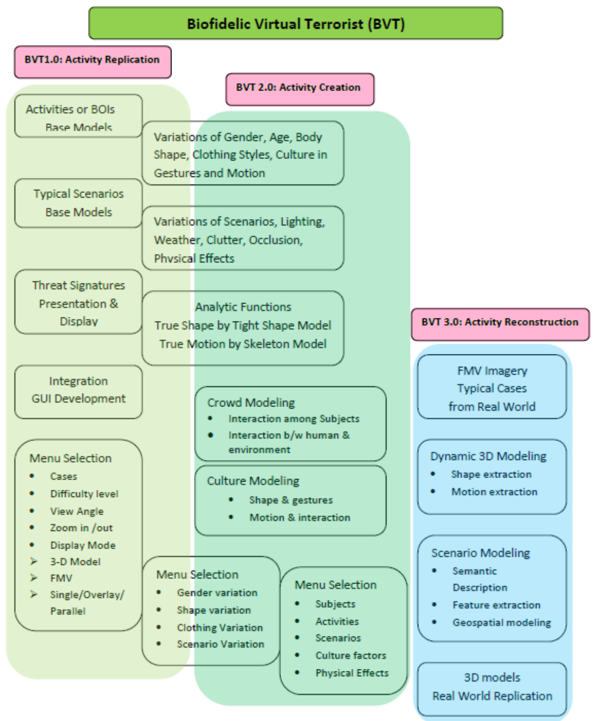


Figure 1. Outline of the BVT development process.

Several open-source software tools were used for activity replication. MeshLab was used to process 3-D scan data, OpenSim was used to derive skeleton models and the associated joint angles from motion capture data, and Blender was used to create an animation model that integrated body shape and motion. The technology development process started with the replication of four simple activities: walking, jogging, limping, and shooting. The tests of human subjects performing the activities and data collection were conducted in the 3-D Human Signatures Laboratory (HSL) at the Air Force Research Laboratory (AFRL). The data included scan and motion capture (MoCap). The human subject, with markers attached, was scanned using the Cyberware whole body scanner. Motion capture data were acquired for the same subject with the same markers attached. The markers allowed the joint centers to be determined for both the scan data and the MoCap data.

The body scan data acquired consists of a large number of data points (vertices) (typically a half-million or more) and may contain holes and large openings. The data was processed so that it could be used for the modeling. MeshLab was used to clean-up the data and to fill holes. Smoothing and approximation functions in MeshLab were implemented to reduce the total number of vertices for each subject scan to 50,000 and to create meshes of the body shape required for the modeling. OpenSim was used to derive a skeleton model from MoCap data (TRC file) and to calculate the joint angles for the skeleton. This process is called inverse kinematics. The skeleton model and associated joint angles were put in a Bio-vision Hierarchical (BVH) file. Both the body surface mesh data (about shape) and the BVH file (about motion) were imported into Blender. Blender was used to integrate the shape with the motion and to create an animation model that replicates an activity. Figure 2 shows the models created for four activities at a particular frame.

(a) **Limping** (b) **Jogging**

(c) **Shooting** (d) **Walking**

Figure 2. Replication of a subject in four activities.

Since the body shape and motion are derived from their respective data files, in certain cases, they can be used independently to create new activities. For instance, the MoCap data of a subject while walking without a vest was used to animate the body shape of the same subject while walking with the vest. Both body shapes were acquired using a 3-D laser scanner. Figure 3 (a) and (b) illustrates a subject walking without and with a vest, respectively.

(a) **Walking without a vest**

(b) Walking with a vest

Figure 3. Dynamic shapes of a subject at several frames during walking.

In the BVT, activity replication was designed to provide the following options:

- One subject performing different activities
- Different subjects performing the same activity
- Different subjects performing different activities
- Complex activities synthesized from simple activities.

Phase II—Activity Creation

It is often desired to virtually create certain activities for certain subjects with prescribed requirements. While replication allows us to replicate a particular activity for a particular subject, it is neither feasible nor possible for us to acquire body shape and motion data for every subject of interest. Virtual human activity creation is common nowadays, as numerous commercial software products are available for use. However, a major challenge is the biofidelity of the virtual activities created, which is essential to human threat recognition training. In the BVT development, high biofidelity was pursued for virtual activity creation via advanced human shape modeling, motion mapping, and 3-D scenario modeling.

Human Shape Modeling

From the perspective of the motion status of the subject to be modeled, human shape modeling can be classified as either static or dynamic. Static shape modeling creates a model to describe human shape in a particular pose, usually a standing pose. Dynamic shape modeling addresses the shape variations due to pose changes or due to the subject being in motion. Extensive investigations have been performed on human shape modeling (Hilton et al, 2002; Allen et al, 2003; Seo et al, 2003; Anguelov et al, 2005; Park & Hodgins, 2006; Balan et al, 2007; De Aguiar et al, 2008). Recent developments in human shape modeling include static shape modeling based on range scan data and dynamic shape modeling from video imagery.

Static Shape Modeling

The human body comes in all shapes and sizes. Our 3-D human static shape modeling is based on the 3-D laser scan data from the CAESAR database. A method developed by the authors (Cheng and Robinette, 2009) uses contour lines as the basic entities for shape modeling, as shown in Figure 4, which provides a compact shape representation and makes shape parameterization possible.

Shape Reconstruction/ Creation

Given a number of scan data sets of different subjects, a novel human shape can be created that has resemblance to the samples but is not the exact copy of any existing ones. This can be realized in the following four ways.

- Interpolation or morphing

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 moving. Figure 5 illustrates the morphing based on contour lines representation of human body shape.

- Reconstruction from eigen-space

After principal component analysis (PCA), the features of sample shapes are characterized by eigen-vectors or eigen-persons which form an eigen-space. Any new shape model can be generated from this space by combining a number of eigen-persons with appropriate weighting factors.

- Feature-based synthesis

Once the relationship between human anthropometric features and eigen-vectors is established, a new shape model can be constructed from the eigen-space with desired features by editing multiple correlated attributes, such as height and weight or fat percentage and hip-to-waist ratio. This approach is particularly useful for activity creation, since the desired features of a subject can be clearly defined.

- Marker-only matching

Marker-only matching is a way to reconstruct using the provided markers. This is important for many applications such as deriving a model from video imagery, since marker data can be obtained using less expensive equipment than a laser range scanner.

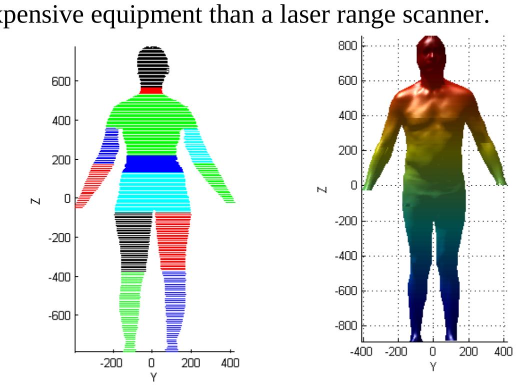


Figure 4. Static shape modeling based on contour lines.

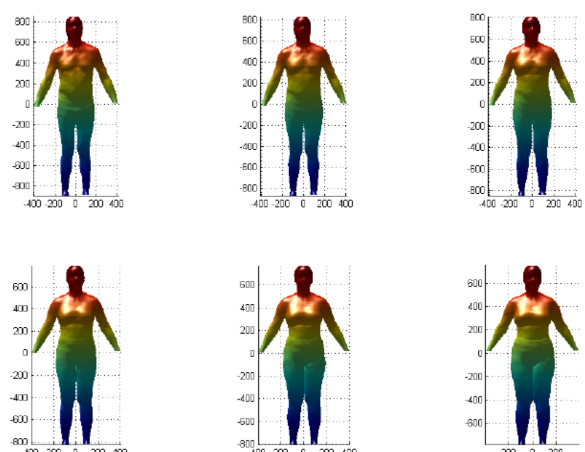


Figure 5. Morphing from a male to a female.

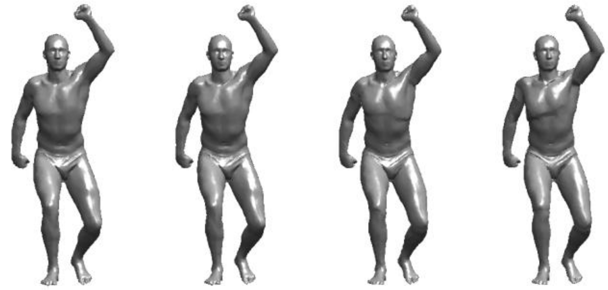
Pose Modeling

During pose changes or body movement, muscles, bones, and other anatomical structures continuously shift and change the shape of the body, that is, the body shape is changing dynamically during pose change or motion. One way to model body shape changes in different poses is called part blending or skinning by interpolating between two adjacent segment surfaces according to certain criteria (e.g., linear interpolation, bone-heat used in Blender, and nonlinear interpolation). While part blending may be able to describe the body shape changes caused by the articulated motion of body segments fairly well (as was in activity replication), it may not be able to describe the shape changes induced by surface deformation (soft tissue deformation).

Alternatively, body shape changes in different poses can be addressed via pose modeling which has recently drawn significant investigations (Allen et al, 2003; Anguelov et al, 2005). For pose modeling, scanning the subject in every pose is impractical; instead, body shape can be scanned in a set of key poses, and then the body shapes corresponding to intermediate poses are determined by smoothly interpolating among these poses. The issues involved in pose modeling include pose definition and identification, skeleton model derivation, shape deformation (skinning), and pose mapping.

In recent years, pose modeling was investigated at the AFRL, and a method was developed by the authors for the characterization and reconstruction of body shape deformation based on PCA (Cheng et al, 2010). The template model used for pose modeling is shown in Figure 6 (Anguelov et al, 2005). The entire body consists of 16 segments, each of which has the pre-defined surface division as shown in the figure. The pose modeling procedure consists of the following steps: (a) coordinate transformation; (b) surface deformation characterization; and (c) surface deformation reconstruction. Figure **7** displays a screen capture of body shape reconstructed for a known pose using principal components.

Figure 6. Template model for pose modeling.



(a) (b) (c) (d)

Figure 7. Shape reconstruction: (a) Original; (b) Full reconstruction; (c) Reconstruction with 20 largest PCs; and (d) Reconstruction with 10 largest PCs.

Shape prediction for new poses

It is not feasible to measure the surface deformation of each subject for all possible poses, because the human body has a large number of degrees of freedom and can take virtually an infinite number of different poses. In fact, only a limited number of poses can be investigated in tests, but it is often required to predict surface deformation for new poses that have not been observed. Three methods can be used to predict surface deformation.

- Method-1: using principal components. Given the joint twist angles for a segment to define a particular pose, projection coefficients can be estimated. Using the full or a partial set of principal components, the surface deformation is reconstructed.
- Method-2: taking the nearest neighbor pose. Given the joint twist angles, find the nearest neighbor to the prescribed pose and take its surface deformation as an approximation. The neighborhood is measured in terms of the Euclidean distance between the joint twist angles for the two poses.
- Method-3: interpolating between two nearest neighbors. Given the joint twist angles, find two nearest neighbors to the prescribed pose. The pose deformation is determined by interpolating between the deformations of these two neighbor poses.

Figure 8 illustrates the predicted shape for 8 different poses using method-2.

Human Motion Mapping

Note that motion capture can only be performed on a limited number of subjects. It is not practical to capture the motion data for each subject for a large number of subjects. In order to create a particular activity for any subject, motion mapping can be used to map the motion from one subject (whose motion is captured) to another (whose motion is not captured) by assuming that changes in major joint angles remain the same for different subjects performing the same activity. While two subjects may perform the same activity quite differently, a common pattern (which consists of several key poses) can be defined for a particular activity in a statistical sense. Although the applicability and usefulness of motion mapping based on this common pattern can be reasonably justified, some advanced methods can be used to recalculate the motion balance and stability of the subject being mapped.

Deriving joint angles from motion capture data is called inverse kinematics. In the BVT, OpenSim is used to derive a skeleton model and the associated joint angles from motion capture data.

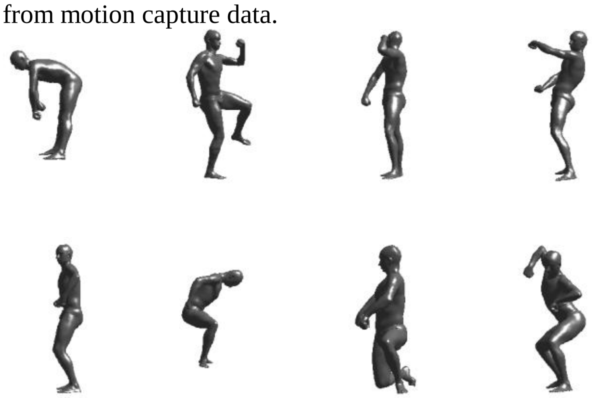


Figure 8. Predicted shape in eight different poses.

3-D Scenario Modeling

Various technologies have been developed in recent years for 3-D scenario modeling, some of which are data-driven. Commercial software packages (e.g., Maya, 3dsMax) are available for use also. In the BVT, Blender was used to create 3-D spatial models. Large real world scenario models were created first. Then small space models with desired details were created and incorporated into large scenario models.

Phase III—Activity Reconstruction

For training purposes, it is often necessary to replay a historical event. Traditionally, this is mainly based on the replay of a 2-D video imagery. Since a recorded 2-D video imagery or full motion video (FMV) has limitations of view angle, distance, occlusion, and cluttering, the actual effectiveness of such training is limited. However, if a dynamic 3-D model can be reconstructed from 2-D video imagery for a typical historical event, the model will provide various viewing operations to the trainees or users, such as free-point view, zoom-in or zoom-out, and rotation and can be incorporated into other scenarios or be integrated with other events or activities.

Activity reconstruction includes the reconstruction of scenarios (space models) and the reconstruction of human activities (human models). The latter problem can be stated as extracting a dynamic human shape model from 2-D video imagery. A dynamic human shape model describes the body shape, the body shape changes due to pose changes and motion, and the body motion. Since the primary interest for human threat recognition and detection is the human subjects and human activities, dynamic shape modeling/extraction from 2-D video imagery was a focal point in the BVT development and will be addressed in detail in the following section.

Dynamic shape extraction from 2-D video imagery

Challenges

Dynamic modeling describes or captures body shape changes while the human is in motion. Dynamic human shape modeling is a challenging topic because (a) the human can take various poses; (b) video imagery provides an incomplete view of the body due to segment occlusion; and (c) video imagery is often contaminated with noise due to light variations, view point/angle, etc.

Approach

Since the human is in motion, video imagery could capture the human body from different viewing angles even if only one camera is used. Therefore, given a multi-viewpoint video record, if the capture time is long enough, it can be reasonably assumed that: (a) the subject exposed every part of his body at some point to the camera (in a common sense); and (b) the subject took all poses associated with the activity played. Robust and efficient dynamic modeling needs to make use of the information contained in each frame of video imagery and fuse the information obtained from all frames. As such, dynamic modeling will be more capable of capturing the human body shape and recognizing human activities.

A strategy for dynamic modeling developed by the authors is illustrated in Figure 9**.** It uses 2-D video imagery as the input and provides a dynamic model as the output. Dynamic modeling is treated as an iterative process that consists of multiple steps. The details of the entire scheme are described in (Cheng et al, 2009), whereas the model fitting process is described as follows.

Model Fitting

Each instance model provides a set of initial values for the control parameters of the model, which are usually not sufficient enough for the description of the ground truth of the shape. The estimation of the parameters of the true model is via fitting the projection of the instance model to the silhouette extracted from video imagery. The problem of model fitting can be formulated as an optimization problem.

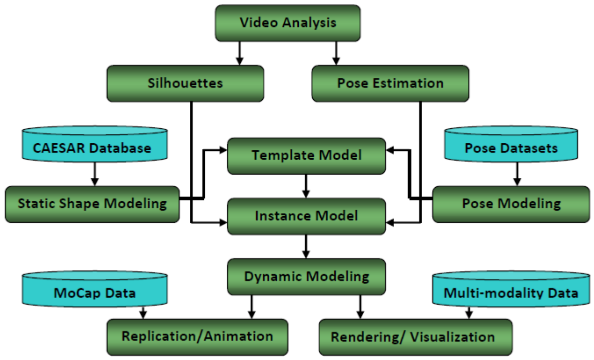


Figure 9. Dynamic modeling scheme.

Denote

$$\mathbf{p}_m = \{\boldsymbol{\beta}^T \ \boldsymbol{\alpha}^T\}^T \qquad (1)$$

as the vector of model parameters, where

$$\boldsymbol{\beta} = \{\beta_1 \ \beta_2 \ldots \beta_n\}^T \qquad (2)$$

as the control parameters for the shape variation and

$$\boldsymbol{\alpha}^T = \{\alpha_1 \ \alpha_2 \ldots \alpha_m\}^T \qquad (3)$$

as the control parameters for the pose-dependent surface deformation. Then, from the template model,

$$\mathbf{S} = S(\mathbf{p}_m) \qquad (4)$$

which is a shape descriptor vector and represents the shape model corresponding to control parameters $\mathbf{p}_m$.

For a given camera view, a foreground silhouette F^I, which extracts the subject from the background, is computed using standard background subtraction methods. The hypothesized shape model is projected onto the plane which is defined by F^I:

$$F^M = P(\mathbf{S}, \gamma) \qquad (5)$$

where γ is the parameter related to camera view which may or may not be known. The projection F^M can be considered as the estimated silhouette in the same frame. In the BVT, the extraction of a dynamic model from video imagery is conducted by fitting F^M to F^I for a sequence of image frames. The method proposed by Balan et al (2007) was used. The cost function is a measure of similarity between these two silhouettes. For a given camera view, a foreground silhouette F^I is computed using standard background subtraction methods, which is then compared with the model silhouette F^M. The pixels in non-overlapping regions are penalized in one silhouette by the shortest distance to the other silhouette and vice-versa (Sminchisescu & Telea, 2002). To do so, a Chamfer distance map (Stauffer & Grimson, 1999) is computed for each silhouette, C^M for the hypothesized model and C^I for the image silhouette. The predicted silhouette should not exceed the image foreground silhouette (therefore minimizing $F^M C^I$), while at the same time try to explain as much of it as possible (thus minimizing $F^I C^M$). Both constraints are combined into a cost function that sums the errors over all image pixels px:

$$f(\mathbf{p}) = \frac{1}{|px|} \sum_{px} (\delta F^M_{px} C^I_{px} + (1-\delta) F^I_{px} C^M_{px}) \qquad (6)$$

where

$$\mathbf{p} = \{\mathbf{p}_\mathbf{m}^\mathbf{T} \ \gamma\}^T \qquad (7)$$

including the control parameters of the model and the parameters related to camera view, and δ weighs the first term more heavily because the image silhouettes are usually wider due to the effects of clothing. When multiple views are available, the total cost is taken to be the average of the costs for the individual views.

Now, model fitting can be formulated as an optimization as follows:

$$\begin{aligned} &\text{Design variables} : \mathbf{p} \\ &\text{Objective function} : \text{Min}\{f(\mathbf{p})\} \\ &\text{Constraints} : \mathbf{p}^L \le \mathbf{p} \le \mathbf{p}^U \end{aligned} \qquad (8)$$

Where $\mathbf{p}^L$ and $\mathbf{p}^U$ are lower and upper bounds on the design variables, respectively.

THREAT RECOGNITION TRAINING

The BVT can be used as an assistant tool for human threat recognition training. It is not a complete training tool at its current stage, since a training tool needs to be designed based on cognitive science and principles. However, certain features are developed and introduced to the BVT to enhance training performance and efficiency.

Model Correlation

One way to enhance cognitive performance is model correlation, as shown in Figure 10. In the real world, malicious subjects tend to disguise themselves by putting loose garments on their body. Therefore, it is very hard to detect the ground truth of the body shape and motion. By creating the skeleton model, tight-shape model, and loose-shape model for the same subject in the same activity, the correspondence among the three models allows trainees to correlate the loose-shape model with the tight shape model and the skeleton model and thus may help them to identify covered objects or to detect suspicious activities.

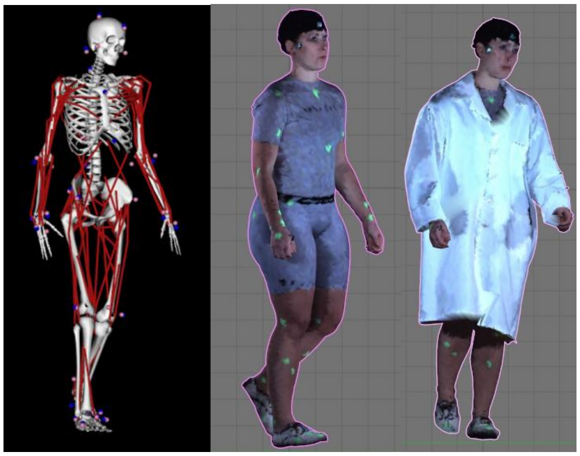


Figure 10. Model correlation among skeleton, tight-shape, and loose-shape models.

Game Features

There has been an increased interest in the use of serious games, or the use of games for non-entertainment purpose (e.g., training) (Ratwani et al, 2010). To enhance training efficiency and to provide some incentive to the trainees, certain game features, e.g., scores and awards, performance feedback, individual competition, and group competition, are planned to be introduced into the BVT framework.

Progressive training process

Training is a learning process. Trainees improve their cognitive performance progressively during training. Therefore, the complexity and difficulty level of human threat must be arranged in accordance with the learning process. In the BVT, human models are developed to represent human activities in the order of

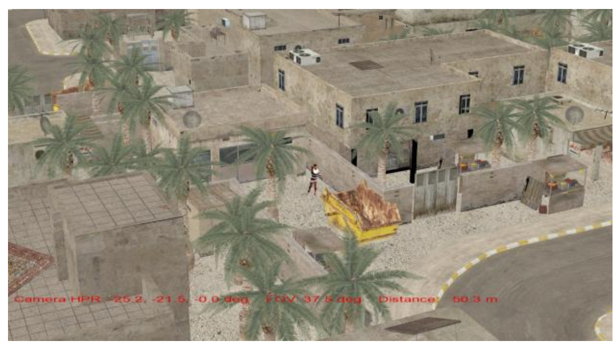
- From simple to complex;
- From single to synthesized;
- From individual subject to multiple subjects;
- From multiple subjects to crowds;
- From skeleton model to tight-shape model;
- From tight-shape models to loose-shape models;
- From particular activities to activity patterns.

BVT ILLUSTRATION

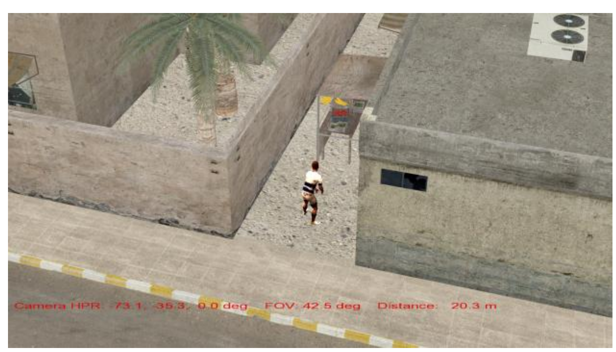
Currently, the BVT is put on the platform provided by SubrScene, an open source software system that has been developed at AFRL for virtual reality modeling and simulation. However, the BVT can be easily incorporated into other platforms, such as VBS2, MetaVR, and RealWorld. Figure 11 (a)-(d) are screen captures of the BVT, illustrating four human activities taking place at different locations and viewed from different distances.

(a) **Walking**

(b) Running



(c) Throwing



(d) Digging

Figure 11. Illustration of BVT.

CONCLUSIONS

Human threat signatures are important to human threat recognition and dismount detection. In order to represent and demonstrate the nuances of human activities and signatures that are indicative of human threats, digital human models must provide highly bio-fidelic descriptions of human body shape and faithful simulations of human motion.

To address the needs from the human threat recognition training community, the BVT was designed and developed to create virtual human activities with high bio-fidelity at three levels: activity replication, activity creation, and activity reconstruction. Methods and techniques were developed to address key technical issues involved in human activity modeling, such as static shape modeling and morphing, pose modeling and deformation mapping, and dynamics model extraction from 2-D video imagery. These human activity models were incorporated into highly fidelic 3-D scenario models, providing natural and realistic exposure and experience to trainees/users. To enhance cognitive training effectiveness, game features, model correlation, and progressive learning were introduced into the BVT. The initial models from the BVT were provided to a user group, and positive feedback and evaluation was received. In addition to being used as a training tool, the BVT can be used as a test bed for evaluating the performance of tools developed for automatic human threat detection. The BVT can also be used to generate data for rare activities and scenarios which otherwise would be impossible or hard to acquire in the real world.

While various modeling and simulation tools and serious games have been developed for mission training, due to inherent complexities, technology using M&S for human threat recognition training is still in the early development stage. Due to great variation and uncertainty of human activities and behaviors, more research and development efforts are needed to address many technical challenges, such as human behavior modeling, group activity modeling, and crowd modeling. One recent development of human modeling is the autonomous digital human agent that integrates motor, perceptual, behavioral, and cognitive components into a comprehensive human model so that it can perform a variety of natural activities (Terzopoulos 2009). Enhanced with high bio-fidelity and incorporated with human threat signatures, such an agent can be used in wide applications including human threat recognition training.

REFERENCES

Allen, B., Curless, B., Popvic, Z. (2003). The space of human body shapes: reconstruction and parameterization from range scans. *ACM SIGGRAPH 2003*, 27-31.

Anguelov, D., Srinivasan, P., Koller, D., Thrun, S., Rodgers, J., Davis, J. (2005). SCAPE: Shape Completion and Animation of People. *ACM Transactions on Graphics (SIGGRAPH)* 24(3).

Balan, A., Sigal, L., Black, M., Davis, J., and Haussecker, H. (2007). Detailed Human Shape and Pose from Images. *IEEE Conf. on Comp. Vision and Pattern Recognition* (CVPR).

Cheng, Z., Smith, J., Parakkat, J., Cheng H., and Robinette, K. (2010). Human Shape Modeling in Various Poses, a chapter in Advances in Applied Digital Human Modeling, CRC ISBN: 978-1-4398-3511-1, eBook ISBN: 978-1-4398-3512-8.

Cheng, Z., Portny, J., Smith, J., Mosher, S. (2009). Human Modeling and Monitoring in 4-Dimension, Phase I Final Report, AFRL-RH-WP-TR-2010-0039.

De Aguiar, E., Stoll, C., Theobalt, C., Ahmed, N., Seidel, HP, Thrun, S. (2008). Performance capture from sparse multi-view video, *ACM Transactions on Graphics*.

Hilton, A., Starck, J., Collins, G. (2002). From 3D Shape Capture to Animated Models. *Proceedings of First International Symposium on 3D Data Processing Visualization and Transmission*, 246- 255.

Park, S.I. and Hodgins, J.K. (2006). Capturing and Animating Skin Deformation in Human Motion. *ACM Transaction on Graphics (SIGGRAPH 2006)*, 25(3), 881-889.

Ratwani, K., Orvis, K., and Knerr, B. (2010). An Evaluation of Game-based Training Effectiveness: Context Matters, *Proceedings of I/ITSEC2010*.

Sminchisescu, C. and Telea, A. (2002). Human pose estimation from silhouettes a consistent approach using distance level sets. *WSCG Int. Conf. Computer Graphics, Visualization and Computer Vision*.

Stauffer, C. and Grimson, W. (1999). Adaptive background mixture models for real-time tracking. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol. 2, 246-252.

Seo, H., Cordier, F., Thalmann, N.M. (2003). Synthesizing Animatable Body Models with Parameterized Shape Modifications. *Eurographics /SIGGRAPH Symposium on Computer Animation*.

Terzopoulos, D. (2010). A reality Emulator Featuring Autonomous Virtual Pedestrians and its Application to Distributed Visual Surveillance, *Proc. IEEE Virtual Reality 2008 Conf. (VR 08),* Reno, NV.