A Surveillance System for the Recognition of Intent within Individuals and Crowds

Charles J. Cohen
Cybernet Systems Corporation
727 Airport Blvd.
Ann Arbor, MI 48108
(734) 668-2567
ccohen @ cybernet.com

Frank Morelli
U.S. Army Research Laboratory
Aberdeen Proving Ground, MD 21005
(410) 278-8824
frank.morelli @ us.army.mil

Katherine A. Scott
Cybernet Systems Corporation
727 Airport Blvd.
Ann Arbor, MI 48108
(734) 668-2567
kscott @ cybernet.com

Abstract— *Differentiating between normal human activity and aberrant behavior via closed circuit television cameras is a difficult and fatiguing task. The vigilance required of human observers when engaged in such tasks must remain constant, yet attention falls off dramatically over time. In this paper we propose an automated system to monitor video sensors and tag aberrant human activities for immediate review by human monitors. From the psychological perspective, isolated human motion depicted by point-light walker (PLW) displays have been shown to be salient for recognition of action [17] and determination of emotional state [18]. We propose that by using the motion data that immediately precedes hostile behavior, it may be possible to classify hostile intent before destructive actions take place. These hostile intent gestures can be used to assign individuals a threat assessment level and improve remote sensor monitoring. Such assessments are useful for monitoring human activities and could potentially provide early warning of IED emplacement activities.*

INTRODUCTION

Closed circuit television (CCTV) cameras have become a pervasive component of modern society's arsenal of crime deterrent devices. When employed as preventative devices, CCTVs are only as effective as the human monitor who views the data and orchestrates a response. The human operator in a CCTV system is subject to boredom and fatigue, and usually has to shift his or her attention between a large number of CCTV camera inputs – the combination of fatigue and distraction makes it difficult for human CCTV monitors to detect criminal and terrorist behavior. Additionally, the cues leading up to these behaviors may be subtle, easily escaping human detection during cursory monitoring. Here we detail a software system that assists the CCTV monitor by "flagging" potentially hostile behavior for further scrutiny, in order to increase operational efficiency and potentially save lives.

A system that detects and highlights hostile behavior in CCTV camera data will improve monitoring only if it is able to reliably detect hostile behavior without distracting the human monitor with false-positive results (non-hostile behavior tagged as hostile, see Table 1). For a multiple camera system, false-positives are acceptable so long as they do not distract the user from true hostile behavior – or in other words, the false positives are not hazardous false-positives. Benign false positives could potentially be distracting, but an optimal intent detection system should closely mirror the behavior of a skilled CCTV camera operator, and therefore not present too many distractions.

Table 1: Detection Criteria

Detection Type	Description
True Positive	The system detects hostile actions.
True Negative	The system detects the lack of hostile actions.
False Positive	The system detects hostile actions where none exist.
False Negative	The system fails to detect hostile actions where they do exist.
Hazardous False Positive	The system simultaneously detects a false positive and a false negative on different data sources.
Benign False Positive	The system simultaneously detects false positives and true negatives.

The ability to extract semantic information solely from human biological motion has been well known for some time [14]. In his seminal work, Johansson revealed that presenting coordinated human joint motion was sufficient for rendering the impression of a human being walking or running through space, despite the lack of any associated visual cues such as body form or clothing. Johansson's point-light walker displays (see Figure 1) provided a novel stimulus platform from which a rich body of research has developed, leading to methodologies for examining visual motion perception in isolation – without interference from extraneous visual information. The demonstrated ability to determine human gender [16],[23],[28], recognize action [17] and determine emotional state [18] from motion cues alone illustrates the sensitivity of the human visual system to motion-derived information. Perception of biological motion has also been shown to be a process so robust that even incidental exposure commands the resources of human visual attention [26]. This fundamental, yet malleable, human perceptual ability improves with both exposure to specific motions and action experience [27], [4]. By

978-1-4244-1978-4/08/$25.00 ©2008 IEEE

empirically investigating the component processes involved in perception-based anticipatory inference of human intent, we will attempt to improve "intent-detection" based on behavioral indicators and subtle threat cues.

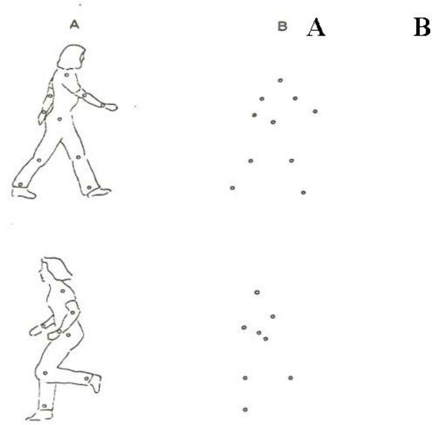

Figure 1. Outline contours of a walking and running subject (A) and the corresponding point-light representations (B) [Johansson, 1973].

With respect to detecting hostile intent, each point in the PLW might have its own "gesture" motion – which when examined in relation to the other links in the object, can be used to determine the overall state of the system. For example, when solely examining the gestures created by the foot and knee joint of a human, one can see that the motions of those features are different depending on whether a person is walking or running.

The gesture recognition module presented here can determine if a dynamic motion is occurring, and using that information with associated kinematic link relationships, develop a hypothesis about the object's overall behavior. Such state recognition is not limited to identifying human (and other animal) motion – a vehicle's state can also be determined by examining the various moving parts, such as the trajectory and tire motions. Even unknown devices (such as mobile robots) could be classified by examining their motion and behavioral characteristics.

In this paper, we demonstrate how to apply a physics-based gesture recognition system (developed to identify human oscillatory hand motions) that can identify human behaviors from data obtained in a CCTV surveillance environment. First, a general kinematic relationship is modeled. Next, specific link combinations are parameterized and modeled, which enables the system to recognize such motions as human gaits (if the links are legs) or the throwing of an object (if the links are arms). Finally, a whole body (human, vehicle, and other) link representation and dynamic model are detailed.

OVERVIEW

The body of this work is divided into four sections that describe the key components and motivations in creating an intent detection system. The first section details the psychology literature on intent detection and outlines our rationale on why such a system is possible, along with our potential data sources. Section II details Cybernet Systems' approach to gesture recognition, and how we will apply that technology to intent detection. The third section details how we intend to extract data about human behavior from live CCTV video for classification by the gesture recognition system described in Section II. The fourth and final section details our conclusions, the applications of our research system, and our future work.

SECTION I: HUMAN PERCEPTION OF INTENT

The U.S. Army Research Laboratory is examining the component processes that underlie human visual perception of biological motion, specifically for determining hostile intent. While human beings have a demonstrated capacity to recognize and classify emotion based on non-verbal communicative cues [8],[9],[12] with evidence to support a degree of universality across cultures [10], the accurate classification of intent based on non-verbal visual cues varies in its reliability. Research plans are underway to examine the role of exposure to specified classes of human biological motion as a modulator for proficiency in rapidly classifying individuals with respect to intent.

While general experience is surely a main factor in the development of such a skill, the specific processes that lead to this type of perceptual capacity are not clear. Focused repetition of elemental concepts and actions is surely critical to learning in general. However, the perceptual learning necessary for the reliable detection of intent partially stands apart from this axiom. Skill development, characterized by perceptual learning, is largely task and stimulus specific, non-transferable to unrelated skills, not subject to explicit knowledge for "how" skill develops, and results in structural and functional neural change [11]. The most important distinction is that perceptual learning is an implicit process often not subject to awareness of the component elements of experiential repetition. Explicit knowledge of the elemental components of a person's mannerisms or style of walk (i.e., gait) after repeated visual exposure is not necessary for development of the ability to recognize that person, for instance, simply by remotely observing them walking at a great distance away. Yet these component elements provide the bottom-up sensory information necessary for the perception and identification of that remote individual as a person whom you know. In the same way, the ability to exhaustively declare the rules of grammar is not a pre-requisite for fluent and grammatical speech, yet exposure to grammatical structure is necessary for grammatical language acquisition and production. It is the eventual aim of this research to explore how mere exposure to human biological motion might modulate intent perception skill development.

This research seeks to specify the sensory and perceptual dynamics of rapid decision-making based on human biological motion cues alone, with the delineation of the dynamic elements of *hostility* in biomotion noted as critical to this effort. The influence of emotional content has been shown to modulate the visual perception of human biological motion – with biomotion sequences featuring "anger" exerting considerable influence over detection performance, even in the absence of whole-body point-light walker displays [5]. This demonstrated sensitivity to point-light renderings of human motion, despite stimulus degradation or based on only partial body component movement [24], underscores the inference-based processing that defines human perception and action – based on ever-changing, often incomplete sensory information evident in the physical environment [21], [22], [13]. The notion that an already robust perceptual sensitivity may be heightened for emotional stimuli [5] is intriguing from the perspective of military application.

By focusing on human anticipatory inference of intentionality, we will examine the human ability to classify threatening actions based solely upon incomplete human motion cues, with a specific focus on human motion characteristics that *lead up* to a violent or threatening action. The goal is a greater understanding of the behavioral tendencies that precede a threatening act, with the hope that such knowledge will provide an advantage to an operator – based on clear perception that leads to rapid, anticipatory, and proactive decisions in order to thwart hostility and prevent battlefield casualties. The perceptual skill necessary for rapid identification and action relative to the precursors of a hostile act, if partially learned prior to deployment, would potentially shorten the learning curve for a soldier new to a battlefield – improving confidence, increasing mission effectiveness and possibly saving lives.

Markerless Motion Capture and Stimulus Presentation

To create a stimulus set that isolates human motion from the cacophony of visual information that typically encompasses a complex scene, video footage of hostile and non-threatening action sequences will be translated into point-light walker (PLW) displays. Video editing software will be utilized in conjunction with customized software based on open-source code generated at the Temple University Vision Laboratory [25] to enable markerless motion-capture for point-light displays. Selected actors will be isolated from video sequences and rendered into point-light walker displays – animations that are devoid of background detail or motion information from alternate objects or individuals that may be present in the original footage. The stimuli will depict isolated point-light renderings of actions ranging from overtly hostile to innocuous. Given that the nature of the depicted actions is a subjective categorization determined by the experimenters, an initial experiment will validate experimenter categorization of thematic action content contained in the displays. While video of action sequences featuring violent or threatening content would surely be rated accordingly, and though human observers are sensitive to the semantic content conveyed by PLW displays, this initial experiment will test stimulus validity by asking subjects to passively observe each display and rate the actions they depict with respect to perceived threat content.

In a subsequent experiment, we will examine human response timing and accuracy when presented with PLW displays such as those detailed above. Subjects will be asked to indicate as quickly and as accurately as possible whether a PLW stimulus depicts a threatening or innocuous event, using a forced-choice paradigm that will record accuracy and response timing. Given that accurate identification of threat within an operational scenario is less useful subsequent to a hostile action than it could be prior to execution, where preventative measures might still be viable, the emphasis for this experiment is the examination of response characteristics that fall within the temporal window preceding the terminal event in an action sequence. To ensure rapid decision-making on the part of the observer that falls within that temporal period, the PLW displays that will be presented to subjects will not feature the final frames that animate a particular goal action. For example, if the PLW rendering depicts an individual placing a mug of coffee on a countertop, the stimulus may show the individual walking toward the counter and executing the preliminary arm and hand motions that are required to set the mug down – but the display will stop short of the final execution of the terminal event, in this case the release of the mug as it settles on the surface of the counter. In this way, observers will not be inclined to wait for the completion of an executed motion before rendering their classification of "threatening" or "innocuous" – since the executed motion they are presented with will never be fully enacted.

Demographic information will be collected to record personal and experiential information that may impact performance data – included will be questions regarding computing, video gaming, as well as living and working experience that relates broadly as possible modifiers for behavioral analysis proficiency. Additionally, the predictive power of efficacy expectations regarding behavior or task performance will be examined by recording data from a Situational Self-Efficacy Scale [1], [2], [3]. For the purposes of this research, it will be used to evaluate response confidence when rating experimental stimuli and providing rapid evaluation of stimulus content. Participants will rate (on a scale of 1 to 10) their level of confidence in their ability to accurately perceive the thematic nature of stimulus content over the course of the stimulus set.

SECTION II: MODELING HUMAN GESTURES FOR BEHAVIOR

In this system, the gesture recognition algorithms are located in an Identification Module – this module uses the position and velocity information provided by the sensor module to identify the gesture. The Identification Module contains a bank of predictor bins (see Figure 2), each containing a dynamic system model with parameters preset to a specific gesture. We assume that the motions of human circular gestures are decoupled in x and y, therefore there are separate predictor bins for the x and y axes (as well as one for z axis when dealing with three dimensional motions). A predictor bin is required for each gesture type for each dimension – the position and velocity information from the sensor module is fed directly into each bin.

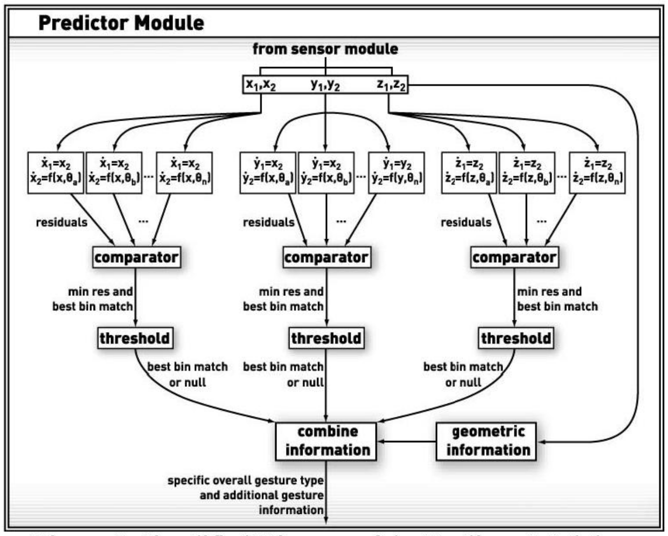

Figure 2: Simplified Diagram of the Predictor Module which Determines Gesture Characteristics.

The idea for seeding each bin with different parameters was inspired by Narendra and Balakrishnan's work on improving the transient response of adaptive control systems. In this work, they created a bank of indirect controllers that were tuned online, but whose identification models had different initial estimates of the plant parameters. When the plant was identified, the bin that best matched that identification supplied a required control strategy for the system [20].

Each bin's model, which has parameters that tune it to a specific gesture, is used to predict the future position and velocity of the motion – the prediction of a particular gesture is made by feeding the current state of the motion into the gesture model. This prediction is compared to the next position and velocity, a residual error is computed, and the bin for each axis with the least residual error is the best gesture match. If the best gesture match is not below a predefined threshold (which is a measure of how much variation from a specific gesture is allowed), then the result is ignored and no gesture is identified. Otherwise, geometric information is used to constrain the gesture further – a single gesture identification number, which represents the combination of the best x bin, the best y bin, and the geometric information, is output to the transformation module. This number (or NULL if no gesture is identified) is output immediately upon the initiation of the gesture, and is continually updated. The parameters used to initially seed each predictor bin can be calculated by feeding the data of each axis from previously categorized motions into the recursive linear least squares identifier.

The identification module contains the majority of the required processing. Compared to most of the systems developed for gesture recognition (for example, see [7], and [19]), the identification module requires relatively little processing time and memory to identify each individual gesture feature.

BEHAVIOR RECOGNITION SYSTEM

Just as the gesture recognition module is built on a bank of predictor bins, the behavior recognition system is composed of a bank of gesture recognition modules. Each module focuses on a specific point of the body (such as features acquired from a PLW system). As that point moves through space, a "gesture" is generated and identified. The combination of gestures from those points is what we define as a motion behavior, which can be categorized and identified. The system, illustrated in Figure 3, details the behavior recognition system – the simplicity of the behavior recognition system is possible because of the strength and utility of the gesture recognition modules.

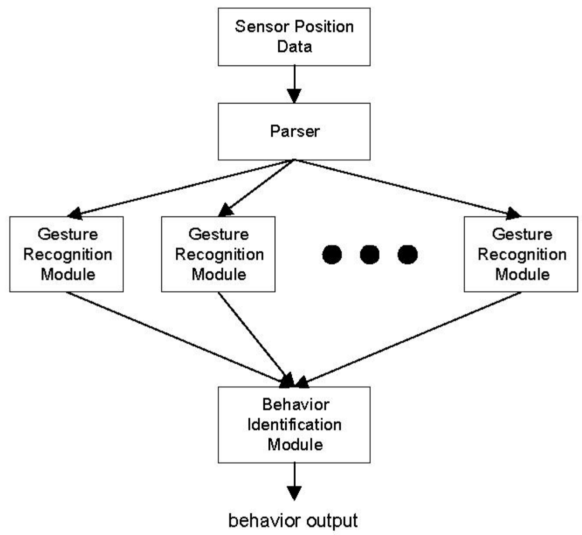

Figure 3: Behavior Recognition System.

OVERALL SYSTEM FLOW

The signal flow proceeds as follows: a user is tagged at various body locations using a series of image processing operations on CCTV video, and the data is acquired as fast as possible (normally at 30 Hz, but a minimum of only 10 Hz is required) and sent to a parser that splits off the data from each specific body location to its own gesture recognition module (GRM) – there is one GRM for each tagged feature. Each GRM outputs the gesture it recognized, if any, to an identification module that matches the gestures to their body location – defining a behavior. If this behavior matches one from a set of predefined behaviors, then this information is outputted.

THE PARSER

In the Parser module, the data time coordinates from each tagged body location (which are input as a stream of consecutive x,y,z), are split up according to body location and sent to an appropriate GRM. This module needs to be changed whenever the input data is of a different format. Runtime variables define how many body parts are being tracked, and therefore the parser uses this information to determine the number of GRM bins and how to split up the data properly.

GESTURE RECOGNITION MODULES (GRMS)

The time and coordinate data from each body feature is used as inputs to an appropriate GRM. Each GRM module is exactly as described earlier, with these modules handling three-dimensional points as a function of time.

BEHAVIOR IDENTIFICATION MODULE

The Behavior Identification Module accepts as inputs gesture types and body locations from the GRMs. Various combinations of gestures at specific body locations are designated as behaviors, and if a match is found the program outputs that match.

SECTION III: CCTV DATA EXTRACTION

The collection of CCTV camera data for our intent detection system is performed at both the system level (i.e. the collection of CCTV cameras at a facility) and at the camera level. The rationale for using a multiple camera system is that it we believe that behaviors exhibited between multiple cameras may yield a secondary source of intent data. For example, multiple cameras make it possible to track an individual moving through a facility or along a road and observe surveillance behaviors that may be indicative of IED emplacement.

SINGLE CAMERA DATA EXTRACTION

At the camera level, human motion data for our gesture recognition system is extracted from CCTV cameras using a series of image processing operations that yield a set of features for processing by our gesture recognition system. For a single CCTV camera the chain of image processing operations is as follows:

1. Capture raw image data.
2. Pre-process, data - resize and flip as necessary.
3. Segment foreground and background data using the codebook method [15].
4. Threshold the foreground imagery into binary connected-component "blobs" or features.
5. Dilate or merge the features and ignore any feature that meets our criteria for noise (e.g. too small, or unrealistic aspect ratio).
6. Perform a rough classification of the binary image components using parameters like size, aspect ratio, position, and color profile to yield a classification confidence pair. (e.g. 0.6 confidence that a feature is a torso).
7. Attempt to associate current frame image features with "similar" features from previous frames. Re-evaluate the feature classification based on previous frame data.
8. Attempt to de-compose large features into smaller sub-features using convex hull detection (e.g. breaking torsos, that include legs and arms, into constituent parts).
9. Feed refined image feature position data and classification into the gesture recognition system mentioned previously.

MULTIPLE CAMERA DATA EXTRACTION

For multiple camera environments, the system consists of multiple camera sensor nodes, software applications, and a computer network (see Figure 4). This modular approach is highly scalable; new sensor nodes and applications can be added as needed. Intelligent pre-processing of data at each sensor node minimizes required network bandwidth. Each sensor node contains a standard functionality suite that includes data capture, data recording, data processing, configuration, and networking (see Figure 4).

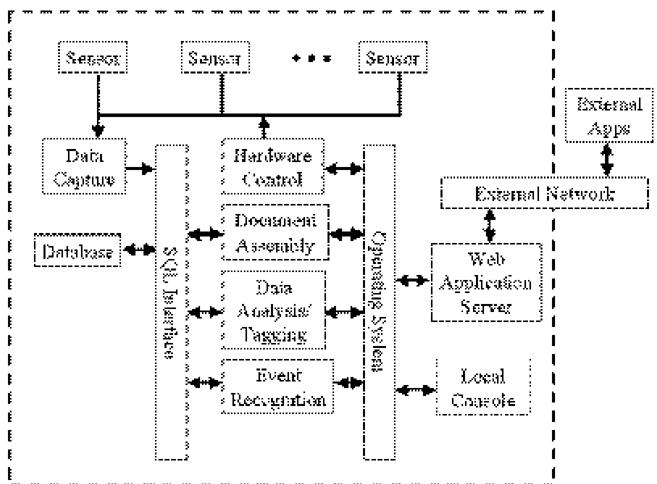

Figure 4: Block Diagram of a Sensor Node

A suite of specific programs, rather than a single monolithic application, controls the configuration of each node. Each of these programs has a set of operational parameters that are stored such that they can be updated in real-time via some API, and is accessible to the database tagging applications. The programs can be stopped and started remotely.

Data captured on a sensor node is stored in a SQL database. SQL Data are divided temporally at some sensible resolution (e.g. 5 or 10 seconds of video data) and stored as "clips" within the SQL database. Each clip can then be tagged with an arbitrary amount of metadata using relational methods. The power of this approach comes from the ability of multiple, independent data processing applications browsing the data that has been captured and then further adding to the metadata. This approach provides the capability to have applications that assemble data streams in to an arbitrary collection that match some specified criteria, and then make those streams available for remote viewing.

The network interface takes advantage of industry standards. We will use IP for all network communications, and HTTP for data transactions wherever possible. By using a web application server such as Apache, we decouple the viewing applications from the actual applications on each node, which allows us to modify either without disrupting the other. This approach also supports the modular addition of new capabilities without changing the infrastructure.

The HTML/CSS forms provide the user interfaces to all programs (configuration, control, data processing, data viewing, etc). This provides the capability for viewing whatever data a node has to offer without needing a priori knowledge of its capabilities; the display application is stored on the sensor node. This approach allows us to leverage existing web browser plug-ins to view video data, receive real-time data updates (via AJAX), and aggregate multiple data sources on one display (using portal technology). At the same time, this does not preclude writing non-web applications – stand-alone applications can still access data using HTTP, then process and display it in whatever manner is needed.

SECTION IV: EXPERIMENTS, FUTURE DIRECTIONS, AND CONCLUSIONS

We performed experiments to test the behavior recognition system. First, a variety of behaviors were performed and served as the baseline for identification. Then these gestures were repeated and the data sent through the system. The behaviors centered on repeated leg and waist motion. The three main types of behaviors performed were walking, running, and jumping in three dimensions.

Rigorous experiments were performed using these behaviors. Originally, a Linear with Offset Component model (2 parameters) was used for gesture/behavior differentiation. However, that low number of parameters representation was not sufficient to capture the richness of motion present in these behaviors, and the system failed. When we instead used a Velocity Damping Terms model for gesture/behavior differentiation, there was clear discrimination between the motions identified by the various identification bins and behavior recognition was possible.

The research and development conducted during this program has significantly advanced the gesture recognition of dynamic gestures and the recognition of human behaviors – it has created the foundation from which advanced behavior recognition systems can be constructed. Though this research and development provides a vertical slice behavior recognition system, it is only a first step. We intend to advance the research to develop a reliable, efficient, and robust behavior recognition system.

In particular, our system's utility would be greatly enhanced if the system could automatically expand its gesture lexicon, and we have begun steps to incorporate such a methodology. This would “teach” the system new gestures rather than having to hard-code additional predictor bins, since the classification scheme cannot identify a motion that does not match any of the currently seeded parameter bins. However, we can make use of on-line free running identification bins [6] to determine parameters for any input motion – if the bank of predictor bins does not recognize a gesture but the identification bin produces parameters that yield consistent prediction results, then the system recognizes the motion as a new gesture. The system would then add a new predictor, seeded with the computed parameters, to the bank of predictor bins. Just as certain voice recognition systems can be trained to recognize specific voice patterns, our system would be able to calibrate itself to recognize user-specific gestures.

We are also exploring the identification of group behaviors, based on combinations of behaviors from one or more sensors in a CCTV environment. Such group behaviors would include the recognition of IED placement activities that require more than one person to bury the device and plant the transmitter or lay and bury the triggering wire.

Any future developments must involve verification of the behavior recognition system. Verification involves both the testing of the algorithms and the use of modules in a completely integrated system. To verify the system, our plan is to:

Test the system with a wide variety of users.

Test the system using a large number of behaviors in our lexicon.

In conjunction with other gesture projects, we will be able to track untagged features – with this ability we will test the system using data from video surveillance cameras.

We strongly believe that these three steps will verify the system, thus facilitating its development into a full system useful in a wide variety of applications.

Check point duty and security tasks represent high-risk missions for soldiers. The increase of asymmetric threats to military and civilian assets throughout the world have generated a critical need for enhanced perimeter security systems, such as automated surveillance systems. Current U.S. Military operations, such as those in Iraq and Afghanistan, have greatly increased the number of U.S. assets in harms way – which in turn has put a strain on our ability to adequately protect all of our facilities. Appropriate perimeter and facility security requires continuous indoor and outdoor monitoring under a wide variety of environmental conditions – typically these locations are complex in that they include open perimeters and high traffic flow that would strain even human observers.

To maximize the effectiveness of military and security personnel, automated real-time methods for tracking, identifying, and predicting the activities of individuals and groups are crucial. The human resources available to such tasks are constantly being reduced, while the prevention of negative actions must be increased. Small numbers of individuals are able to cause a large amount of damage and terror by going after non-military secondary targets that still have an adverse affect on the military’s abilities to fulfill its missions – the system presented here can be used to identify behaviors that would be flagged for observation and evaluation by human monitors based on their observed threat, thereby reducing the tedium and making it possible for one operator to accurately and attentively observe large numbers of CCTV sensors.

REFERENCES

[1] Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*, 191-215.

[2] Bandura, A. (1995). *Self-efficacy in changing societies.* Cambridge, MA: Cambridge University Press.

[3] Bandura, A. (Ed.). (1997). *Self-efficacy: The exercise of control*. NY: Freeman.

[4] Cavanagh, P. (1991). Short-range vs. long-range motion: not a valid distinction. *Spatial Vision*, 5, 303-309.

[5] Chouchourelou, A., Matsuka, T., Harber, K. & Shiffrar, M. (2006). The visual analysis of emotional actions. *Social Neuroscience*, 1(1), 63-74.

[6] Cohen, C. (1996) "Dynamical System Representation, Generation, and Recognition of Basic Oscillatory Motion Gestures and Applications for the Control of Actuated Mechanisms." Ph.D. Dissertation, University of Michigan.

[7] Darrell, Trevor J. and Penland, Alex P. (1993) "Space-Time Gestures." In /IEEE Conference on Vision and Pattern Recognition, /NY, NY.

[8] Ekman, P. (1992). Facial expressions of emotion: New findings, new questions. *Psychological Science*, 3(1), 34-38.

[9] Ekman, P., Friesen, W.V. & Ancoli, S. (1980). Facial Signs of Emotional Experience. *Journal of Personality and Social Psychology*, 39(6), 1125-1134.

[10] Ekman, P. & Friesen, W.V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology*, 17(2), 124-129.

[11] Fahle, M. & Poggio, T. (2002). *Perceptual Learning*. MIT Press, Cambridge, MA.

[12] Frank, M.G., Ekman, P. & Friesen, W.V. (1993). Behavioral markers and recognizability of the smile of enjoyment. *Journal of Personality and Social Psychology*, 64(1), 83-93.

[13] Intraub, H. (1997). The representation of visual scenes. *Trends in Cognitive Sciences*, 1, 217–221.

[14] Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics*, 14 (2), 201-211.

[15] K. Kim (2005) Real-time Foreground-Background Segmentation using Codebook Model, *Real-time Imaging*, Volume 11, Issue 3, Pages 167-256

[16] Kozlowski, L.T. & Cutting, J.E. (1977). Recognizing the sex of a walker from a dynamic point-light display. *Perception & Psychophysics*, 21 (6), 575-580.

[17] Loula, F., Prasad, S., Harber, K. & Shiffrar, M. (2005). Recognizing People From Their Movement. *Journal of Experimental Psychology: Human Perception and Performance,* 31(1), 210–220.

[18] Montepare, J.M., Goldstein, S.B. & Clausen, A. (1987). The identification of emotions from gait information. *Journal of Nonverbal Behavior*, 11(1), 33-42.

[19] Murakami, Kouichi and Taguchi, Hitomi (1991). "Gesture Recognition Using Recurrent Neural Networks." */Journal of the ACM/*, 1(1):237-242.

[20] Narendra, Kumpati S. and Balakrishnan, Jeyendran (1994). "Improving Transient Response of Adaptive Control Systems using Multiple Models and Switching." */IEEE Transactions on Automatic Control/*, 39:1861-1866.

[21] Noë, A. (2005). *Action in Perception*. MIT Press, Cambridge, MA.

[22] O'Regan, J.K. & Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences*, 24(5), 939-973.

[23] Pollick, F.E., Kay, J.W., Heim, K. & Stringer, R. (2005). Gender recognition from point-light walkers. *Journal of Experimental Psychology: Human Perception and Performance*, 31(6), 1247–1265.

[24] Pollick, F.E., Paterson, H.M., Bruderlin, A. & Sanford, A.J. (2001). Perceiving affect from arm movement. *Cognition*, 82, B51-B61.

[25] Shipley, T.F. & Brumberg, J.S. (2003). *Markerless motion-capture for point-light displays*. Technical Report, Temple University Vision Laboratory, Philadelphia, PA.

[26] Thornton, I.M. & Vuong, Q.C. (2004). Incidental Processing of Biological Motion. *Current Biology*, 14, 1084-1089.

[27] Thornton, I.M., Rensink, R.A. & Shiffrar, M. (2002). Active versus passive processing of biological motion. *Perception*, 31, 837-853.

[28] Troje, N.F. (2002). Decomposing biological motion: A framework for analysis and synthesis of human gait patterns. Journal of Vision, 2, 371-387.