DOI: 10.2352/ISSN.2470-1173.2016.11.IMAWM-471

Deep network for analyzing gait patterns in low resolution video towards threat identification.

Binu M Nair and Kimberly D Kendricks
University of Nevada Las Vegas, Nevada,USA

Abstract

We propose a deep regression-based neural network to analyze gait of individuals in low resolution video for real-time detection of abnormal activity/threats in surveillance applications. In todays commercial setting, extracting such gait patterns require motion capture devices such as Kinect or VICON, and is restricted to indoor and controlled scenarios such as in gaming applications or clinical studies. The network is trained by incorporating an inverse kinematic Groebner-based model to estimate the body joint angles from the pose. These angle trajectories of the upper and lower extremities of the body serve as gait signatures for identifying threat patterns. The first few layers model the relationship between motion and image features of the individual using a deep belief network. The next few layers model the relationship between the latent features generated from deep belief nets and the inverse kinematic model using a regression-based deep network. This network characterizes the relationship between the low-level image/motion features and the kinematics associated with the movement of an individual. The estimated joint angle trajectories are then classified as threats (person wearing a loaded vest) and non-threats (person without any load on body) using a K-Nearest Neighbor classifier. Experimental results on the INSPIRE dataset released by Air force institute of technology and its analysis show the effectiveness of deep learning concepts for gait analysis.

Introduction

Gait biometrics have attained wide recognition in research as suitable ways of evaluating motion of individuals for person identification [29]. Furthermore, gait is the one of the key biometrics capable of executing such assessments at a long distance, especially in behavioral analysis for homeland security applications [17]. To contribute to national security efforts in identifying threats, this research work is aimed at analyzing gait patterns for low resolution video surveillance. This work is focused on the computation and evaluation of specific body joint angles as signatures for real-time detection of abnormal gait patterns. Our work is based on a previous research study on the influences of load such as an improvised explosive device on the gait of the individual to determine threats [26].

The aim of this project is to analyze the gait of individuals in low resolution video surveillance footage by evaluating the change of his/her pose as a result of added load on the body. A sample of the video scene and the set of body joints to be tracked is illustrated in Figure 1. Our earlier work proposed an inverse kinematic model using Groebner basis theory to analyze an individual arm/leg swing movements [16, 15]. This work was extended for developing mathematical models to extract gait signatures from motion capture data for analyzing the effect of load on gait cycles of an individual [1]. However, this gait analysis required very accurate body joint locations of the individual (pose) and therefore, constrained it to indoor and controlled scenarios such as in gaming applications or clinical studies. Such accurate pose can be estimated on high resolution images and video using methods such as Deep Pose [27],[10] and articulated parts model [28]. However, most low-cost surveillance video feeds are of low resolution. The pose estimates using these methods are noisy and not suitable for the application of mathematical models for gait signature extraction.

Figure 1: Illustration of specific joints to be tracked and evaluated for gait analysis

In our previously published work [22, 24, 23], we developed methodologies where we can obtain close approximation to the continuous body joint trajectories by using the discrete pose estimates as a prior. Optical flow, along with HOG [6] and LBP [25] descriptors were used to track the body joints and evaluate its position at each frame [22]. In our other work, we developed the Improved Region-based Kalman filter to obtain precise estimates of the body joint locations or pose on low resolution video [24, 23]. This work considered the displacements in the joint position trajectories as a signature for threat classification. We also investigated the effect of a tracking scheme in obtaining continuous position trajectories and compared these with discrete pose estimates to determine threat classification accuracy. But, these

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.11.IMAWM-471

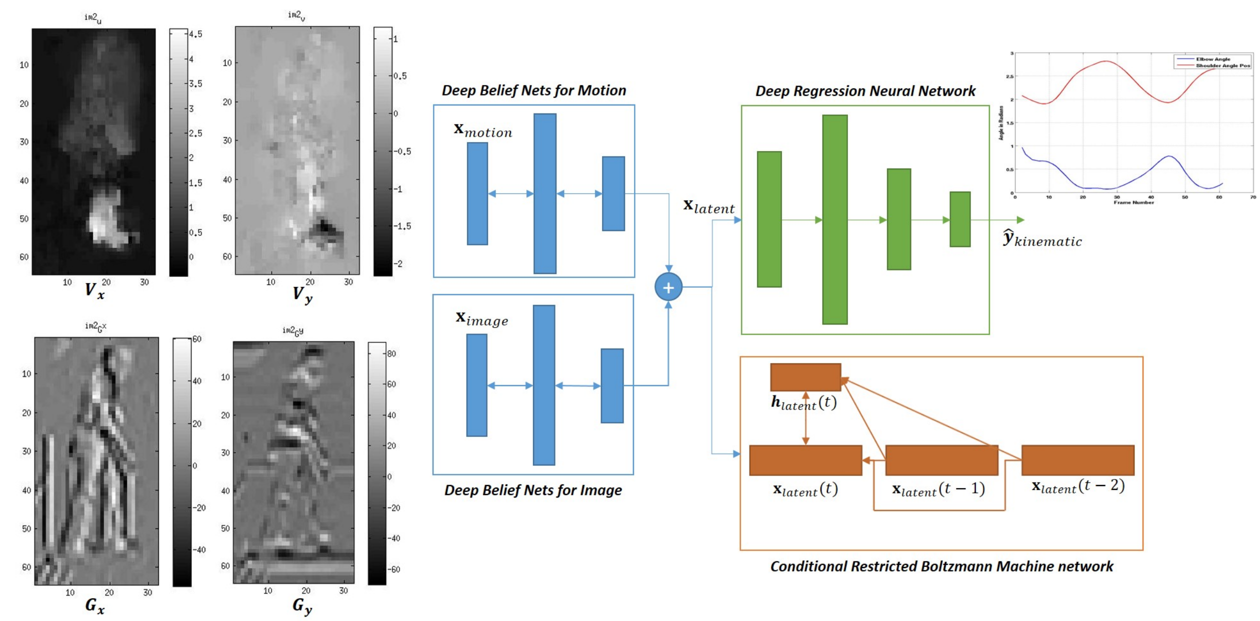


Figure 2: Deep network for analyzing gait patterns.

methods depended on the accuracy and representational power of the various region descriptors. So, an automated feature extraction mechanism tuned towards gait-based threat identification is required.

Our investigation into the gait-based threat identification problem leads to the development of a deep neural network (shown in Figure 2) which can extract motion/shape latent features and model its relationship to an inverse kinematic model for estimating gait signatures. Our novelty lies in the development and usage of a deep network to represent human pose by incorporating an inverse kinematic model of true human locomotion.

Deep architecture for gait analysis

We propose a deep belief network [14] for latent feature extraction and a deep regression-based neural network which can model the relationship between the latent features and the inverse kinematic model. We also propose the idea of a temporal learning mechanism such as conditional restricted Boltzmann machine (CRBM) to investigate the possibility of tracking gait patterns. The use of CRBM is to be considered as part of the future work and will not be investigated in this manuscript. The complete deep neural network is shown in Figure 2.

Feature extraction using deep belief nets

The first module in our proposed network is the feature extraction from motion and image features using Deep Belief nets [14]. Since we are dealing with low resolution, we directly use the optical flow field and the image gradient at a particular instant. Motion components are obtained by computing the optical flow in the x and y directions using the warping technique by Brox et al. [4] between two consecutive frames. This technique is more suited for extracting motion components, as it provides much more smoother motion flow field with few mismatches compared to other optical flow methods such as Lucas-Kanade [21, 3] and Farneback [9]. Image components are computed from the image gradients in x and y directions. As shown in Figure 2, the motion components are represented by flow velocities (V_x, V_y) and image components are represented by gradients (G_x, G_y). We convert these components into vectors $\mathbf{v} = (V_x(:)V_y(:))$ and $\mathbf{g} = (G_x(:)G_y(:))$ and perform PCA-based whitening to decorrelate the pixels. Then, for the motion vectors $\mathbf{v}$ and image vectors $\mathbf{g}$, we train a separate deep belief network. In other words, we train a pair of deep belief networks, one for computing motion latent features and the other for computing image latent features. These two deep belief nets are independent of each other where each deep belief network is a set of stacked RBMs. To understand how latent features are computed, we first give an overview of the RBM.

RBM overview

Restricted Boltzmann Machines (RBM) are generative neural networks which model the relationship between the set of features and its latent representations. The basic RBM network contains two layers, the visible layer $V = \mathbf{v} \in \mathbb{R}^{N_V}$ and the hidden layer $H = \mathbf{h} \in \mathbb{R}^{N_H}$, each containing a set of binary units with only inter-layer connections through weights $W \in \mathbb{R}^{N_H \times N_V}$ and no intra-layer connections. The energy of the RBM $E(V,H)$ can be defined as given in Equation 1.

$$Energy; E(V,H) = (\mathbf{v}^T W^T \mathbf{h} + \mathbf{v}^T \mathbf{bv} + \mathbf{h}^T \mathbf{bh}) \tag{1}$$

The probability of a setting the state of a single hidden unit state to 1 when the states of the visible units are set can be defined as $P(h_j = 1|V = \mathbf{v})$. Similarly, the probability of setting the state of a single visible unit state to 1 when the states of the hidden units are set can be defined as $P(v_i = 1|H = \mathbf{h})$. These are given in Equations 3 and 2 where they can be derived from the joint distribution given in Equation 1. The training of the RBM is then to attain the states having minimum energy and this corresponds to optimizing the weights W using the Contrastive Divergence (CD) technique proposed by Hinton [13]. It follows the Gibbs sampling [5] or Markov Chain Monte Carlo sampling (MCMC) [7]

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.11.IMAWM-471

for obtaining the hidden and visible states in an iterative manner.

$$P(h_j = 1|V = \mathbf{v}) = \frac{1}{(1+\exp(-(bh_j + W_{j,:}\cdot\mathbf{v})))} \quad (2)$$

$$P(v_i = 1|H = \mathbf{h}) = \frac{1}{(1+\exp(-(bv_i + (W_{:,i})^T\cdot\mathbf{h})))} \quad (3)$$

Use of RBM in deep belief nets

Since the RBM models the joint distribution and obtains optimal weights between any pair of layers (visible and hidden layer), it has been widely used in deep multi-layer neural networks with more than one hidden layer. Such networks are very difficult to train using the traditional back-propagation algorithm. The RBMs provide a way to obtain optimized weights between a pair of consecutive layers, one pair at a time, in a deep neural network [2]. Once these pre-trained weights for a pair of layers are obtained, these can be fine-tuned using the error-back propagation algorithm to get the final optimized weights of the network [12, 18]. This particular property of the RBM now enables us to build a deep neural network which can model layers of hidden/latent feature representations. Some examples of such neural networks are convolutional networks [19, 2], Stacked autoencoders [12] and deep belief nets [14, 20]. Therefore, in this approach, we construct a deep belief net by stacking RBMs for the motion features and image features independently. After pre-training of the layers, the motion and image vectors are passed through the respective deep belief networks to obtain motion and image latent features.

Kinematic model representation

The kinematic model employed in this network is a Groebner-based inverse kinematic model [16] which has been evaluated for gait analysis in clinical studies [1]. A 2D joint angle configuration using this kinematic model in the saggital plane is shown in Figure 3. According to this configuration, the inverse kinematic model can be divided into two portions: the upper extremity and the lower extremity. The upper extremity is related to the lengths of the upper/fore arms and the body joint angles of the wrist/elbow with respect to the shoulder joint. The lower extremity is related to the lengths of the thigh/legs and the body joint angles of the knee and ankle with respect to the hip joint. In our earlier work [23], we developed a tracking mechanism to obtain precise body joint position trajectories (given as x, y coordinates) from the pose. From these set of joint position trajectories, we can estimate the relative joint angles by solving the inverse kinematic problem. The solution to the inverse kinematic problem in human motion analysis can be stated as finding relative angles between the joints given the length of segments and the reference joint position. The trajectory formed by these relative body joint angles across time contains the necessary gait signatures which can be classified for gait-based threat identification.

The wrist is considered as the end-effector in the upper extremity model. Solving the forward kinematic problem corresponds to finding the location of the wrist $(\hat{x}_w, \hat{y}_w)$ given the segment lengths (L_{SE}, L_{EW}) and joint angles (θ_S, θ_E). Therefore, the inverse kinematic problem is defined as finding the estimate of the joint angles $\hat{\theta}_S, \hat{\theta}_E$ at instant t, given the segments lengths and the wrist position (x_w, y_w). Similarly, for the lower extremity model,

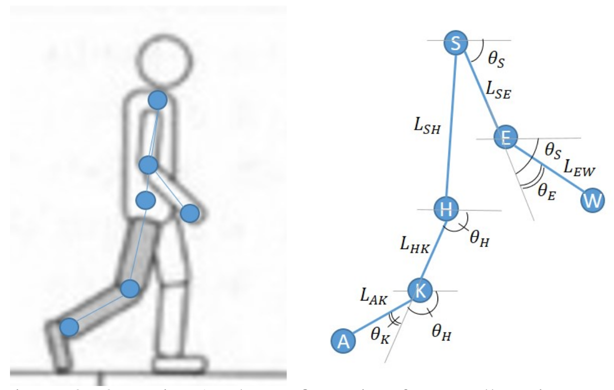


Figure 3: 2D Joint Angle configuration for a walk action at an instant t

the inverse kinematic problem is defined as finding the estimate of the joint angles $\hat{\theta}_H, \hat{\theta}_K$ at instant t, given the segments lengths (L_{HK}, L_{KA}) and the ankle position (x_a, y_a). By considering the human pose as a kinematic chain with the wrist and the foot as the end-effectors, a set of non-zero polynomial equations can be derived from the resulting joint geometry at instant t. The system of equations characterizing the upper and lower extermity models are given in Equation 4 and Equation 5. Here, we use the notations as $c_l \backsim \cos(\theta_l)$, $s_l \backsim \sin(\theta_l)$ where $l \in \{S, E, H, K\}$ which refers to the shouder, elbow, hip and knee respectively.

$$\begin{aligned}
f_1 &:= \tilde{L}_{SE}c_E + \tilde{L}_{EW}c_S c_E + \tilde{L}_{EW}s_S s_E - \tilde{x}_W \\
f_2 &:= \tilde{L}_{SE}s_E + \tilde{L}_{EW}s_S c_E - \tilde{L}_{EW}c_S s_E - \tilde{y}_W \\
f_3 &:= c_E^2 + s_E^2 - 1 \\
f_4 &:= c_S^2 + s_S^2 - 1
\end{aligned} \quad (4)$$

$$\begin{aligned}
f_1 &:= \tilde{L}_{HK}c_H + \tilde{L}_{KA}c_H c_K - \tilde{L}_{KA}s_H s_K - \tilde{x}_A \\
f_2 &:= \tilde{L}_{HK}s_H + \tilde{L}_{KA}s_H c_K + \tilde{L}_{KA}c_H s_K - \tilde{y}_A \\
f_3 &:= c_H^2 + s_H^2 - 1 \\
f_4 &:= c_K^2 + s_K^2 - 1
\end{aligned} \quad (5)$$

Equation 4 is a set of 4 polynomial equations with unknown variables c_E, s_E, c_S, s_S and coefficients $\tilde{L}_{SE}, \tilde{L}_{EW}, \tilde{x}_W, \tilde{y}_W$. By converting this set to a grobner basis, this set of equations can be solved easily for the unknowns at instant t. The groebner basis theory is an algorithm which has been used widely for solving the inverse kinematic problems for robotic manipulators where by knowing the position and orientation of the end-effector, the various joint parameters can be computed. Similarly in the case of the lower extremity model, the system of equations (Equation 5 can be converted to a Grobner basis and the following unknowns can be solved easily. Thus, we get joint angle trajectory at every frame t given as $\tilde{\theta}_l(t)$ where $l \in \{S, E, H, K\}$. In this analysis, we only consider the upper extremity model where the parameters are given by $(x_w, y_w, L_{SE}, L_{EW}, \Theta_S, \Theta_E)$. By apply groebner basis on the above system of equation, the solution to the inverse kinematic model obtained is given by the Equations 6 - 9. Then the joint angles Θ_S and Θ_E can easily be computed and the inverse kinematic vector can be obtained.

$$\cos(\Theta_S) = \frac{x_w^2 + y_w^2 - L_{SE}^2 - L_{EW}^2}{2\cdot L_{SE}\cdot L_{EW}} \quad (6)$$

DOI: 10.2352/ISSN.2470-1173.2016.11.IMAWM-471

$$\sin(\Theta_S) = \frac{-(L_{SE}^4 + L_{EW}^4 + x_w^4 + y_w^4 + 2 \cdot (x_w^2 \cdot y_w^2 - L_{SE}^2 \cdot L_{EW}^2)}{4 \cdot L_{SE}^2 \cdot L_{EW}^2} + \frac{(L_{SE}^2 + L_{EW}^2) \cdot (y_W^2 + .x_W^2))}{4 \cdot L_{SE}^2 \cdot L_{EW}^2} \quad (7)$$

$$\cos(\Theta_E) = \frac{L_{SE} \cdot x_w}{x_w^2 + y_w^2} \pm \frac{y_w(L_{EW}^2 - L_{SE}^2 - (x_w^2 + y_w^2))}{2 \cdot L_{SE} \cdot (x_w^2 + y_w^2)} \quad (8)$$

$$\sin(\Theta_E) = \frac{L_{SE} \cdot y_w}{x_w^2 + y_w^2} \pm \frac{x_w(L_{EW}^2 - L_{SE}^2 - (x_w^2 + y_w^2))}{2 \cdot L_{SE} \cdot (x_w^2 + y_w^2)} \quad (9)$$

Inverse kinematics-based deep neural network

As mentioned earlier, the usage of inverse kinematics in estimating pose and extracting gait patterns is restricted to indoor scenarios. Application of any inverse kinematic model would require accurate pose estimation and this is not possible in a low resolution surveillance footage. It is essential to model the relationship between the motion/image features from video to the inverse kinematic model so that efficient gait signatures can be computed.

We train a deep regression-based neural network to model the relationship between the latent features computed from the deep belief network and the inverse kinematic model parameters. Since the kinematic parameters are of continuous nature, the modeling of the relationship corresponds to a regression problem. Therefore, we use the mean squared error as the objective function to be minimized in the error-back propagation learning algorithm. We also use rectified linear units [11] for faster learning and guaranteed convergence. The algorithm to train the proposed deep neural network is given in Algorithm 1. Similarly, the algorithm to infer the kinematic model vector, to compute the gait patterns and classify them as threat is given in Algorithm 2.

Algorithm 1 Learning

```
Compute optical flow field and gradient at frame t
Apply person detector to obtain local region of interest.
Obtain motion/image vectors from optical flow/gradient within local region.
Using algorithm proposed in [23], compute the pose.
Solve the inverse kinematic model to get the kinematic vector.
Normalize the motion and image vectors.
for l = 1 : L do
    Train RBM layer l of motion deep belief net.
    Forward data to next layer l + 1
end for
for l = 1 : L do
    Train RBM layer l of image deep belief net.
    Forward data to next layer l + 1
end for
Compute the latent features using trained deep belief nets.
Set latent features as regressors.
Set kinematic model vector as response variables.
Train a deep regression neural network.
Estimate kinematic model parameters from training data using the deep network.
Compute and store the trajectory descriptor [23] from the estimated kinematic model parameters.
```

Algorithm 2 Inference

```
Compute optical flow field and gradient at frame t
Apply person detector to obtain local region of interest.
Obtain motion/image test vectors from optical flow/gradient within local region.
Normalize the motion and image test vectors.
for l = 1 : L do
    Apply RBM layer l of motion deep belief net.
    Forward data to next layer l + 1
end for
for l = 1 : L do
    Apply RBM layer l of image deep belief net.
    Forward data to next layer l + 1
end for
Concatenate the motion and image latent features.
Estimate the kinematic model vector from the trained deep regression neural network.
Compute the trajectory descriptor on the kinematic model parameters.
Using K-Nearest Neighbor, classify the trajectory descriptors as threat or no threat.
```

Analysis

The proposed tracking scheme was tested on a private dataset by the Air Force Institute of Technology (AFIT) at Wright Patterson Air Force Base, OH. It consisted of the walking activity of 12 individuals on an outdoor circular track that included flat grass, a staircase attached to a raised platform, and a down ramp. The subjects walked clockwise and counterclockwise around the track to collect a total of 100 video sequences in each direction. Subjects wore civilian clothes and shoes during the experiment. The video sequences were captured using two Canon GL cameras that faced the track at a distance of 50 ft. Each sequence was divided into 5 phases A, B, C, D, and E, a sequence of each phase was selected from either the left camera or right camera depending on what part of the track the walking activity was taking place. For the purposes of our analysis, we don't consider which camera the sequence was shot from. There are two variations of the sequences; one which contains an individual wearing a weighted vest of around 3-5 Kg and the other the same individual not wearing this heavy weighted vest. We perform our analysis on Phase A,C, and E of the sequence which simulates different walking conditions. We exclude the phases B and D since the available data are erroneous for analysis. The frames of interest are of the subject walking on the cross over platform.

Validation of inverse kinematic model

We validate the use of the inverse kinematic model based on Groebner basis for computing gait features from surveillance video. Using the improved region-based Kalman filter (IRKF) [23, 24] developed by us, we can obtain body joint position trajectories or pose at each frame. The tracking scheme employed different types of descriptors in the analysis such as Histogram of Oriented Gradients (HOG), Local Binary Patterns (LBP), SIFT, SURF, BRIEF, BRISK and ORB. The IRKF estimates a set of body joint location coordinates which varies with the type of de-

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.11.IMAWM-471

Θ (Joint Angle)	PLS	HOG	LBP	SIFT	SURF	BRIEF	BRISK	ORB
Θ_S	54.95	52.25	52.25	51.35	64.86	59.45	52.25	54.95
Θ_E^+	**64.86**	53.15	62.16	60.36	58.10	**68.46**	60.36	54.95
Θ_E^-	53.15	53.15	60.36	57.65	55.85	53.15	58.10	**64.86**
Θ_S Θ_E^+	54.95	54.05	60.36	56.75	**69.36**	**66.66**	60.36	54.50
Θ_S Θ_E^-	54.95	53.15	54.95	54.05	55.85	55.85	53.15	55.85
Θ_S Θ_E^+ Θ_E^-	54.95	56.30	**63.06**	53.15	58.55	59.45	60.36	52.25

(a) Phase A

Θ (Joint Angle)	PLS	HOG	LBP	SIFT	SURF	BRIEF	BRISK	ORB
Θ_S	53.57	50	**71.42**	64.28	53.57	50	50	50
Θ_E^+	57.14	57.14	**64.28**	53.57	57.14	46.42	62.50	50
Θ_E^-	60.71	50	53.57	53.57	57.14	53.57	62.5	60.71
Θ_S Θ_E^+	53.57	60.71	**67.85**	53.57	60.71	53.57	60.71	53.57
Θ_S Θ_E^-	50	46.42	57.14	57.14	50	53.57	57.14	57.14
Θ_S Θ_E^+ Θ_E^-	50	57.14	57.14	53.57	60.71	60.71	60.71	53.57

(b) Phase C

Θ (Joint Angle)	PLS	HOG	LBP	SIFT	SURF	BRIEF	BRISK	ORB
Θ_S	61.53	61.53	66.82	61.53	61.53	**70.19**	61.53	61.53
Θ_E^+	61.53	61.53	66.34	64.42	**68.26**	65.38	64.42	**68.26**
Θ_E^-	**64.42**	61.53	63.46	62.5	60.57	61.53	61.53	61.53
Θ_S Θ_E^+	61.53	59.61	64.42	64.42	67.30	64.42	64.42	**69.23**
Θ_S Θ_E^-	61.53	61.53	**68.26**	61.53	60.57	67.30	67.30	61.53
Θ_S Θ_E^+ Θ_E^-	61.53	60.57	67.30	65.38	**68.26**	65.38	64.42	**68.26**

(c) Phase E

Table 1: Classification results obtained using different combinations of joint angle trajectories for the three phases A, C and E. Phase B and D is not included as the discrete pose estimates provided by PLS are errorneous.

scriptor used in the tracking scheme. Therefore, for each descriptor, the estimated kinematic model parameters computed using the inverse kinematic Groebner-based model also varies. Trajectory descriptors are computed from the body joint angles and classified as a threat/non-threat using a kernal SVM. In Table 1, we provide the accuracies in distinguishing threats for a specific combination of the joint angles and a single descriptor.

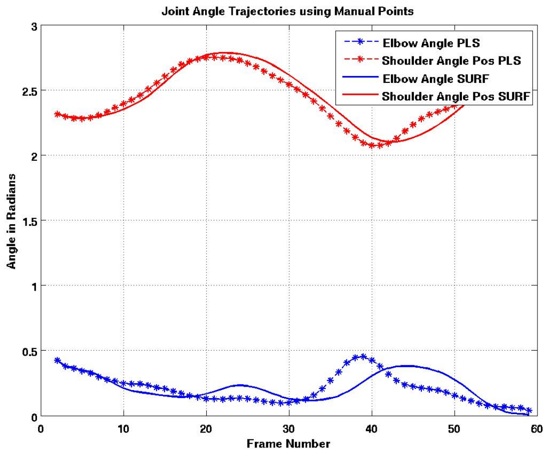
In phase A, the positive solution of the elbow joint angle Θ_E^+ gives little better accuracy of 64% using the discrete body joint positions as provided by the Point Light Software (PLS). Selection of other combinations of the joint angles computed from PLS provide no better than 55%. When using the region-based tracking with other descriptor combinations, we find that the SURF and BRIEF descriptors give much better accuracy of 69% and 68%. However, only the Θ_S and Θ_E^+ seem to have any significant effect on the accuracy. The negative solution Θ_E^- seems to have no effect except for when using the ORB feature descriptor which gets an accuracy of 64.86%. We can visually evaluate the body joint angle trajectories for some subjects in Phase A to determine the correlation between the tracked and the discrete trajectories. An illustration of the body joint angle trajectories computed from SURF descriptor-based IRKF and the discrete pose is shown in Figure 4. From this analysis, we found that the accuracy can be related to the smoothness of the joint angle curve, the amplitude of the curve and the closeness to the true sinusoidal trajectory.

We also see that the LBP obtains a high accuracy of 71.42% using only the joint angle Θ_S in phase C. This is a big improvement with 18% more than the accuracy obtained with the discrete pose. We see the same improvement with Θ_E^+ joint angle. However, the negative solution Θ_E^- gives poor accuracy and no improvement. Thus, with the LBP descriptor and using the configurations Θ_S and (Θ_S, Θ_E^+) has better accuracy and bigger improvement over the PLS method. To summarize, it is shown that the joint angle configuration (Θ_S, Θ_E^+) is more effective that the others. The other configurations obtain accuracies of $68\% - 69\%$ but this depends on the scenerio and the descriptor. From this analysis, we can assume that some sort of weighting factor or function should be used in the selection of the appropriate descriptor and the joint angle configuration. This weighting factor should reflect the scenario through contextual descriptors in the scene or using pre-trained models to detect a specific scenario based on the persons movement. The deep neural network proposed by us in this manuscript helps in obtaining specific weighting functions (weights of the network) which can extract optimal latent features and correlate it with the inverse kinematic model parameters.

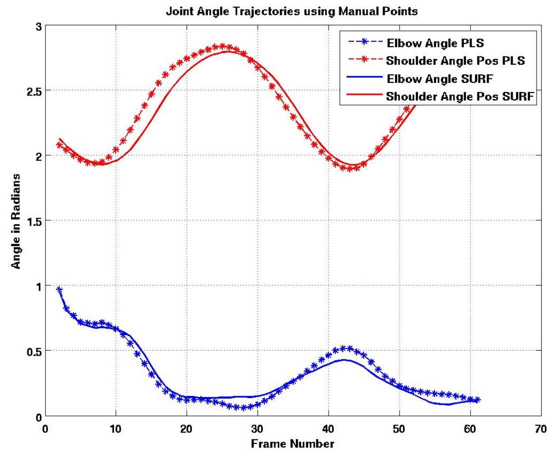
Validation of the deep neural network for inverse kinematic modeling

In the proposed deep learning neural network, we use a 2 layered architecture for the deep belief net based feature extraction and a 3-layered architecture for the deep regression neural network. The sizes of the deep belief nets for both motion and image is [2048 512] with the input size being 4096. Since the inputs to the deep regression neural network is the concatenated motion and image latent features, we set the sizes as [2048 256 6] with the input size being 1024 and the output being 6. For the

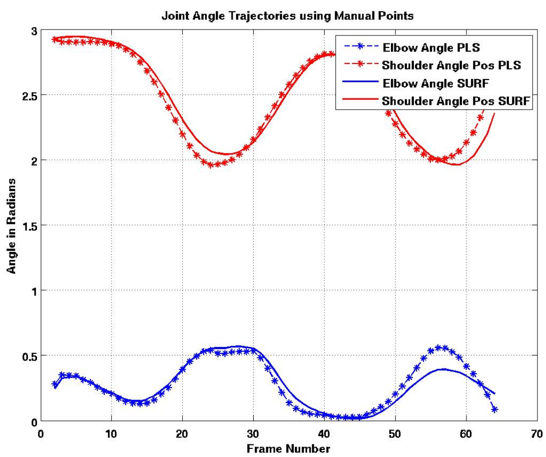
©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.11.IMAWM-471



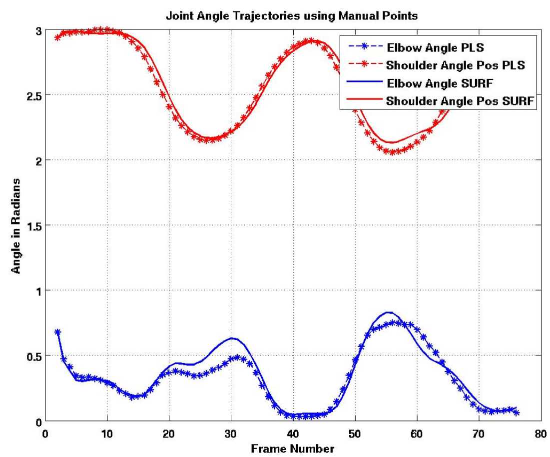
(a) Subject 11 with loaded vest.



(b) Subject 11 with no load.



(c) Subject 3 with loaded vest.



(d) Subject 3 with no load.

Figure 4: Comparison of body joint angle trajectories computed from SURF-based IRKF [23] method and the PLS method for Phase A sequences

entire network, we set the learning rate at 0.01. We train the network on a randomly selected $\frac{9}{10}$ of the total number of samples and test on the remaining $\frac{1}{10}$. We repeat this sampling 20 times and average the accuracies obtained. After estimating the inverse kinematic model parameters, we use the trajectory descriptor to describe each body joint angle trajectory. By using a k- nearest neighbor classifier from the YAEL library [8], we classify the trajectory descriptors of a video sequence as a threat or non threat.

To train the deep regression neural network, we used different sets of joint angle trajectories, each set computed using a particular method (different descriptor in the improved region-based Kalman filter). We also used the median (or average) value of the body joint angles across the different methods for every frame of every video sequence. In addition, we accumulate all of the video sequences from phases A,C and E for training the network. Using $K = 10$ in the $K-NN$ classifier, the highest accuracy we obtained in threat classification was 60.36%. This was using the median joint angle trajectories as the response variable in the deep regression network. The next closest accuracy obtained was 59.36% under the training scheme of using the joint angle trajectories computed from SURF-based IRKF method. However, by applying the K-NN classifier on the joint angle trajectories from tracked pose without the proposed neural network, we obtain only 52%. This is in contrast as the estimated trajectories should provide less accuracy that the ones used for training the algorithm. But, note that the trajectories used for training are noisy. So, during the optimization (learning phase), the network not only learns the underlying distribution of the latent features but also models its relationship to the inverse kinematic model. This modeled relation in the form of weights are not strictly tied to the training set, and therefore, provides better discriminative inverse kinematic model parameters. This establishes the fact that deep neural networks are not only used for classification based on training data but also provides better discriminative features which spans multiple types of scenes.

Conclusions and Future Work

In this manuscript, we extended our work done in the area of threat identification from low resolution video sequences. Our previous work used improved region-based Kalman filter (IRKF) to track the pose and obtain continuous body joint trajectories. In this work, we used an inverse kinematic model based on Groebner basis to compute body joint angles in the upper extremity from video sequences. These body joint angles correspond to the gait patterns needed for this problem. By evaluating the various descriptors in the IRKF, we validated this inverse kinematic model for classifying threats based on the change in the body joint angle trajectories. To investigate the use of the inverse kinematic model without depending on too much on the tracking mechanism, we developed a deep neural network, which extracts motion and image features using deep belief nets. These features are then correlated to the inverse kinematic model using a deep regression neural network with rectilinear units. We proved that by using the proposed deep neural network, we not only obtained a good classification accuracy of 60.32% but also provided an insight into its capability of having an inverse kinematic model influence the optimization of the weights. Our future work will involve the use of Conditional Restricted Boltzmann machines to infer temporal gait signatures for future prediction of a threat from streaming low resolution surveillance videos.

References

[1] Anum Barki, Kimberly Kendricks, Ronald F. Tuttle, David J. Bunker, and Christoph C. Borel. Extraction of human gait signatures: an inverse kinematic approach using groebner basis theory applied to gait cycle analysis. volume 8734, pages 87340L–87340L–14, 2013.

[2] Yoshua Bengio. Learning deep architectures for ai. *Found. Trends Mach. Learn.*, 2(1):1–127, jan 2009.

[3] Jean-Yves Bouguet. Pyramidal implementation of the lucas kanade feature tracker description of the algorithm, 2000.

[4] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert. High accuracy optical flow estimation based on a theory for warping. In *European Conference on Computer Vision (ECCV)*, volume 3024 of *Lecture Notes in Computer Science*, pages 25–36. Springer, May 2004.

[5] George Casella and Edward I. George. Explaining the Gibbs Sampler. *American Statistician*, 46:167–174, 1992.

[6] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 886–893 vol. 1, 2005.

[7] Persi Diaconis. The Markov chain Monte Carlo revolution. *Bulletin of the American Mathematical Society*, 46(2):179–205, November 2008.

[8] Matthijs Douze and Hervé Jégou. The yael library. In *Proceedings of the 22Nd ACM International Conference on Multimedia*, MM '14, pages 687–690, New York, NY, USA, 2014. ACM.

[9] Gunnar Farnebck. Two-frame motion estimation based on polynomial expansion. In Josef Bigun and Tomas Gustavsson, editors, *Image Analysis*, volume 2749 of *Lecture Notes in Computer Science*, pages 363–370. Springer Berlin Heidelberg, 2003.

[10] V. Ferrari, M. Marin-Jimenez, and A. Zisserman. Progressive search space reduction for human pose estimation. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8, June 2008.

[11] Xavier Glorot, Antoine Bordes, and Yoshua Bengio. Deep sparse rectifier neural networks. In Geoffrey J. Gordon and David B. Dunson, editors, *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics (AISTATS-11)*, volume 15, pages 315–323. Journal of Machine Learning Research - Workshop and Conference Proceedings, 2011.

[12] G. E. Hinton and R. R. Salakhutdinov. Reducing the dimensionality of data with neural networks. 313(5786):504–507, 2006.

[13] Geoffrey E. Hinton. Training products of experts by minimizing contrastive divergence. *Neural Comput.*, 14(8):1771–1800, August 2002.

[14] Geoffrey E. Hinton, Simon Osindero, and Yee-Whye Teh. A fast learning algorithm for deep belief nets. *Neural Comput.*, 18(7):1527–1554, July 2006.

[15] Kimberly D. Kendricks. A kinematic analysis of the gmf a-510 robot: An introduction and application of groebner basis theory. *Journal of Interdisciplinary Mathematics*, 16(2-03):147–169, 2013.

[16] Kimberly D. Kendricks, Adam M. Fullenkamp, Robert Mcgrellis, Jonathan Juhl, Ronald F. Tuttle, Jonathan Juhl, and Ronald F. Tuttle. An inverse kinematic mathematical model using groebner basis theory for arm swing movement in the gait cycle, 2010.

[17] T. Ko. A survey on behavior analysis in video surveillance for homeland security applications. In *Applied Imagery Pattern Recognition Workshop, 2008. AIPR '08. 37th IEEE*, pages 1–8, Oct 2008.

[18] Hugo Larochelle and Yoshua Bengio. Classification using discriminative restricted boltzmann machines. In *In ICML 08: Proceedings of the 25th international conference on Machine learning. ACM*, 2008.

[19] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, Nov 1998.

[20] Honglak Lee, Roger Grosse, Rajesh Ranganath, and Andrew Y. Ng. Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. In *Proceedings of the 26th Annual International Conference on Machine Learning*, ICML '09, pages 609–616, New York, NY, USA, 2009. ACM.

[21] Bruce D. Lucas and Takeo Kanade. An iterative image registration technique with an application to stereo vision. In *Proceedings of the 7th International Joint Conference on Artificial Intelligence - Volume 2*, IJCAI'81, pages 674–679, 1981.

[22] Binu M. Nair, Kimberley D. Kendricks, Vijayan K. Asari, and Ronald F. Tuttle. Optical flow based kalman filter for body joint prediction and tracking using hog-lbp matching. In *Proceedings of SPIE, Video Surveillance and Transportation Imaging Applications*, volume 9026, pages 90260H–90260H–14, 2014.

[23] Binu M. Nair and Kimberly D. Kendricks. chapter Improved Region-Based Kalman Filter for Tracking Body Joints and Evaluating Gait in Surveillance Videos", title="Advanced Concepts for Intelligent Vision Systems: 16th International Conference, ACIVS 2015, Catania, Italy, October 26-29, 2015. Proceedings, pages 311–322. Springer International Publishing, 2015.

[24] BinuM. Nair, KimberlyD. Kendricks, VijayanK. Asari, and RonaldF. Tuttle. Body joint tracking in low resolution video using region-based filtering. In *Advances in Visual Computing*, volume 8887 of *Lecture Notes in Computer Science*, pages 619–628. Springer International Publishing, 2014.

[25] T. Ojala, M. Pietikainen, and T. Maenpaa. Multiresolution grayscale and rotation invariant texture classification with local binary patterns. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(7):971–987, 2002.

[26] Kimberly D. Kendricks Sean S Kohles, Anum Barki and Ronald F. Tuttle. Biomechanical analysis of concealed pack load influences on terrorist gait signatures derived from grobner basis theory. *Journal*

of Forensic Biomechanics, 5(0):–, 2014.

[27] A. Toshev and C. Szegedy. Deeppose: Human pose estimation via deep neural networks. In *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*, pages 1653–1660, June 2014.

[28] Yi Yang and D. Ramanan. Articulated pose estimation with flexible mixtures-of-parts. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 1385–1392, June 2011.

[29] Zhaoxiang Zhang, Maodi Hu, and Yunhong Wang. *Biometric Recognition: 6th Chinese Conference, CCBR 2011, Beijing, China, December 3-4, 2011. Proceedings*, chapter A Survey of Advances in Biometric Gait Recognition, pages 150–158. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011.

Author Biography

Binu Nair graduated with a PhD in Electrical Engineering from University of Dayton in May 2015 where his dissertation was on localizing human actions from streaming videos using deep networks. He is a post-doctoral research associate with University of Nevada, Las Vegas during the summer of 2015. His research is focused on the motion analysis for gait-based threat identification and action localization from streaming videos.

Kimberly Kendricks is the STEM coordinator/proposal writer in the Center for Mathematics, Science, and Engineering Education at University of Nevada, Las Vegas. Prior to joining UNLV, Kendricks was also an interim department chair and associate professor of mathematics in the department of mathematics and computer science at Central State University, Ohio. For her research, which has been supported by the National Science Foundation, the American Society for Engineering Education, the National Signature Program, and Clarkson Aerospace Engineering, she takes mathematical theories and applies them to real world problems in robotics and gait analysis.