# Exploring the Space of an Action for Human Action Recognition

Yaser Sheikh Mubarak Shah

Computer Vision Laboratory,
School of Computer Science,
University of Central Florida,
Orlando, FL 32826

## Abstract

*One of the fundamental challenges of recognizing actions is accounting for the variability that arises when arbitrary cameras capture humans performing actions. In this paper, we explicitly identify three important sources of variability: (1) viewpoint, (2) execution rate, and (3) anthropometry of actors, and propose a model of human actions that allows us to address all three. Our hypothesis is that the variability associated with the execution of an action can be closely approximated by a linear combination of action bases in joint spatio-temporal space. We demonstrate that such a model bounds the rank of a matrix of image measurements and that this bound can be used to achieve recognition of actions based only on imaged data. A test employing principal angles between subspaces that is robust to statistical fluctuations in measurement data is presented to find the membership of an instance of an action. The algorithm is applied to recognize several actions, and promising results have been obtained.*

## 1. Introduction

Developing algorithms to recognize humans actions has proven to be an immense challenge since it is a problem that combines the uncertainty associated with computational vision with the added whimsy of human behavior. Even without these two sources of variability, the human body has no less than 244 degrees of freedom ([19]) and modeling the dynamics of an object with such non-rigidity is no mean feat. Further compounding the problem, recent research into anthropology has revealed that body dynamics are far more complicated than was earlier thought, affected by age, ethnicity, class, family tradition, gender, sexual orientation, skill, circumstance and choice, [4]. Human actions are not merely functions of joint angles and anatomical landmark positions, but bring with them traces of the psychology, the society and culture of the actor. Thus, the sheer range and complexity of human actions makes developing action recognition algorithms a daunting task. So how does one appropriately model the non-rigidity of human motion? How do we account for the personal styles (or motion signatures, [17]) while recognizing actions? How do we account for the diverse shapes and sizes of different people? In this paper, we consider some of these questions while developing a model of human actions that approaches these issues. To begin with, it is important to identify properties that are expected to vary with each observation of an action, but which should not affect recognition:

**Viewpoint** The relationship of action recognition to object recognition was observed by Rao and Shah in [13], and developed further by Parameswaran and Chellappa in [9], [10] and Gritai *et al* in [6]. In these papers, the importance of view invariant recognition has been stressed, highlighting the fact that, as in object recognition, the vantage point of the camera should not affect recognition. The projective and affine geometry of multiple views is well-understood, see [7], and various invariants have been proposed.

**Anthropometry** In general, an action can be executed, irrespective of the size or gender of the actor. It is therefore important that action recognition be unaffected by so-called “anthropometric transformations”. Unfortunately, since anthropometric transformations do not obey any known laws, formally characterizing invariants is impossible. However, empirical studies have shown that these transformations are not *arbitrary* (see [3]). This issue has previously been addressed by Gritai *et al.* in [6].

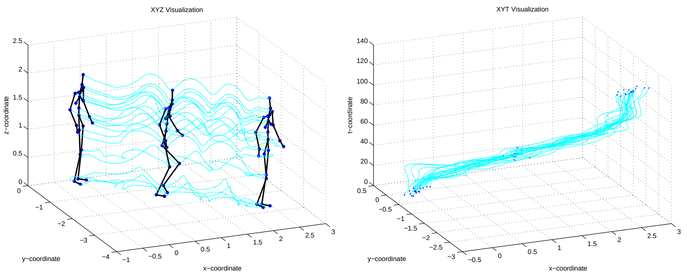
**Execution Rate** With rare exceptions such as synchronized dancing or army drills, actions are rarely executed at a precise rate. It is desirable, therefore, that action recognition algorithms remain unaffected by some set of temporal transformations. The cause of temporal variability can be two fold, caused by the actor or by differing camera frame-rates. Dynamic time warping has been a popular approach to account for highly non-linear transformations, [13].

Recognition presumes some manner of grouping and the question of what constitutes an action is a matter of perceptual grouping. It is difficult to quantify exactly, for instance, whether “walking quickly” should be grouped together with “walking” or with “running”, or for that matter whether walking and running should be defined as a single action or not. Thus grouping can be done at different levels of abstraction and, more often than not, depends on circumstance and context. In this paper, rather than arbitrarily defining some measure of similarity between actions, we allow membership to be defined through exemplars of a group. Our hypothesis is that the variability associated with the execution of an action can be closely approximated by a linear combination of action bases in joint spatio-temporal space. We demonstrate that such a model bounds the rank of a matrix of image measurements and that this bound can be used to achieve recognition of actions based only on imaged data. A test employing principal angles between subspaces that is robust to statistical fluctuations in measurement data is presented to find the membership of an instance of an action. The algorithm is applied to recognize several actions, and promising results have been obtained. As in [9] and [6], we do not address lower-level processing tasks such as shot segmentation, object detection, and body-joint detection. Instead, we assume the image-positions of anatomical landmarks on the body are provided, and concentrate on how best to model and use this data to recognize actions. Johansson demonstrated that point-based representations of human actions were sufficient for the recognition of actions, [8]. In our work, the input is the 2D motion of a set of 13 anatomical landmarks, $\mathcal{L} = \{1, 2, \cdots 13\}$, as viewed from a camera.

The rest of the paper is organized as follows. We situate our work in context of previous research in Section 2. In Section 3, we present our model of human actions and discuss some properties of the proposed framework, followed by the development of a matching algorithm in Section 4. Results are presented in Section 5, followed by conclusions in Section 6.

## 2 Previous Work

Action recognition has been an active area of research in the vision community since the early 90s. A survey of action recognition research by Gavrila, in [5], classifies different approaches into three categories: (1) 2D approaches without shape models, (2) 2D approach with shape models and (3) 3D approaches. Since the publication of this survey, a newer approach to action recognition has emerged: 2D approaches based on 3D constraints, which maintain invariance to viewpoint, while avoiding difficulties of 3D reconstruction of non-rigid motion. The first approach to use 3D constraints on 2D measurements was proposed by Seitz and Dyer in [14], where sufficient conditions for determining whether measurements were 2D affine projections of 3D cyclic motion. Rao and Shah extended this idea in [13], to recognize non-cyclic actions as well, proposing a representation of action using dynamic instances and intevals and proposing a view invariant measure of similarity. Syeda-Mahmood and Vasilescu proposed a view invariant method of concurrently estimating the fundamental matrix and recognizing actions in [15]. In [9], Parameswaran and Chellappa use 2D measurements to match against candidate action volumes, utilizing 3-D model based invariants. In addition to view invariance, Gritai *et al.* proposed a method that was invariant to changes in the anthropometric proportions of actors. As in [13] and [9], they ignored time and treated each action as an object in 3D.



(a) (b)

Figure 1: Representation of an action in 4-space. (a) Action in $XYZ$ space, (b) Action in $XYT$ space. The actor is shown at frame 1, 50 and 126.

## 3 The Space of an Action

By marginalizing time, several papers have represented actions essentially as objects in 3D ([13], [6] and [9]). While some success has been achieved, ignoring temporal information in this way and focussing only on order, makes modeling of temporal transformations impossible. Instead, since an action is a function of time, in this work an instance of an action is modeled as a spatio-temporal construct, a set of points, $\mathbf{A} = [\mathbf{X}_1, \mathbf{X}_2, \ldots \mathbf{X}_p]$, where $\mathbf{X}_i = (X^j_{T_i}, Y^j_{T_i}, Z^j_{T_i}, T_i)^\intercal$ and $j \in \mathcal{L}$ (see Figure 1) and $p = 13n$, for $n$ recorded postures of that action[1]. An *instance* of an action is defined as a linear combination of a set of *action-basis* $\mathbf{A}_1, \mathbf{A}_2, \ldots \mathbf{A}_k$. Consequently, any instance of an action can be expressed as,

$$\mathbf{A}' = \sum_{i=1}^{k} a_i \mathbf{A}_i, \qquad (1)$$

where $a_i \in \mathbb{R}$ is the coefficient associated with the action-basis $\mathbf{A}_i \in \mathbb{R}^{4\times p}$. The space of an action, $\mathcal{A}$ is the span of all its action bases. By allowing actions to be defined in this

[1] The construction of $\mathbf{A}$ must respect the ordering of $\mathcal{L}$.

way by action bases we do not impose arbitrary definitions on what an actions is. The action is defined entirely by the constituents of its action bases. The variance captured by the action bases can include different (possibly non-linearly transformed) execution rates of the same action, different individual styles of performance as well as the anthropometric transformations of different actors (see Figure 2). In general, the number of samples per execution of an action will not necessarily be the same. In order to construct the bases, the entire duration of each action is sampled the same number of times.

### 3.1 Action Projection

When change in depth of a scene is small compared to the distance between the camera and the scene, affine projection models can be used to approximate the process of imaging. In this paper, we assume a special case of affine projection - weak-perspective projection. Weak-perspective projection is effectively composed of two steps. The world point is first projected under orthography, followed by a scaling of the image coordinates. This can be written as,

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{bmatrix} \alpha_x \mathbf{r}^{1\,\mathsf{T}} \\ \alpha_y \mathbf{r}^{2\,\mathsf{T}} \end{bmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} + \mathbf{D} \qquad (2)$$

where $\mathbf{r}^{i\mathsf{T}}$ is the $i$-th row of the rotation matrix $\mathbf{R}$, $\alpha_i$ is a constant scaling factor and $\mathbf{D}$ is the displacement vector. Here, a fixed camera is observing the execution of an action across time. For our purposes we find it convenient to define a canonical world time coordinate $T$, where the imaged time coordinate is related to the world time coordinate $t$ by a linear relationship, $t = \alpha_t T + d_t$ where $\alpha_t$ is temporal scaling factor and $d_t$ is a temporal displacement. This transformation in time can occur because of varying frame rates, because the world action and the imaged action are separated in time or because of linear changes in the speed of execution of an action. *If two actions differ only by a linear transformation in execution rate, we consider them equivalent.* We can define a space-time projection matrix $\widehat{\mathbf{R}}_{3\times4}$ that projects a point $(X, Y, Z, T)^{\mathsf{T}}$ to it's image $(x, y, t)$,

$$\begin{pmatrix} x \\ y \\ t \end{pmatrix} = \begin{bmatrix} \alpha_x \mathbf{r}^{1\,\mathsf{T}} & 0 \\ \alpha_y \mathbf{r}^{1\,\mathsf{T}} & 0 \\ \mathbf{0}^{\mathsf{T}} & \alpha_t \end{bmatrix} \begin{pmatrix} X \\ Y \\ Z \\ T \end{pmatrix} + \begin{bmatrix} \mathbf{D} \\ d_t \end{bmatrix}$$

or

$$\mathbf{x} = \widehat{\mathbf{R}}\mathbf{X} + \widehat{\mathbf{D}}.$$

As in [2] and [16], we can eliminate $\widehat{\mathbf{D}}$ by subtracting the mean of all imaged points. Thus, in our setup, where each instance of an action, $\mathbf{A}_i$, is being observed by a stationary camera, we have $\mathbf{a}_i = \widehat{\mathbf{R}}\mathbf{A}_i$. Available data is usually in terms of these imaged position of the landmarks across time. A matrix $\mathbf{W}$ can be constructed from these measurements as,

$$\mathbf{W} = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,n} \\ y_{1,1} & y_{1,2} & \cdots & y_{1,n} \\ t_{1,1} & t_{1,2} & \cdots & t_{1,n} \end{bmatrix}. \qquad (3)$$

We now show that simply given sufficient *imaged* exemplars of an action in the form of $\mathbf{W}$, a new action, $\mathbf{W}$ can be recognized.

**Proposition 1** If $\mathcal{W}$ is constructed of images of several instances of an action that span the space of that action, and $\mathbf{W}'$ is another instance of that action, then

$$\mathrm{rank}(\mathcal{W}) = \mathrm{rank}([\mathcal{W}\ \ \mathbf{W}']).$$

Under affine projection, we have,

$$\mathbf{W} = \mathbf{R}\mathbf{A} = \mathbf{R}\sum_{i=1}^{k} a_i \mathbf{A}_i = \underbrace{[a_1\mathbf{R}\ \ \cdots\ \ a_k\mathbf{R}]}_{4k} \begin{bmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \\ \vdots \\ \mathbf{A}_k \end{bmatrix} \qquad (4)$$

When several observations are available,

$$\mathcal{W} = \begin{bmatrix} \mathbf{W}_1 \\ \mathbf{W}_2 \\ \vdots \\ \mathbf{W}_n \end{bmatrix} = \begin{bmatrix} a_{1,1}\mathbf{R}_1 & \cdots & a_{k,1}\mathbf{R}_1 \\ \vdots & & \vdots \\ a_{1,n}\mathbf{R}_n & \cdots & a_{k,n}\mathbf{R}_n \end{bmatrix} \begin{bmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \\ \vdots \\ \mathbf{A}_k \end{bmatrix} \qquad (5)$$

Since the columns of $\mathcal{W}$ are spanned by $\mathcal{A}$, the rank of $\mathcal{W}$ is at most $4k$. Now if an observed action $\mathbf{A}'$ is an instance of the same action, then it too should be expressible as a linear combination (Equation 1) of $\mathbf{A}_i$, and therefore the the rank of $[\mathcal{W}\ \ \mathbf{W}']$ should remain $4k$. If it is not the same action, i.e. that is not expressible as a linear combination, then the rank should increase.

An important consequence of Proposition 1 is that we do not need to explicitly compute either the action bases or the space-time projection matrix. In the next section, we show how membership can be tested using only imaged measurements.

## 4 Recognizing an Action Using the Angle between Subspaces

Given a matrix $\mathbf{W}'$ containing measurements of the imaged position of the anatomical landmarks of an actor $e$, we wish to find which of $c$ possible actions was performed. The measured image positions are described in terms of the true po-

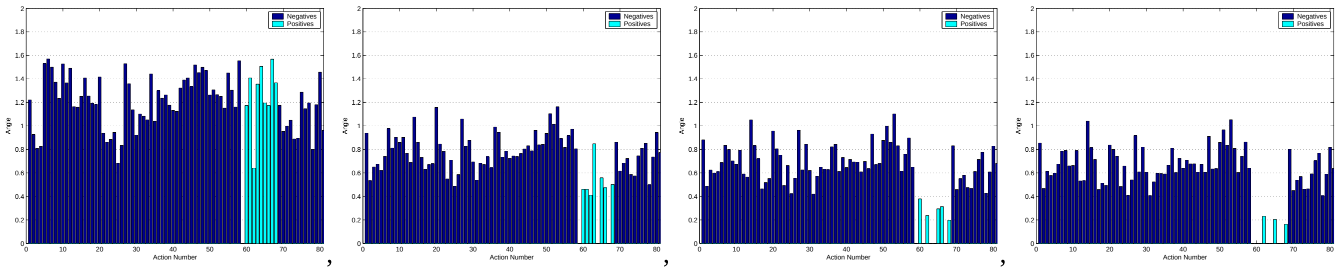


Figure 3: Change in Subspace angles as number of action exemplars are increased. We incrementally added positive examples to the training set and as the span of the action bases increased the angle associated with the positive examples decreased much more than that of the negative examples.

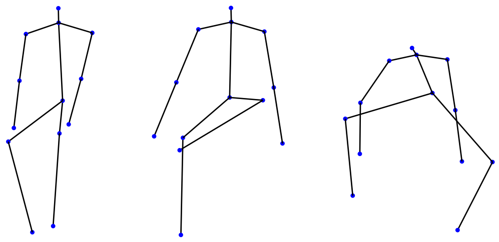


Figure 2: Different postures associated with sitting. Our hypothesis is that there exists a set of action-basis that can compactly describe different styles and rates of execution of an action.

sitions, $\bar{\mathbf{W}}$ with independent normally distributed measurement noise, $\mu = 0$ and variance $\sigma^2$, that is

$$\mathbf{W}' = \bar{\mathbf{W}}' + \epsilon, \epsilon \sim \mathcal{N}(0, \sigma). \tag{6}$$

Our objective is to find the Maximum Likelihood estimate of $j^*$ such that,

$$j^* = \arg\max_{j \in c} p(\mathcal{A}_j | \mathbf{W}'). \tag{7}$$

Because of Proposition 1, we do not need to have the actual action bases to evaluate $p(\mathcal{A}_j|\mathbf{W}')$. Instead, each action is defined by a set of imaged exemplars, that describe the possible variation in the execution of that action. This variation may arise from any one of the many reasons discussed in the introduction. Thus for each action $\mathcal{A}_j$ we have a set of exemplars of that action, $\mathcal{W}_j = [\mathbf{W}_{j,1}, \mathbf{W}_{j,2} \cdots \mathbf{W}_{j,n}]$, where $n \geq k$ and $\dim(\mathcal{A}_j) = k$.

Now, $\mathcal{W}$ and $\mathbf{W}'$ are matrices defining two subspaces, and without loss of generality assume that, $\dim(\mathcal{W}) \geq \dim(\mathbf{W}') \geq 1$. The smallest angle $\theta_1 \in [0, \pi/2]$ between $\mathcal{W}$ and $\mathbf{W}'$ is,

$$\cos\theta_1 = \max_{u \in \mathcal{W}} \max_{v \in \mathbf{W}'} u^\intercal v, \tag{8}$$

where $\|u\|_2 = 1, \|v\|_2 = 1$. It was shown by Wedin in [11] that the angle between $\mathcal{W}$ and $\mathbf{W}'$ gives an estimate of the amount of new information afforded by the second matrix not associated with measurement noise. They show that for two matrices $A$ and $B$, where $B$ is a perturbation of $A$, i.e. if $A = B + \epsilon$, the subspace angle between $\mathrm{range}(B)$ and $\mathrm{range}(A)$ is bounded as,

$$\sin(\theta) \leq \frac{\|\epsilon\|_2}{\sigma_r(A)},$$

where $\sigma_r(A)$ is the $r$-th eigenvalue of $A$. The stability to measurement error makes the angle between subspaces an attractive alternative to standard re-projection errors. Thus,

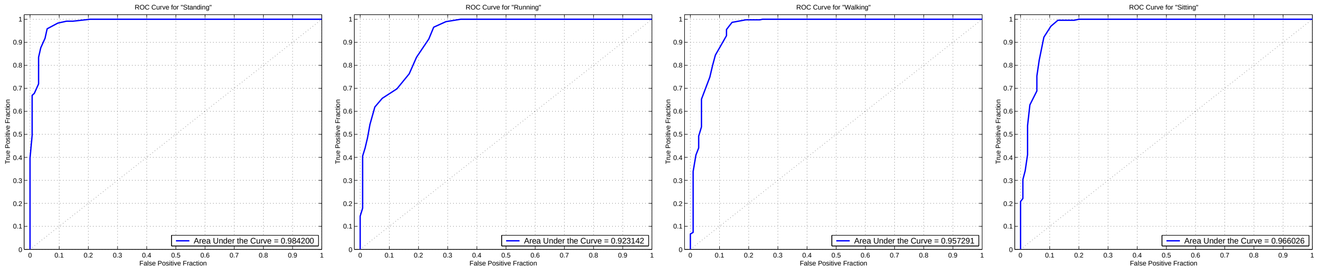
$$p(\mathcal{A}_j|\mathbf{W}) = L(\mathbf{W}|\mathcal{W}_j) \propto \cos\theta_1. \tag{9}$$

where $L$ is the likelihood function. The recognition algorithm, along with the algorithm described by Björk and Golub in [1] for the numerically stable computation of the angle between the subspaces, is given in Figure .

## 5 Results

During experimentation our goal was to verify the claims in this paper, namely that the proposed algorithm can recognize actions despite changes in viewpoint, anthropometry and execution rate. Furthermore, through experimentation, we validate our conjecture that an action can be described by a set of exemplars of that actions. In our experiments, we used a number of complex sequences which were a mix of real motion capture data[2] and also direct video measurements, performed by different actors in many different ways. The test data included the following actions: Sitting, Standing, Falling, Walking, Dancing and Running. In addition to differences in anthropometry and the execution rates, we also generated different views by changing projection matrices for the motion-captured data. For the imaged data, too, we captured the sequences from several different views. Figure 7(a) shows some examples of "Sitting" while Figure 7(b) shows some examples of "Walking".

[2] We did not use any $Z$ information while testing the recognition.



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

Figure 4: ROC Curves for four actions. The dotted diagonal line shows the random prediction. (a) ROC Curve for Standing. The area under the ROC curve is 0.9842. (b) ROC Curve for Running. The area under the ROC Curve is 0.9231. (c) ROC Curve for Walking. The area under the ROC Curve is 0.9573. (d) ROC Curve for Sitting. The area under the ROC Curve is 0.9660.

**Objective**

Given a matrix $\mathbf{W}'$ corresponding to the projection of an action instance, and matrices $\mathcal{W}_1, \mathcal{W}_2, \cdots \mathcal{W}_N$ each modeling the $N$ different actions, find which action was mostly likely executed.

**Algorithm**

For each action $\mathcal{A}_i, i \in 1, 2, \cdots N$ do,

1. **Normalization**: Compute a similarity transform, transforming the mean of the points to the origin and making the average distance of the points from the origin equal to $\sqrt{2}$. This should be done separately for each action instance.
2. **Compute Subspace Angle between $\mathbf{W}$ and $\mathcal{W}_i$**:
   - **Compute Orthogonal Bases:** Use SVD to reliably compute orthonormal bases of $\mathbf{W}'$ and $\mathcal{W}_i$, $\widetilde{\mathbf{W}}'$ and $\widetilde{\mathcal{W}_i}$.
   - **Compute Projection:** Using the iterative procedure described in [1], for $j \in 1, \cdots p$

$$\mathbf{W}'_{i+1} = \mathbf{W}'_i - \mathcal{W}\mathcal{W}^{\mathsf{T}}\mathbf{W}'_i$$

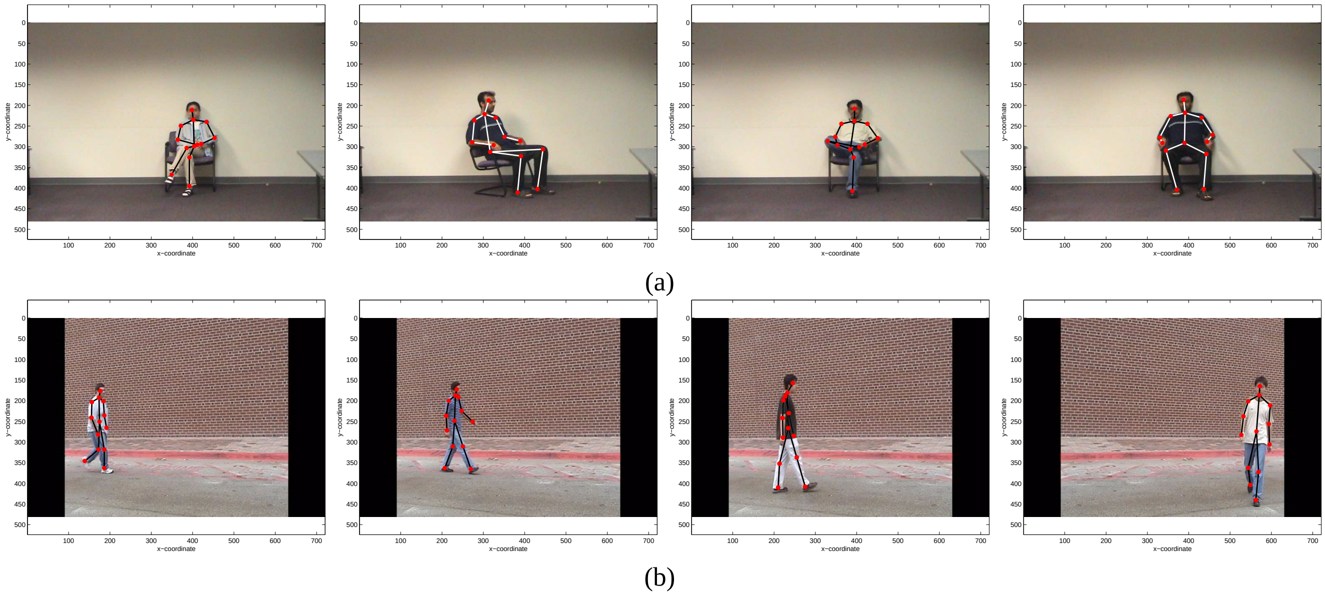
   - **Find Angle:** Compute $\theta = \arcsin \quad \min(1, \|\mathbf{W}'_p\|_2)$ .

Select $i^* = \arg\max_{i \in \{1, \cdots, N\}} \cos(\theta_1)$.

Figure 5: Algorithm for Action Recognition

## 5.1 Action Recognition Results

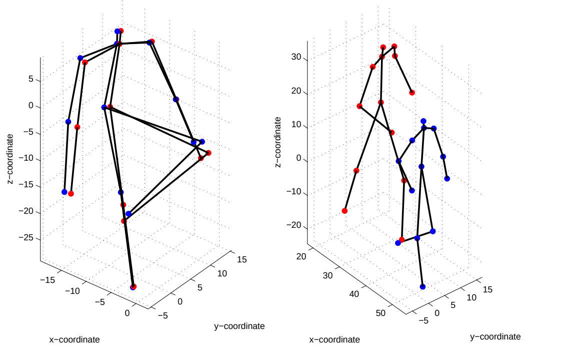
The set of action exemplars for each action is composed by adding exemplars iteratively until these sufficiently span the action space. To achieve this using the minimum number of training samples, we iteratively picked the action sequence from the corpus of that action which has the maximum angle between it and the action subspace (at that point) and continue until the rank of the action subspace is unaffected by additions of further exemplars from the corpus. This greedy method allows us to minimize the number of training samples required to span the action space instead of just selecting an arbitrary number of exemplars. The effect of increasing the number of exemplars in this way is shown in Figure 3. Clearly, the angles of the positive exemplars are ultimately significantly lower than those of the negative exemplars. To test our approach, for each action, we take all the instances of that action in the corpus as positive sequences and the sequences for all the other actions as negative sequences. Table 1 shows the number of training and testing sequences that were eventually used to obtain the final results. For each action in our testing set, we computed the angle between the action space and the subspace spanned by that action instance. This result is thresholded to give the final binary classification. The ROC curves based on this classification for “Walking”, “Sitting”, “Standing” and “Running” are shown in Figure 4. As can be seen from the area under these ROC curves, using our approach we have been able to correctly classify most of the actions. These figures also show that ”Standing” and ”Sitting” are better classified than ”Walking” and ”Running”. As discussed earlier in Section 1, this is because ”Walking” and ”Running” are similar actions and it is comparatively difficult to distinguish between them, although our method is



(a)

(b)

Figure 8: Video sequences used in the experiments. Actors of both genders, and of different body proportions were observed performing actions in different styles. (a) Sitting sequences. Clearly, each actor sits in a unique way, with legs apart, or one leg crossed over the other. (b) Walking sequences. Sources of variability include arm swing, speed, and stride length.



(a)

(b)

Figure 6: Reconstructing an action in XYZT. We do not require such reconstruction in our recognition algorithm. In this figure, we are simply demonstrating the validity of our hypothesis. (a) Accurate reconstruction. The last frame of an instance of “Sitting” is shown. The blue-markered skeleton represents the original measurements, the red-markered skeleton represents the reconstruction after projection on the “Sitting” action basis. (b) The 32th frame of an instance of “Running” is shown. The blue-markered skeleton represents the original measurements and the red-markered skeleton represents the reconstruction after projection onto the “Walking” action basis.

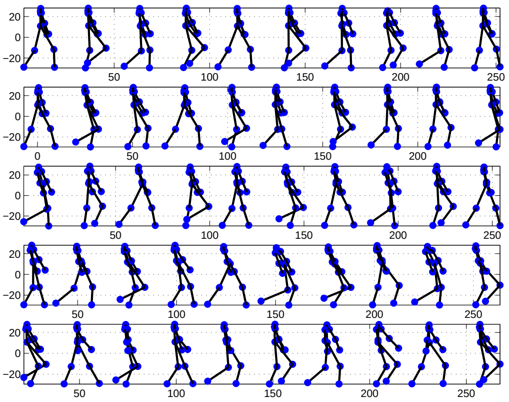
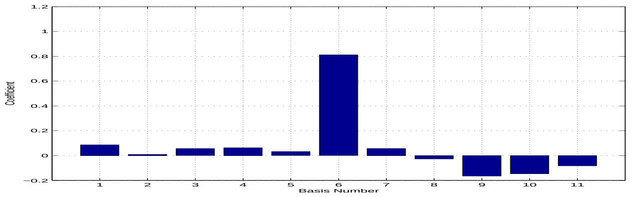


Figure 7: Variations in Walking. Ten evenly sampled postures were taken from the duration of the action.

still able to distinguish between these to a large extent. Reconstruction of an imaged action after projection onto the action basis for “Sitting” is shown, along with its coefficients for each of its 11 basis is shown in Figure 9.

# 6 Summary and Conclusions

In this paper we have developed a framework for learning the variability in the execution of human actions that is unaffected by the changes. Our hypothesis is that any instance of an action can be expressed as a linear combination of



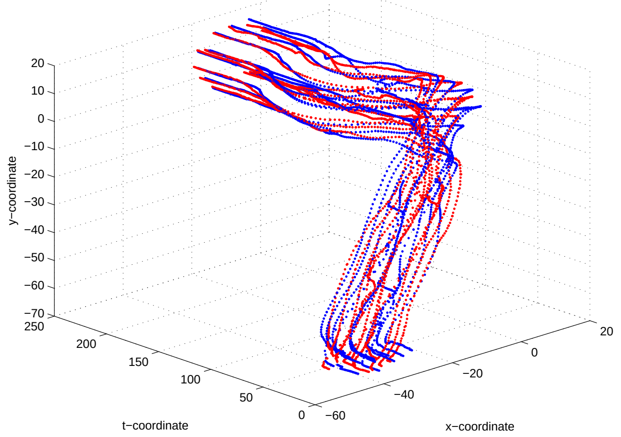


Figure 9: Action reprojection of "Sitting" in $xyt$ space. The red points show the original action in $xyt$ and the blue points show close reconstruction after projection onto the action bases of sitting. Note this is based on imaged exemplars only.

| | Action Exemplars | # of Positive | # of Negative |
|---|---|---|---|
| Sitting | 11 | 230 | 127 |
| Standing | 12 | 120 | 138 |
| Running | 17 | 290 | 121 |
| Walking | 11 | 450 | 105 |

Table 1: Number of training and testing samples for each of the four actions recognized during the experiments.

spatio-temporal action basis, capturing different personal styles of execution of an action, different sizes and shapes of people, and different rates of execution. We demonstrate that using sufficient *imaged* exemplars of actions, an action as view from a camera can be recognized using the angle between the subspace of the exemplars and the subspace of the inspection instance. This concept is related to earlier factorization approaches proposed by (among others) Tomasi and Kanade in [16], and Bregler *et al.* in [2]. In particular in [2], non-rigid motion viewed by a single camera over time was modeled as a linear combination 3D shape basis. However, rather than factorizing measurement matrices constructed from a single camera, in the case of objects, we factorize measurement matrices captured across multiple cameras. In this work, we are not interested in explicitly recovering the actual three (or four) dimensional actions or action bases, but instead to use the constraints they provide to perform recognition. Future directions could involve recovering the 4D structure of an action explicitly, aided by action bases.

# References

[1] A. Björk and G. Golub, "Numerical Methods for Computing Angles between Linear Subspaces," Mathematics of Computation, 1973.

[2] C. Bregler, A. Hertzmann and H. Biermann, "Recovering Non-Rigid 3D Shape from Image Streams," *IEEE Conference of Computer Vision and Pattern Recognition*, 2000.

[3] R. Easterby, K. Kroemer and D. Chaffin, "Anthropometry and Biomechanics - Theory and Appplication," Plenum Press, 1982.

[4] B. Farnell, "Moving Bodies, Acting Selves," *Annual Review of Anthropology*, Vol. 28, 1999.

[5] D. Gavrila, "The Visual Analysis of Human Movement: A Survey," *Computer Vision and Image Understanding*, 1999.

[6] A. Gritai, Y. Sheikh, M. Shah, "On the Use of Anthropometry in the Invariant Analysis of Human Actions," *International Conference on Pattern Recognition*, 2004.

[7] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, 2000.

[8] G. Johansson, "Visual perception of biological motion and a model for its analysis," *Perception and Psychophysics*, 1973.

[9] V. Parameswaran, R. Chellappa, "View Invariants for Human Action Recognition," *IEEE Conference on Computer Vision and Pattern Recognition*, 2003.

[10] V. Parameswaran, R. Chellappa, "Quasi-Invariants for Human Action Representation and Recognition," *IEEE International Conference on Pattern Recognition*, 2002.

[11] P.-A. Wedin, "On angles between subspaces of a finite dimensional inner product space," Matrix Pencils, Lecture notes in Mathematics, Kagstrom and Ruhe (Eds.), 1983.

[12] C. Rao, A. Gritai, M. Shah, "View-invariant Alignment and matching of Video Sequences," *IEEE International Conference on Computer Vision*, 2003.

[13] C. Rao, M. Shah, "View-Invariance in Action Recognition," *IEEE Conference on Computer Vision and Pattern Recognition*, 2001.

[14] S. Seitz and C. Dyer, "View-Invariant Analysis of Cyclic Motion," *International Journal of Computer Vision*, 1997.

[15] T. Syeda-Mahmood and A. Vasilescu, "Recognizing action events from multiple viewpoints," *IEEE Workshop on Detection and Recognition of Events in Video*, 2001.

[16] C. Tomasi and T. Kanade, “Shape and Motion from Image Streams under Orthography,” International Journal of Computer Vision, 1992.

[17] M. Vasilescu, “Human Motion Signatures: Analysis, Synthesis, recognition,” *International Conference on Pattern Recognition*, 2002.

[18] A. Veeraraghavan, A. Roy Chowdhury and R. Chellappa,“Role of Shape and Kinematics in Human Movement Analysis,” *IEEE Conference on Computer Vision and Pattern Recognition*, 2004.

[19] V. Zatsiorsky, “Kinematics of Human Motion,” *Human Kinetics*, 2002.