# Estimation of Skill Levels in Sports Based on Hierarchical Spatio-Temporal Correspondences

**Conference Paper** · September 2003

DOI: 10.1007/978-3-540-45243-0_67 · Source: DBLP



**3 authors**, including:

Winfried Ilg
University of Tuebingen
**73** PUBLICATIONS **1,032** CITATIONS

SEE PROFILE

Martin A. Giese
University of Tuebingen
**374** PUBLICATIONS **5,240** CITATIONS

SEE PROFILE

# Estimation of Skill Levels in Sports based on Hierarchical Spatio-Temporal Correspondences

Winfried Ilg[1] and Johannes Mezger[2] and Martin Giese[1]

[1] Laboratory for Action, Representation and Learning
Department for Cognitive Neurology, University Clinic Tübingen, Germany
{wilg,giese}@tuebingen.mpg.de
[2] Graphical-Interactive Systems, Wilhelm Schickard Institute for Computer Science
University Tübingen, Germany

## Abstract

*We present a learning-based method for the estimation of skill levels from sequences of complex movements in sports. Our method is based on a hierarchical algorithm for computing spatio-temporal correspondence between sequences of complex body movements. The algorithm establishes correspondence at two levels: whole action sequences and individual movement elements. Using Spatio-Temporal Morphable Models we represent individual movement elements by linear combinations of learned example patterns. The coefficients of these linear combinations define features that can be efficiently exploited for estimating continuous style parameters of human movements. We demonstrate by comparison with expert ratings that our method efficiently estimates the skill level from the individual techniques in a "karate kata".*

## 1 Introduction

The analysis of complex movements is an important problem for many technical applications in computer vision, computer graphics, sports and medicine (see reviews in [6] and [10]). For several applications it is crucial to model different styles of movements, for example to quantify the movement disorders in medical gait analysis, or for the classification and description of different skill-levels in sports. In the literature different methods for the parameterization of styles of complex movements have been proposed, e.g. based on hidden Markov models [2][13], principal component analysis [15][1] or fourier coefficients [12].

An efficient method for the synthesis of movements with different styles is the linear combination of example trajectories. Such linear combinations can be defined efficiently on the basis of spatio-temporal correspondence. The technique of Spatio-Temporal Morphable Models (STMMs) defines linear combinations by weighted summation of spatial and temporal displacement fields that morph the prototypical movement trajectories into a reference pattern. This method has been successfully applied for the generation of cyclic movements in computer graphics (motion morphing [3],[14]) as well as for the recognition of movements and movement styles from trajectories in computer vision [8].

To generalize the method of linear combination for complex sequences containing many complex movements we extend the basic STMM algorithm by introducing a second hierarchy level that represents motion primitives. Such primitives correspond to

parts of the approximated trajectories, e.g. techniques in a sequence of karate movements. These movement primitives are then modeled using STMMs by linearly combining example movements. This makes it possible to learn generative models for sequences of movements with different styles. We apply this hierarchical algorithm to model sequences of complex karate movements and to estimate the skill levels of different actors based on the trajectory information obtained by motion capturing.

## 2 Algorithm

An overview of the hierarchical algorithm is shown in figure 1. The next sections describe the extraction of the movement elements and the modeling of the individual elements by STMMs.

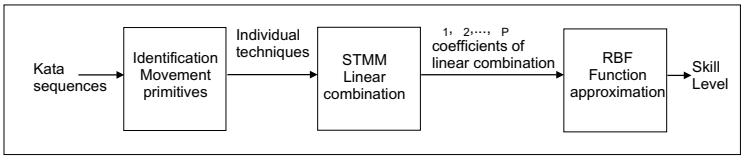


**Fig. 1.** Schematic description of the method and data flow. First, individual techniques are identified and segmented using invariant key features. Then the segmented techniques are represented by linear combination of prototypical example trajectories using STMMs. The resulting linear coefficients $\omega_1 \dots \omega_P$ are mapped onto the estimated skill level with RBF networks that are trained with expert ratings for the skill levels of each individual karate technique.

### 2.1 Identification of movement primitives

For the identification of movement primitives within a complex movement sequence an appropriate description of the spatio-temporal characteristics of the individual movement elements must be found that is suitable for a robust matching with stored example templates. Based on such features spatio-temporal correspondence between new movement sequences and stored example sequences can be established on a coarse level. The underlying features must be invariant against changes of the style of the individual movements elements. Different elementary spatio-temporal and kinematic features, like angular velocity [5][11] or curvature and torsion of the 3D trajectories [4] have been proposed in the literature. The key features of our algorithm are zeros of the velocity in few "characteristic coordinates" of the trajectory $\boldsymbol{\zeta}(t)$. For the matching process, which is based on dynamic programming, we represent the features by discrete events. Let $m$ be the number of the motion primitive and $r$ the number of characteristic coordinates of the trajectory. Let $\boldsymbol{\kappa}(t)$ be the "reduced trajectory" of the characteristic coordinates that takes the values $\boldsymbol{\kappa}_i^m$ at the velocity zeros[3]. The movement primitive is then characterized by the vector differences $\Delta\boldsymbol{\kappa}_i^m = \boldsymbol{\kappa}_i^m - \boldsymbol{\kappa}_{i-1}^m$ of subsequent velocity zeros (see fig. 2).

A robust identification of movement primitives in noisy data with additional or missing zero-velocity points $\kappa_i^s$ can be achieved by dynamic programming. The purpose of the dynamic programming is an optimal sequence alignment between the key features of the prototypical movement primitive $\kappa_0^m \dots \kappa_q^m$ and the key features of a search window $\kappa_0^s \dots \kappa_p^s$ (see fig. 2b). This is accomplished by minimizing a cost function $\delta$ that

[3] Zero-velocity is defined by a a zero of the velocity in at least one coordinate of the reduced trajectory.

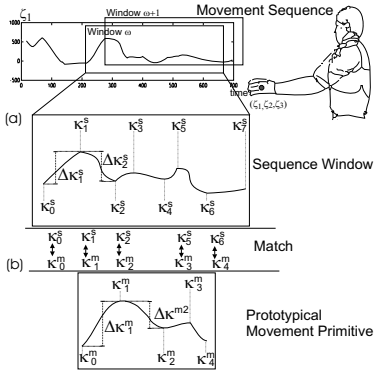


**Fig. 2.** Illustration of the method for the automatic identification of movement primitives: (a) In a first step all key features $\boldsymbol{\kappa}_i^s$ are determined. (b) Sequences of key features from the sequences (s) are matched with sequences of key features from the prototypical movement primitives (m) using dynamic programming. A search window is moved over the sequence. The length of the window is two times the number of key features of the learned movement primitive. The best matching trajectory segment is defined by the sequence of feature vectors that minimizes $\sum_j ||\Delta\boldsymbol{\kappa}_i^s - \Delta\boldsymbol{\kappa}_j^m||$ over all matched key features. With this method spatio-temporal correspondence at a coarse level is established.

is given by the sum of $||\Delta\boldsymbol{\kappa}_i^s - \Delta\boldsymbol{\kappa}_j^m||$ over all matched key features. A more formal description of the algorithm is given in [9].

### 2.2 Morphable Models for modeling movement primitives

The technique of *Spatio-Temporal Morphable Models* [7],[8] is based on linearly combining the movement trajectories of prototypical motion patterns in space-time. Linear combinations of movement patterns are defined on the basis of spatio-temporal correspondences that are computed by dynamic programming [3]. Complex movement patterns can be characterized by trajectories of feature points. The trajectories of the prototypical movement pattern $p$ can be characterized by the time-dependent vector $\boldsymbol{\zeta}_p(t)$. The correspondence field between two trajectories $\boldsymbol{\zeta}_1$ and $\boldsymbol{\zeta}_2$ is defined by the spatial shifts $\boldsymbol{\xi}(t)$ and the temporal shifts $\tau(t)$ that transform the first trajectory into the second. The transformation is specified mathematically by the equation:

$$\boldsymbol{\zeta}_2(t) = \boldsymbol{\zeta}_1(t + \tau(t)) + \boldsymbol{\xi}(t) \tag{1}$$

By linear combination of spatial and temporal shifts the Spatio-Temporal Morphable Model allows to interpolate smoothly between motion patterns with significantly different spatial structure, but also between patterns that differ with respect to their timing. The correspondence shifts $\boldsymbol{\xi}(t)$ and $\tau(t)$ are calculated by solving an optimization problem that minimizes the spatial and temporal shifts under the constraint that the temporal shifts define a new time variable that is always monotonically increasing. For further details about the underlying algorithm we refer to [7],[8]. Signifying the spatial and temporal shifts between prototype $p$ and the reference pattern by $\boldsymbol{\xi}_p(t)$ and $\tau_p(t)$, linearly combined spatial and temporal shifts can be defined by the two equations:

$$\boldsymbol{\xi}(t) = \sum_{p=1}^{P} w_p\, \boldsymbol{\xi}_p(t) \qquad \tau(t) = \sum_{p=1}^{P} w_p\, \tau_p(t) \tag{2}$$

The weights $w_p$ define the contributions of the individual prototypes to the linear combination. We always assume convex combinations with $0 \leq w_p \leq 1$ and $\sum_p w_p = 1$. After linearly combining the spatial and temporal shifts the trajectories of the morphed pattern can be recovered by morphing the reference pattern in space-time using the spatial and temporal shifts $\boldsymbol{\xi}(t)$ and $\tau(t)$. The space-time morph is defined by equation (1) where $\boldsymbol{\zeta}_1$ is the reference pattern and $\boldsymbol{\zeta}_2$ has to be identified with trajectory of the linearly combined pattern.

**Fig. 3.** A Snapshot from motion capturing karate movements with 11 cameras. The subjects had 41 markers and perform the karate kata "Heian Shodan"

## 3 Experiments

We demonstrate the function of the algorithm by modeling movement sequences from martial arts. Using a motion capture system (VICON 612) with 11 cameras we captured the movements of 7 actors performing the karate kata "Heian Shodan". 14 movement sequences (two sequences per actor) were captured at a sampling frequency of 120 Hz using 41 passively reflecting markers. The actors had different belt levels (Kyu degrees) in karate (Shotokan) (see tab. 1). The kata was decomposed into 20 movement primitives (karate techniques). The total duration of the whole sequences was between 25 and 35 s ( $\doteq$ 3000-4200 captured frames). Each individual technique was rated by an expert on a scale from 0 to 10.

| actor | den | mar | tho | joh | ste | chr | joa |
|---|---|---|---|---|---|---|---|
| Kyu | 7 | 6 | 5 | 5 | 3 | 2 | 1 |
| $\bar{r}_{exp}$ | 8.1 | 7.0 | 6.0 | 4.2 | 2.2 | 1.9 | 0.9 |

**Table 1.** Official belt levels (Kyu degrees) of the seven actors and average of the expert ratings for the individual techniques $\bar{r}_{exp}$ on a scale from 0 to 10 (0 signifying optimal performance, and 10 signifying worst performance).

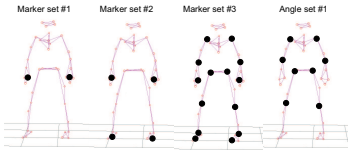


**Fig. 4.** Visualization of different feature sets used for the analysis. Feature sets #1-#3 were based on selected 3D markers. Feature set # 4 is based on joint angles. The black dots illustrate the positions of the selected markers respectively joints.

### 3.1 Identification of individual techniques

The individual techniques were extracted automatically from the kata sequences using the method described in section 2.1. For the representation of the relevant key features we have examined different feature sets based on 3D markers and joint angles (fig. 4). The prototypes for the identification of individual techniques were generated from manually segmented trajectories. As prototypes we used techniques from individual actors and also the averages over all actors generated by time alignment using STMMs. Fig. 5 shows the error measure $\delta$ for all frames for a kata sequence using one particular prototypical movement primitive for matching.

The results of the automatic segmentations are as follows: Out of the 280 individual techniques in the data set 96% were correctly classified. The best segmentation results

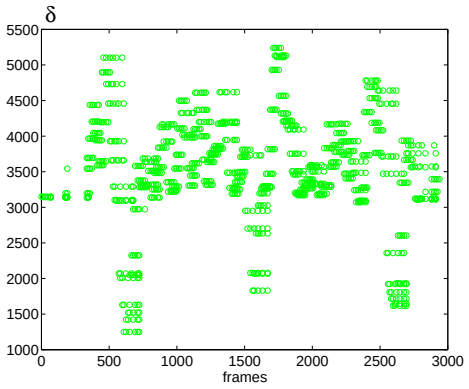


**Fig. 5.** Results of the automatic segmentation for one sequence identifying one technique for actor (chr) using the prototypical movement primitive generated by a 3 movements average $3av$. The diagram shows the distance measure of the dynamic programming method, $\delta$, for different matches of the corresponding movement primitive over the whole sequence. The circles mark the times of the matched key feature $\kappa_i^m$ in the sequence (see fig. 2). Each match of a whole movement primitive is illustrated by a row of circles with the same $\delta$. The number of circles corresponds to the number of key features of the movement primitive. There are three distinct minima in the $\delta$-function corresponding to a technique that occurs multiple times in the kata with slightly different rotation steps. The movement primitive (with the correct rotation step) corresponds to the smallest minimum of the error function that occurs at frame 595.

were obtained using prototypes generated by averaging the trajectories of all actors. Using an average of three actors including a beginner, a medium skilled karateka, and a master, we obtained comparable results. In general, using prototypes generated by averaging we obtained significantly better segmentations than for prototypes that were derived from individual actors. Segmentation errors typically arise when the same technique occurs multiple times in different contexts in the kata. Such errors can be easily removed by taking into account the overall sequence of the techniques in the kata. The best segmentation performance was obtained with marker set # 1. The reason for this result might be that the movements of the feet during many of the techniques were very similar. Our segmentation algorithm was sensitive enough to detect if actors forgot individual movements during the kata.

### 3.2 Modeling of movement element by linear combination

In the next step, the segmented movement elements were approximated by linear combinations of prototypical movement primitives. The weights of these linear combinations are useful (1) for actor identification, if the movement comes from an actor in the prototype set (2) for the estimation of skill levels of actors, which is not in the prototype set.

*Actor identification* Based on the representation by linear combinations a robust actor identification can be realized. The STMM is trained with the automatically segmented movements from all actors. A new movement sequence from actor ($tho$), that was not part in the training set, is approximated by the linear combination of the training movement sequences. Fig. 6.a shows that the linear coefficients $\omega_{p,i}$ peak exactly for the weight of the prototype (tho). In addition, the estimated weight of this prototype is

close to one for all techniques. This actor identification works for both types of feature sets (3D markers and joint angles). The identification is thus not based on the specific kinematic structure of the actor, but rather on specific spatio-temporal characteristics.

*Coefficients of linear combination as basis for skill level estimation* The linear coefficients can also be used as basis for the estimation of skill levels. For this purpose the mapping between the linear coefficients and an estimate for the skill level is learned (sec. 3.3). The coefficients reflect the weighting of the specific spatio-temporal characteristics of the prototypes for the rated movement pattern. We evaluated the method using a leave-one-out paradigm. The STMM was trained with the automatically segmented movements from a set of six prototypes excluding the actor that performed the test sequence. Fig. 6.b shows a typical example of the estimated coefficients $\omega_{p,i}$ for a new actor. Interestingly, the weight estimations for the different movement primitives are similar even though the actor was not in the training set. This indicates correlations in movement styles of different actors that are similar for different movement elements.

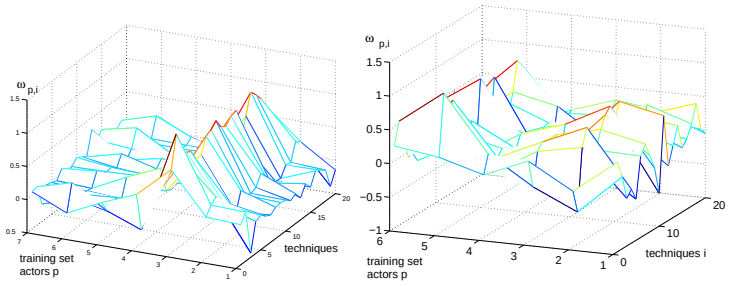


**Fig. 6.** Left panel: Coefficients $\omega_{p,i}$ of the linear combination that approximates a new sequence of actor $tho$. The weight corresponding to this actor has the index $p = 3$. (The tested sequence was not in the training set for the STMM). Right panel: Coefficients $\omega_{p,i}$ of linear combination approximating a sequence of the actor $joh$ after the STMM has been trained with a set of six prototypes excluding actor $joh$ . Consistently, for most movement primitives high coefficients arise for $p = 5$ and $p = 2$.

### 3.3 Skill-level estimation on segments

For the estimation of the skill-levels based on the linear coefficients $\omega_{p,i}$ (eq. 2) we used RBF networks. For each karate technique a separate network was trained that realizes the mapping $RBF_p^i : [\omega_1 \dots \omega_P] \rightarrow r_{est}^{p,i}$, where $r_{est}^{p,i}$ denotes the estimate. The networks were trained with the coefficient vectors for the prototypes $p_1 \dots p_P$ and the expert ratings.

Fig. 7 (left panel) shows a comparison between the estimated skill levels and the expert ratings for all techniques of a single actor. The right panel shows the averaged deviations $\Delta_i = (\sum_{p \in \mathcal{P}} |r_{est}^{p,i} - r_{exp}^{p,i}|)/\#\mathcal{P}$ for all techniques $i$ averaged over all actors. The figure shows that the reliability varies over the movement primitives. A possible explanation is that the techniques vary with respect to their difficulty. Very simple techniques might not so well differentiate between different skill levels as more difficult ones.

For the further analysis the techniques with a reliability $\Delta_i < 0.2$ were determined. Only the estimates of the RBF networks trained with these techniques were combined into a final skill level estimate by computing the average of their outputs.

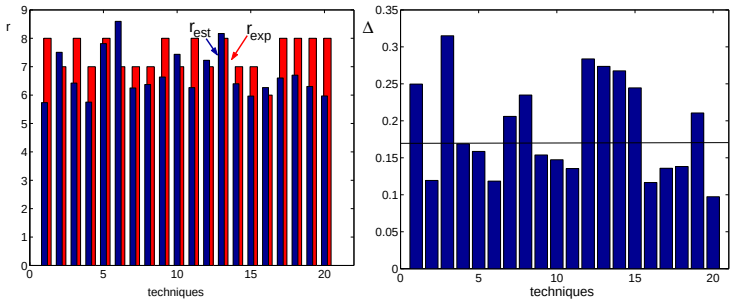


**Fig. 7.** Left panel: Comparison between the estimates $r_{est}$ of the skill level obtained from the individual techniques and the expert ratings $r_{exp}$ for one actor (mar). The averages of the estimated and the real skill levels are quite similar (6.3 vs. 7.0). In particular for techniques that were not executed correctly larger deviations arise. Right panel: Reliability of the skill-level estimates from the different kata techniques. Techniques with $\Delta_i < 0.2$ were used to compute the averaged skill-level.

The overall reliability of the proposed method was then tested with a new data set with sequences from the 7 actors using only the previously selected techniques. The results of the skill-level estimation based on different feature sets and segmentation methods compared with the expert ratings are shown in figure 8 and table 2. The estimates have exactly the same monotonic order as the expert ratings and match them closely in the range of the lower skill levels. Larger deviations occur for some actors with higher ranks (ste and chr). The estimates of the extreme skill levels are shifted towards less extreme values, likely a consequence of the lack of training data outside the range between these extremes.

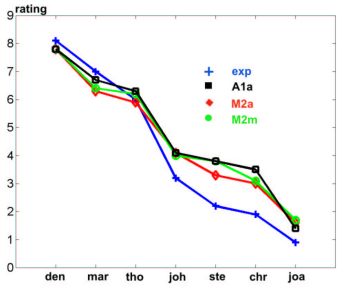


**Fig. 8.** Comparison of the averaged expert ratings (exp) with the automatically estimated ratings using different feature sets (see table 2) for all actors. The automatic estimates obtained with different feature sets are very similar.

| actor | den | mar | tho | joh | ste | chr | joa |
|---|---|---|---|---|---|---|---|
| Kyu | 7 | 6 | 5 | 5 | 3 | 2 | 1 |
| $\text{Kyu}_{est}$ | 7.0 | 5.9 | 5.5 | 3.5 | 3.2 | 2.9 | 1.0 |
| $\bar{r}_{exp}$ | 8.1 | 7.0 | 6.0 | 4.2 | 2.2 | 1.9 | 0.9 |
| $\bar{r}^m_{M2}$ | 7.8 | 6.3 | 5.9 | 4.1 | 3.3 | 3.0 | 1.6 |
| $\bar{r}^a_{M2}$ | 7.8 | 6.4 | 6.2 | 4.0 | 3.8 | 3.1 | 1.7 |
| $\bar{r}^a_{A1}$ | 7.8 | 6.7 | 6.3 | 4.1 | 3.8 | 3.5 | 1.4 |

**Table 2.** Comparison between the belt level (Kyu), expert rating averaged over all techniques $\bar{r}_{exp}$, and the estimated skill levels for different sets of features. Results are shown for automatic and manual segmentation of the movement primitives ($m \doteq manual, a \doteq automatic$). Based on the estimate $\bar{r}^a_{A1}$ an estimated belt level $\text{Kyu}_{est}$ was computed by linear transformation using the extreme skill values (1. and 7. Kyu) as reference points.

## 4 Discussion

We have presented a learning-based method for the quantification of movement styles in sequences of movements that works on small data sets. The proposed method is based on establishing spatio-temporal correspondence between learned prototypical example

sequences and new trajectories exploiting a hierarchical algorithm for the computation of spatio-temporal correspondence. We demonstrated that this technique is suitable for person recognition from individual movement primitives, and for the estimation of skill levels from sequence of complex movements in sports.

Compared to related methods for the representation of movement styles in computer vision and computer graphics (see section 1) the proposed method seems to be interesting for the following reasons: (1) As demonstrated in this paper it works with very small data sets. We applied principle component analysis on the same trajectories using the same type of neural networks and obtained less accurate estimates of the skill level. (2) The coefficients of the STMM are often intuitive to interpret, as shown in figure 6. (3) A further advantage of the proposed method, which applies also to some other techniques, is that representation of movement sequences by linear combinations of learned examples is also suitable for synthesis of movement sequences with defined styles [9]. Future work will have to test the proposed method on bigger data sets.

## Acknowledgments

This work is supported from the Deutsche Volkswagenstiftung. We thank H.P. Thier, B. Eberhardt, W. Strasser, H.H. Bülthoff and the Max Planck Institute for Biological Cybernetics for additional support.

## References

1. A. F. Bobick and J. Davis. An appearance-based representation of action. In *Proceedings of the IEEE Conference on Pattern Recognition*, pages 307–312, 1996.
2. M. Brand. Style machines. In *SIGGRAPH*, 2000.
3. A. Bruderlin and L. Williams. Motion signal processing. In *SIGGRAPH*, pages 97–104, 1995.
4. T. Caelli, A. McCabe, and G. Binsted. On learning the shape of complex actions. In *International Workshop on Visual Form*, pages 24–39, 2001.
5. A. Galata, N. Johnson, and D. Hogg. Learning variable lenth markov models of behavior. *Journal of Computer Vision and Image Understanding*, 81:398–413, 2001.
6. D.M. Gavrila. The visual analysis of human movement: a survey. *Journal of Computer Vision and Image Understanding*, 73:82–98, 1999.
7. M. A. Giese and T. Poggio. Synthesis and recognition of biological motion pattern based on linear superposition of prototypical motion sequences. In *Proceedings of IEEE MVIEW 99 Symposium at CVPR, Fort Collins*, pages 73–80, 1999.
8. M.A. Giese and T. Poggio. Morphable models for the analysis and synthesis of complex motion patterns. *International Journal of Computer Vision*, 38(1):59–73, 2000.
9. W. Ilg and M.A. Giese. Modeling of movement sequences based on hierarchical spatial-temporal correspondence of movement primitives. In *Workshop on Biologically Motivated Computer Vision*, pages 528–537, 2002.
10. T. B. Moeslund. A survey of computer vision-based human motion capture. *Journal of Computer Vision and Image Understanding*, 81:231–268, 2001.
11. T. Mori and K. Uehara. Extraction of primitive motion and discovery of association rules from motion data. In *Proceedings of the IEEE International Workshop on Robot and Human Interactive Communication*, pages 200–206, 2001.
12. M. Unuma, K. Anjyo, and R. Takeuchi. Fourier principles for emotion-based human figure animation. In *SIGGRAPH*, pages 91–96, 1995.
13. A. D. Wilson and A. F. Bobick. Parametric hidden markov models for gesture recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(9):884–900, 1999.
14. A. Witkin and Z. Popovic. Motion warping. In *SIGGRAPH*, pages 105–108, 1995.
15. Y. Yacoob and M. J. Black. Parameterized modeling and recognition of activities. *Journal of Computer Vision and Image Understanding*, 73(2):398–413, 1999.