Segmentation of Action Streams:
Human Observers vs. Bayesian Binning

Dominik Endres, Andrea Christensen, Lars Omlor, and Martin A. Giese

dominik.endres@klinikum.uni-tuebingen.de
{andrea.christensen,martin.giese@uni-tuebingen.de}@uni-tuebingen.de
Theoretical Sensomotorics, Cognitive Neurology, University Clinic Tübingen,
CIN Tübingen and HIH Tübingen,
Frondsbergstr. 23, 72070 Tübingen, Germany

Abstract. Natural body movements are temporal sequences of individual actions. In order to realise a visual analysis of these actions, the human visual system must accomplish a temporal segmentation of action sequences. We attempt to reproduce human temporal segmentations with Bayesian binning (BB)[8]. Such a reproduction would not only help our understanding of human visual processing, but would also have numerous potential applications in computer vision and animation. BB has the advantage that the observation model can be easily exchanged. Moreover, being an exact Bayesian method, BB allows for the automatic determination of the number and positions of segmentation points. We report our experiments with polynomial (in time) observation models on joint angle data obtained by motion capture. To obtain human segmentation points, we generated videos by animating sequences from the motion capture data. Human segmentation was then assessed by an interactive adjustment paradigm, where participants had to indicate segmentation points by selection of the relevant frames. We find that observation models with polynomial order ≥ 3 can match human segmentations closely.

Cite as:
Endres D., Christensen C., Omlor L. and Giese M.A. (2011). Segmentation of Action Streams: Human Observers vs. Bayesian Binning. In *KI2011: Advances in Artificial Intelligence, LNAI 7006, Springer*, 75-86.
The original publication can be found at www.springerlink.com,
DOI: 10.1007/978-3-642-24455-1.

1 Introduction

Temporally segmenting (human) action streams is interesting for a variety of reasons: firstly, if we had a model which reproduced human segmentations closely, might reveal important insights in human action representation. Previous work in this direction has studied in detail the segmentation of sequences of piecewise linear movements in the two-dimensional plane [23,1]. Secondly, a good temporal segmentation would have numerous applications in the field of computer vision. Worth mentioning in this context is the Human Motion Analysis (HMA) [24]. HMA concerns the detection, tracking and recognition of people from image sequences involving humans and finds its application in many areas such as smart surveillance and man-machine interfaces. Thirdly, extraction of important key frames by improved motion segmentation would not only contribute to computer vision research but also to computer graphics and motion synthesis. Animations of human motion data can be done with less computational costs if the key frames are defined optimally (e.g. [6,5]).

While most researchers base their temporal segmentation approaches on real video data and focus on the computer vision problem to analyse human motion data by tracking of skeleton models or feature sequences [21,2,13], we address here specifically the problem of the segmentation of action streams based on motion capture data. We compare Bayesian binning (BB) for segmentation of human full-body movement with human responses, which were assessed in an interactive video segmentation paradigm.

BB is a method for modelling data with a totally ordered structure, e.g. time series, by piecewise defined functions. Its advantages include automatic complexity control, which translates into automatic determination of the number and length of the segments in our context. BB was originally developed for density estimation of neural data and their subsequent information theoretic evaluation [8]. It was later generalised for regression of piecewise constant functions [14] and further applications in neural data analysis [10,9]. Concurrently, a closely related formalism for dealing with multiple change point problems was developed in [11].

We give a concise description of the data recordings in section 2, since these data have not been published before. The psychophysical experiments and their results are described in section 3. We use BB for the segmentation of joint angle data obtained by motion capture in section 4. Furthermore, we show how to use BB with non-constant observations models in the bins. In section 5 we present the segmentations achieved by BB and compare them with the psychophysical results. Finally, we discuss the advantages and limitations of our approach and give an outlook for further investigations in section 6.

2 Kinematical Data

The action streams we studied are solo Taekwondo activities performed by ten internationally successful martial artists. Each combatant performed the same fixed sequence of 27 kicks and punches, forming a so called *hyeong*. A complete hyeong had a full length of about 40 seconds. The kinematical data was obtained by motion capture using a VICON 612 system with 11 cameras, obtaining the 3D positions of 41 passively reflecting markers attached to the combatants' joints and limbs with a 3D reconstruction error of below 1 mm and at a sampling frequency 120 Hz.

The use of the obtained kinematical data was twofold. First, joint angle trajectories were computed from a hierarchical kinematic body model (skeleton) which was fitted to the original 3D marker positions. The rotations between adjacent segments of this skeleton were described by Euler angles, defining flexion, abduction and rotations about the connecting joint (e.g. [19,22]). Second, from the derived joint angle trajectories we created movie clips showing computer animations of the Taekwondo movements. Those videos served as stimuli in our psychophysical experiment to obtain human segmentation ratings.

3 Human Action Segmentation

To test the validity of the segmentation results obtained using our algorithmic approach we conducted a psychophysical study to achieve action segmentations of human observers.

Stimulus Preparation: short video clips displaying the Taekwondo movements served as stimuli in the psychophysical paradigm. Volumetric grey puppet constructed from simple geometric shape elements were animated with the combatants' movements. An illustration of the puppet's appearance is shown in fig. (1)A. To avoid stimuli of uncomfortable length each complete hyeong was split into five sub-sequences of comparable length each containing between three and eight separate Taekwondo actions. We restricted the number of stimuli in the experiment in order to prevent the participants from experimental fatigue and frustration. Thirty video clips corresponding to the complete hyeong of six representative combatants served as stimuli in this study. The puppet within the stimuli subtended approximately 4 x 8.6 degrees of visual angle and was presented on a computer screen viewed from a distance of 50 cm.

Experimental Procedure: the experiment started with a training phase in which participants familiarised themselves with the procedure. Five video clips corresponding to the complete hyeong of one combatant were only shown during this training phase. The remaining 25 movies served as test stimuli. Each was shown three times resulting in 75 segmentation trials per subject. Human observers watched video clips displaying the Taekwondo movements animated as puppets and segmented the complete hyeong into actions. In every trial the current video clip was first presented twice to enable the subjects to acquaint themselves with this action sequence. During the third presentation of the animation participants segmented the action sequence by pressing a marked key on the keyboard at each point which they perceived as the endpoint of one single, separable action. The segmentation was then replayed and if the participants felt insecure about their responses they had the opportunity to correct themselves up to two times. Noteworthy, it was completely left to the participants' own judgement what exactly defines *one single action* and the corresponding *endpoint*. They never received feedback regarding their segmentation neither during training nor during testing.

Participants: thirteen naïve subjects (mean age 26 years 6 month, ranging from 21 years 11 month to 38 years 11 month, 10 female) participated in this study. None of them had experience performing Taekwondo or other sports related to martial arts. All participants had normal or corrected-to-normal vision, gave informed written consent and were paid for their participation.

Segmentation Results of Human Observers: the results of the human action segmentation for the hyeong of one representative Taekwondo combatant are shown in fig. (1)B. Each single black dot represents one key press indicating the perception of an intersection between two Taekwondo actions. The 39 rows correspond to the three segmentation repetitions of each of the thirteen participants. The lack of feedback and explicit definitions of a single action resulted in differences in the interpretation of *one separable action* between participants. Most participants (11) tended to divide the action sequences on a very fine-grained level resulting in many endpoints (mean number of segmentation points = 25.36, standard error = 2.49). Though, two subjects concentrated only on the coarse separation of the hyeong by setting only 5 respectively 8 segmentation points. In direct comparison with the timing of the 27 expected endpoints as defined by the Taekwondo combatants themselves (see coloured bars in fig. (1)B), naıve participants placed 47.1% of the segmentations (standard error 5%) accurately within a very tight time window of +/-250 ms around the expected time point. Although a hit rate of 47.1% seems low at a first glance the following has to be taken into account. First, the complete hyeong was presented in 5 video clips. Each video ended at one expected endpoint. Some participants did not indicate a segmentation point at the video boundaries because they thought it would be redundant. Second, human observers tended to set the endpoint of the actions slightly to early compared to the expected endpoints. This happened especially when the combatant remained still for a longer time after he has completed an action. Shifting the accuracy time window from +/-250 ms to -380 ms to +120ms and excluding the video boundaries from the hit rate analysis results in a hit rate of 56.6%. Despite the slightly shift in timing compared to the expected time points the set segmentations are consistent across subjects (see fig. (1)C and D for the segmentation density). These results are in accordance with previous findings about the agreement of human raters on boundary placing in movement sequences [7,18,25].

4 Bayesian Binning for Action Segmentation

We now briefly specifiy the BB model used for the segmentation of joint angle data. The following sections describe the prior over bin boundaries (section 4.1) and the used observation models (section 4.2). The algorithmic details of evaluating posterior expectations are only outlined schematically, they are detailed in [8].

4.1 The Bin Boundary Prior

Our aim is to model a time series D in the time interval $[t_{min}, t_{max}]$. We want to be able to draw conclusions about change point estimates from small amounts of

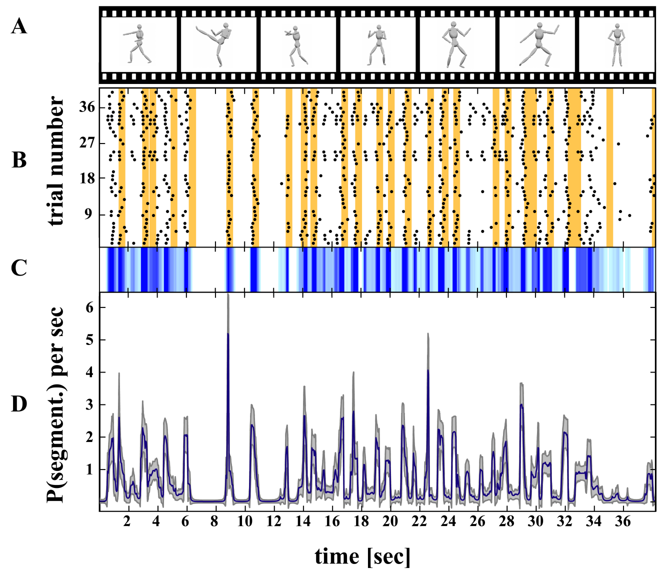

Fig. 1. Human Action Segmentation. A) Illustration of Stimuli. Snapshots taken from the stimuli videos showing the custom-built volumetric grey puppet performing different Taekwondo kicks and punches. B) Subjective Segmentation Points. Black dots correspond to the intersection points participants perceived between two Taekwondo actions. Results for the individual participants are shown row-wise. The coloured areas mark the time windows +/- 250 ms around the expected endpoints as defined by experts. C) Predictive Segmentation Density I. Predictive segmentation density estimated from human key presses. Estimation was carried out by Bayesian binning with a Bernoulli-Beta observation model (see section 4.2). Color saturation indicates density (darker = higher). D) Predictive Segmentation Density II. Same density as in C). Blue line represents the predictive segmentation density using Bayesian binning, the shaded grey area indicates ± one posterior std. dev.

data, let the model complexity be driven by D and handle D corrupted by (large amounts of) noise. We therefore take a Bayesian approach. Let $[t_{min}, t_{max}]$ be discretised into T contiguous intervals of duration $\Delta t = (t_{max} - t_{min})/T$, such that interval j is $[j \cdot \Delta t, (j+1) \cdot \Delta t]$ (see fig. (2)). Assume that Δt is small enough so that all relevant features of the data are captured in the discretised version of D. We model the generative process of D by $M+1$ non-overlapping, contiguous bins, indexed by m and having *inclusive* upper boundaries $k_m \in \{k_m\}$. The bin m therefore contains the time interval $T_m = (\Delta t\, k_{m-1}, \Delta t\, k_m]$. Let D_m be that part of the data which falls into bin m. We presuppose that the probability of D given $\{k_m\}$ can be factorised as

$$P(D|\{k_m\}, M) = \prod_{m=0}^{M} P(D_m|k_{m-1}, k_m, M) \tag{1}$$

where we additionally define $k_{-1} = -1$, $k_M = T - 1$.

$$\begin{aligned} P(D|\{k_m\}, M) &= P(D_0|k_{-1}, k_0)P(D_1|k_0, k_1)\dots P(D_M|k_{M-1}, k_M) \\ &= \textstyle\prod_{m=0}^{M} P(D_m|k_{m-1}, k_m) \end{aligned}$$

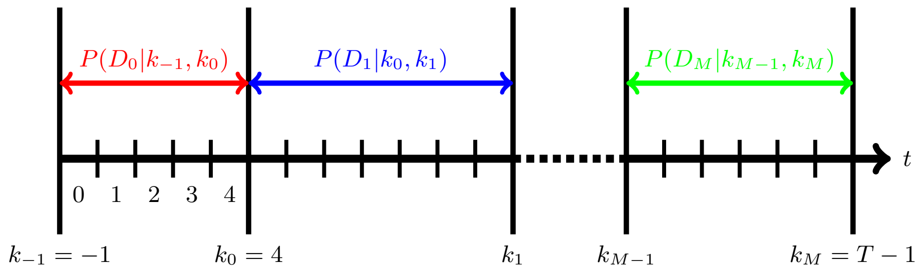

Fig. 2. Exemplary binning of a discrete time series of length T into $M + 1$ contiguous, non-overlapping bins with (inclusive) upper bin boundaries $k_m \in \{k_m\}$. Within each bin m, the observation model for data D is given by $P(D_m|k_{m-1}, k_m)$, where D_m is that part of the data which falls into bin m. We assume that the data are independent across bins given the $\{k_m\}$and M.

Prior on $\{k_m\}$: since we have no preferences for any bin boundary configuration (other than $m' < m \Rightarrow k_{m'} < k_m$), our prior is

$$P(\{k_m\}|M) = \binom{T-1}{M}^{-1} \tag{2}$$

where $\binom{T-1}{M}$ is just the number of possibilities in which M ordered bin boundaries can be distributed across $T - 1$ places (bin boundary M always occupies position $T - 1$, hence there are only $T - 1$ positions left).
Prior on M: we have no preference for any number of bin boundaries (which controls the model complexity). Thus, we let

$$P(M) = \frac{1}{T} \tag{3}$$

since the number of bin boundaries M must be $0 \leq M \leq T - 1$.
For temporal segmentation, the most relevant posterior is that of the $\{k_m\}$ for a given M:

$$P(\{k_m\}|D, M) = \frac{P(D|\{k_m\}, M)P(\{k_m\}|M)}{P(D|M)} \tag{4}$$

This requires the evaluation of $P(D|M)$:

$$P(D|M) = \sum_{k_0=0}^{k_1-1} \sum_{k_1=1}^{k_2-1} \dots \sum_{k_{M-1}=M-1}^{T-1} P(D|\{k_m\}, M) \tag{5}$$

which appears to be $\mathcal{O}(T^M)$ since it involves M sums of length $\mathcal{O}(T)$. However, exploiting the form of $P(D|\{k_m\}, M)$ (eqn. (1)) allows us to "push sums" past all factors which do not depend on the variable being summed over:

$$\begin{aligned} P(D|M) &= \sum_{k_0=0}^{k_1-1} \sum_{k_1=1}^{k_2-1} \dots \sum_{k_{M-1}=M-1}^{T-1} \prod_{m=0}^{M} P(D_m|k_{m-1}, k_m) \\ &= \sum_{k_0=0}^{k_1-1} P(D_0|k_{-1}, k_0) \sum_{k_1=1}^{k_2-1} P(D_1|k_0, k_1) \dots \\ &\quad \dots \sum_{k_{M-1}=M-1}^{T-1} P(D_M|k_{m-1}, k_M) \end{aligned} \tag{6}$$

Now each sum over $\mathcal{O}(T)$ summands has to be evaluated $\mathcal{O}(T)$ times for the possible values of the upper summation boundary. Since there are M sums, this calculation has complexity $\mathcal{O}(MT^2)$, which is feasible. This way of computing $P(D|M)$ is an instance of the **sum-product** algorithm [16]. As detailed in [8], the expectation of any function of the model parameters (e.g. bin boundary position, bin width or probability of a bin boundary at a given point in time) can be evaluated with a similar approach, given that the function depends only on the parameters of one bin for any given $\{k_m\}$.

4.2 Observation Models $P(D|\{k_m\})$ for Action Streams

We employed two different observation models. For both, conjugate priors can be specified on their parameters which allow for an evaluation of expectations and marginal probabilities in closed form. This enables us to compute efficiently the marginal probability of the data given the number of bin boundaries (eqn. (6)).

Bernoulli-Beta: human segmentation events (i.e. key presses by observers) are binary. It is therefore natural to model these data with a Bernoulli process having a conjugate Beta prior (one per bin). This is analogous to modelling neural spike trains with BB [10]. Thus, for a segmentation event $e(t) \in D$ at time t in bin m, i.e. $t \in T_m$ we have

$$P(e(t)|t \in T_m) = P_m \tag{7}$$

$$p(P_m) = \mathrm{B}(P_m; \gamma_m, \delta_m) \tag{8}$$

where $\mathrm{B}(P_m; \gamma_m, \delta_m)$ is the Beta density with parameters γ_m, δ_m (see e.g. [4]).

Multivariate Gaussian with Polynomial Time-Dependence: joint angles are real numbers in $[-\pi, \pi)$. We could thus employ a multivariate von-Mises density or generalisations thereof [17]. Instead, we chose to model joint angles with a multivariate Gaussian whose mean has a polynomial time dependence, because its conjugate priors are tractable analytically. The exponential family conjugate prior on the mean μ and the precision matrix $\mathbf{P}$ (inverse covariance) is then given by an extended Gauss-Wishart density (see e.g. [4]). Let $\boldsymbol{X}_t \in D$ be a L-dimensional vector of joint angles at time $t \in T_m$, and S be the chosen polynomial order. Let $t_m = \Delta t\ k_{m-1}$ be the start time of bin m. Then

$$p(\boldsymbol{X}_t | t \in T_m) = \mathcal{N}(\boldsymbol{X}(t); \boldsymbol{\mu}_m, \mathbf{P}_m^{-1}) \tag{9}$$

$$p(\mathbf{P}_m | \nu_m, \mathbf{V}_m) = \mathcal{W}(\mathbf{P}_m; \nu_m, \mathbf{V}_m) \tag{10}$$

$$\boldsymbol{\mu}_m = \sum_{i=0}^{S} \boldsymbol{a}_{i,m} (t - t_m)^i \tag{11}$$

The $\boldsymbol{a}_m = (\boldsymbol{a}_{i,m})$ are the polynomial coefficients in bin m. Note that this vector has $(S+1) \cdot L$ components. $\mathcal{N}(\boldsymbol{X}, \boldsymbol{\mu}, \boldsymbol{\Sigma} = \mathbf{P}^{-1})$ is a multivariate Gaussian density in $\boldsymbol{X}$ with means $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. $\mathcal{W}(\mathbf{P}; \nu, \mathbf{V})$ is a Wishart density in $\mathbf{P}$ with ν degrees of freedom and scale matrix $\mathbf{V}$. To construct a prior which is conjugate to the likelihood (eqn. 9), we choose a vector $\boldsymbol{\alpha}_m = (\boldsymbol{\alpha}_{i,m})$ with $(S+1) \cdot L$ components, which are the biases on $\boldsymbol{a}_m$. Furthermore, we introduce a symmetric, positive (semi-)definite $(S+1) \times (S+1)$ matrix $\mathbf{B}_m$, which contains the concentration parameters on $\boldsymbol{a}_m$. The prior on $\boldsymbol{a}_m$ given $\mathbf{P}_m$ is then a multivariate Gaussian density

$$p(\boldsymbol{a}_m | \boldsymbol{\alpha}_m, \mathbf{B}_m, \mathbf{P}_m) = \mathcal{N}(\boldsymbol{a}_m; \boldsymbol{\alpha}_m, \mathbf{Q}_m^{-1}) \tag{12}$$

where the $(S+1)L \times (S+1)L$ matrix $\mathbf{Q}_m$ is obtained by block-wise multiplication of the entries $\mathbf{B}_{m,i,j}$ of $\mathbf{B}_m$ with $\mathbf{P}_m$:

$$\mathbf{Q}_m = \begin{pmatrix} \mathbf{B}_{m,0,0}\mathbf{P}_m & \cdots & \mathbf{B}_{m,0,S}\mathbf{P}_m \\ \vdots & \ddots & \vdots \\ \mathbf{B}_{m,S,0}\mathbf{P}_m & \cdots & \mathbf{B}_{m,S,S}\mathbf{P}_m \end{pmatrix} \tag{13}$$

Lengthy but straightforward calculations confirm that the product of the Gaussian (eqn. 12) with the Wishart (eqn. 10) does indeed constitute a conjugate prior on the likelihood given by eqn. 9. We omit these calculations here for brevity. Since the prior is conjugate with a known normalisation constant (i.e. that of the Gaussian times the Wishart), the marginal likelihood of the data in each bin can be computed, and thus Bayesian binning can be applied with this observation model.

5 Results

We applied BB to joint angle trajectories of shoulder and elbow angles, and combinations thereof, to determine the segmentation densities. Fig. (3), left,

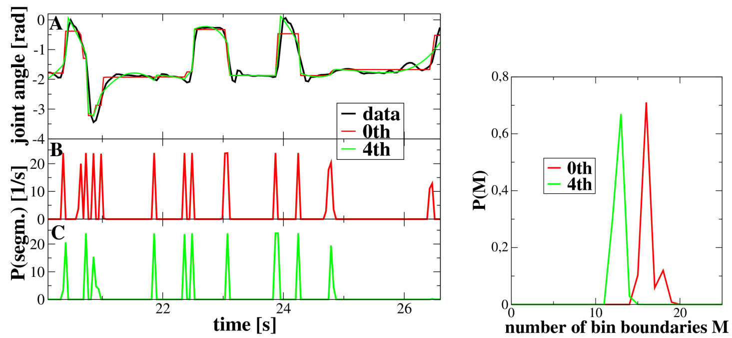

Fig. 3. *Left*: A: fitting a part of a joint angle trajectory with Bayesian binning. Joint angles have not been wrapped around at $-\pi$ to avoid creation of artificial segmentation points. Red lines shows predictive joint angles with a 0th order (i.e. bin-wise constant) observation model (see section 4.2), green lines show predictions from 4th order observation model. B,C: predictive segmentation densities for these two observation models. The 4th order model needs less segmentation points than the 0th order model, and also yields a more faithful fit of the joint angle trajectory. *Right*: posterior distribution of the number of bin boundaries M. The M-posterior of the 4th order observation model peaks at smaller values of M than the 0th order model, indicating that the 0th order model requires more bins to fit the data well. Note that both peaks are far from the maximum $M = 171$, i.e. over-fitting is avoided.

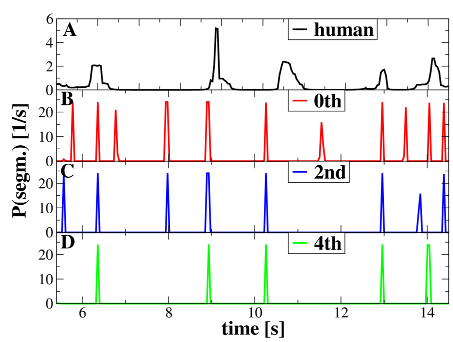

Fig. 4. Comparison of human segmentation densities with those obtained by Bayesian binning. Shown is an interval with a few, relatively clear segmentation points and good agreement between human subjects. Note that the human segmentation density (panel A) peaks usually closely to a peak in the density obtained by Bayesian binning. The 0th order model (panel B) predicts more segmentation points than the higher-order models (panels C,D), and the higher-order models are in better agreement to the human segmentation, both in number and location of the segmentation points.

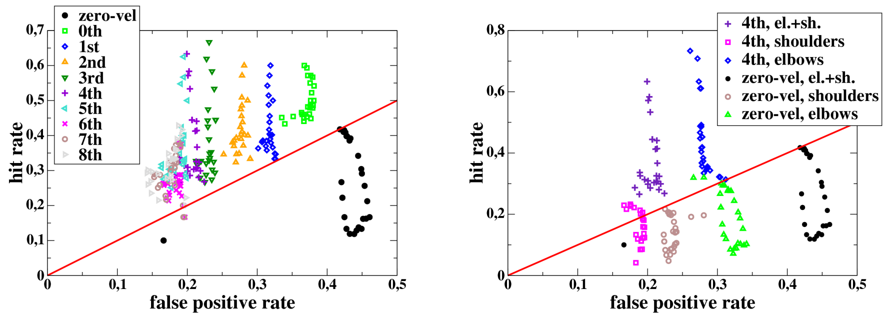

Fig. 5. Hit rate performance analysis. Red line: line of no discrimination. *zero-vel*: segmentation based on zero-crossings of angular velocity. *Left*: comparison between observation models of different polynomial orders (S in eqn. (11)). Elbow and shoulder angles were jointly segmented. An observation model with $S \in \{3, 4, 5\}$ offers the best compromise between a high hit rate and a low false positive rate. *Right*: performance dependence on joint angles for a model with $S = 4$ and the *zero-vel* segmentation. We segmented either elbow angles only, or shoulder angles only, or both together (*el.+sh.* in the legend). The latter yields the best segmentation results. For details, see text.

panel A shows the predictive trajectories of an elbow angle computed with a 0th order and a 4th order observation models. Both models fit the data well, but the 4th order model yields a better fit while needing less bin boundaries, as indicated by the M posterior in fig. (3), right. Panels B and C in fig. (3), left, depict the predicted segmentation densities, showing where the 0th order model inserts the additional boundaries compared to the 4th order model.

Fig. (4) shows comparisons between human and BB segmentation densities. Note that the human (panel A, in fig. (1)) and the BB segmentation densities peak usually in close temporal vicinity. The 0th order model (panel B) over-segments, this over-segmentation is already reduced for the 2nd order model (panel C) and virtually gone for the 4th order model (panel D).

For a more quantitative evaluation of the agreement between human subjects and BB, we performed a hit rate/false positive rate analysis. Hits and false positives were computed by thresholding the segmentation densities (see fig. (4) and fig. (1)D), thereby yielding a binary segmentation event signal for each point in time. Every human segmentation event in a 400 ms accuracy window after a BB segmentation event was counted as a hit, the lack of a human segmentation event in this window counted as a false positive. This choice of accuracy window length was motivated by the comparison between naïve and expert human observers presented in section 3. We varied the threshold between 0.1 and 3.0 to obtain the data shown in fig. (5). As a simple baseline for comparison, we also computed segmentation points by searching for zero-crossings of angular velocity (*zero-vel* in fig. (5)). Using angular velocity zero-crossings as a baseline method was inspired by [15]. The zero-crossing search was carried out by computing local

(300 ms window) parabolic fits to the joint angle data at every point in time, and checking whether the 1st order coefficient of the fit was close to 0.

Fig. (5), *left* shows a comparison between observation models of different polynomial orders (S in eqn. (11)). Elbow and shoulder angles were jointly segmented. Observation models with $S \in \{3, 4, 5\}$ offer the best compromise between a high hit rate and a low false positive rate. For all orders S, BB is a lot better than the baseline method. Fig. (5), *right* depicts the performance dependence on joint angles for a model with $S = 4$ and the *zero-vel* segmentation. We segmented either elbow angles only, or shoulder angles only, or both together. Segmenting both angles together yields the best segmentation results.

The fact that models with $S \in \{3, 4, 5\}$ provide a better match than the lower orders indicates that humans employ (the visual equivalent of) angular acceleration discontinuities, rather than discontinuities in angular velocities when segmenting action streams. This agrees with the 'minimum jerk' hypothesis [12].

6 Conclusion

In this paper, we have shown how to extend Bayesian binning by piecewise polynomial observation models and demonstrated its usefulness for action stream segmentation. Furthermore, we have created a ground truth data set for the evaluation of machine segmentation methods against human observers. Comparing our method to other automatic motion segmentation approaches, e.g. [3], will be interesting future work.

Previously, trajectories were successfully fitted with parabolic pieces [20]. We showed that higher orders yield a yet better agreement with human psychophysical data. One might also consider using a hidden Markov model (HMM) in each bin. The BB prior might be a feasible way of switching between HMMs, which were used for action segmentation in [13].

Our approach does not yet include context information into the segmentation process, we utilised only purely kinematic information. [25] reports that humans use context information for segmentation tasks when such is available, and rely increasingly on kinematics when context is reduced. Thus, including context can be expected to improve performance further.

Acknowledgements

This work was supported by EU projects FP7-ICT-215866 SEARISE, FP7-249858-TP3 TANGO, FP7-ICT-248311 AMARSi and the DFG. We thank Engelbert Rotalsky, Hans Leberle and the Taekwondo Unions of Nordrhein-Westfalen and Baden-Württemberg for cooperation on the data acquisition. We thank S. Cavdaroglu, W. Ilg and T. Hirscher for their help with data collection and post-processing.

References

1. Agam, Y., Sekuler, R.: Geometric structure and chunking in reproduction of motion sequences. Journal of Vision 8(1), 1–12 (2008)
2. Albu, A.B., Bergevin, R., Quirion, S.: Generic temporal segmentation of cyclic human motion. Pattern Recognition 41(1), 6 – 21 (2008)
3. Barbič, J., Safonova, A., Pan, J.Y., Faloutsos, C., Hodgins, J.K., Pollard, N.S.: Segmenting motion capture data into distinct behaviors. In: Proceedings of Graphics Interface 2004. pp. 185–194. GI '04, Canadian Human-Computer Communications Society, School of Computer Science, University of Waterloo, Waterloo, Ontario, Canada (2004), `http://portal.acm.org/citation.cfm?id=1006058.1006081`
4. Bishop, C.M.: Pattern Recognition and Machine Learning. Springer (2007)
5. Bruderlin, A., Williams, L.: Motion signal processing. In: SIGGRAPH. pp. 97–104 (1995)
6. Chen, W., Zhang, J.J.: Parametric model for video content analysis. Pattern Recognition Letters 29(3), 181 – 191 (2008)
7. Dickman, H.R.: The perception of behavioral units. In: Barker, R.G. (ed.) The stream of behavior, pp. 23–41. Appleton-Century-Crofts, New York (1963)
8. Endres, D., Földiák, P.: Bayesian bin distribution inference and mutual information. IEEE Transactions on Information Theory 51(11), 3766 – 3779 (2005)
9. Endres, D., Oram, M.: Feature extraction from spike trains with bayesian binning: latency is where the signal starts. Journal of Computational Neuroscience 29, 149–169 (2010)
10. Endres, D., Oram, M., Schindelin, J., Földiák, P.: Bayesian binning beats approximate alternatives: estimating peri-stimulus time histograms. In: Platt, J., Koller, D., Singer, Y., Roweis, S. (eds.) Advances in Neural Information Processing Systems 20. MIT Press, Cambridge, MA (2008)
11. Fearnhead, P.: Exact and efficient bayesian inference for multiple changepoint problems. Statistics and Computing 16(2), 203–213 (2006)
12. Flash, T., Hogan, N.: The coordination of arm movements: an experimentally confirmed mathematical model. J. Neurosci. (5), 1688–1703 (1985)
13. Green, R.D.: Spatial and temporal segmentation of continuous human motion from monocular video images. In: Proceedings of Image and Vision Computing. pp. 163–169. New Zealand (2003)
14. Hutter, M.: Exact bayesian regression of piecewise constant functions. Journal of Bayesian Analysis 2(4), 635–664 (2007)
15. Ilg, W., Bakir, G., Mezger, J., Giese, M.: On the representation, learning and transfer of spatio-temporal movement characteristics. International Journal of Humanoid Robotics 1(4), 613–636 (2004)
16. Kschischang, F., Frey, B., Loeliger, H.A.: Factor graphs and the sum-product algorithm. IEEE Transactions on Information Theory 47(2), 498–519 (2001)
17. Marida, K.V., Jupp, P.E.: Directional Statistics. Wiley (2000)
18. Newtson, D., Engquist, G.: The perceptual organization of ongoing behavior. Journal of Experimental Social Psychology 12(5), 436 – 450 (1976)
19. Omlor, L.: New methods for anechoic demixing with application to shift invariant feature extraction. PhD in informatics, Universität Ulm. Fakultät für Ingenieurwissenschaften und Informatik (2010), urn:nbn:de:bsz:289-vts-72431
20. Polyakov, F., Stark, E., Drori, R., Abeles, M., Flash, T.: Parabolic movement primitives and cortical states: merging optimality with geometric invariance. Biol. Cybern. 100(2), 159–184 (2009)

21. Quirion, S., Branzan-Albu, A., Bergevin, R.: Skeleton-based temporal segmentation of human activities from video sequences. In: Proceedings WSCG'2005 - 13-th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision'2005 (2005)
22. Roether, C.L., Omlor, L., Christensen, A., Giese, M.A.: Critical features for the perception of emotion from gait. Journal of Vision 9(6), 1–32 (2009)
23. Shipley, T.F., Maguire, M.J., Brumberg, J.: Segmentation of event paths. Journal of Vision 4(8) (2004)
24. Wang, L., Hu, W., Tan, T.: Recent developments in human motion analysis. Pattern Recognition 36(3), 585 – 601 (2003)
25. Zacks, J.M., Kumar, S., Abrams, R.A., Mehta, R.: Using movement and intentions to understand human activity. Cognition 112(2), 201 – 216 (2009)