Spatiotemporal Deformable Prototypes for Motion Anomaly Detection

Robert Bensch
bensch@cs.uni-freiburg.de
Thomas Brox
brox@cs.uni-freiburg.de
Olaf Ronneberger
ronneber@cs.uni-freiburg.de

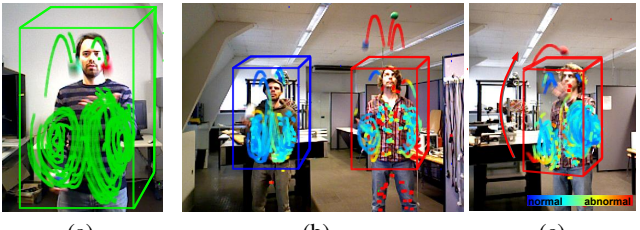
Department of Computer Science and BIOSS Centre for Biological Signalling Studies
University of Freiburg
Germany
http://lmb.informatik.uni-freiburg.de

Abstract

This paper presents an approach for motion-based anomaly detection, where a prototype pattern is detected and elastically registered against a test sample to detect anomalies in the test sample. The prototype model is learned from multiple sequences to define accepted variations. "Supertrajectories" based on hierarchical clustering of dense point trajectories serve as an efficient and robust representation of motion patterns. An efficient hashing approach provides transformation hypotheses that are refined by a spatiotemporal elastic registration. We propose a new method for elastic registration of 3D+time trajectory patterns that induces spatial elasticity from trajectory affinities. The method is evaluated on a new motion anomaly dataset and performs well in detecting subtle anomalies. Moreover, we demonstrate the applicability to biological motion patterns.

1 Introduction

An anomaly is generally a deviation from what is regarded as normal. Since there are no examples from which distinct features of the anomaly could be learned, anomaly detection cannot be modeled as a discriminative classification task. We must rather learn a precise



(a) (b) (c)

Figure 1: Motion anomaly detection in juggling patterns. (a) Motion pattern prototype: standard 3-ball cascade pattern. (b-c) Detection of local anomalies in context of prototype detections (*bounding boxes*) of different jugglers in different 3D poses. The anomaly score is plotted for supertrajectories from low/normal (*blue*) to high/abnormal (*red*).

generative model of normal patterns, for which we have examples, and detect anomalies as cases that are not sufficiently explained by this model. In this paper, we consider motion-based anomaly detection from video and present a new setting of anomaly detection. Compared to existing approaches in the literature, we formulate anomaly detection as the task of detecting subtle anomalies in the context of a well-defined, reproducible motion pattern, termed motion pattern prototype. We want to detect instances of this prototype and localize anomalies in its context. Anomaly detection is performed by reconstructing an unseen motion pattern by prototype placements. Subpatterns that remain poorly reconstructed are detected as abnormal. For reconstruction, we start with a robust detection followed by a spatiotemporal elastic registration of a deformable prototype. Our method copes with 3D+time data, where *3D+time* denotes motion trajectories in full 3D space, in contrast to 2D+depth. For detection, we allow for temporal shift and 3D spatial translation and rotation. 3D scaling is not modelled explicitly and therefore will be recognized anomalous in principle. However, elastic registration is able to compensate for scaling to some extent. A statistical prototype model is learned from training data. It defines the accepted spatiotemporal deformations and deviations. Fig. 1 illustrates our setting of anomaly detection with an example of juggling patterns. In this example, the prototype defines a standard 3-ball juggling pattern. It is robustly detected under various transformations. Deviations from the standard pattern are localized as anomalies. We stress that, in this paper, we are interested in anomalies in the motion pattern rather than the object appearance. For example, we do not want to detect an anomaly, if a "normal" motion pattern is performed but the person is wearing a different shirt. An overview of our method is given in Fig. 2. The experimental section evaluates the performance of our approach on a new motion anomaly dataset. Moreover, we demonstrate the relevance and general applicability of our method in experiments on biological motion patterns. While in the motion anomaly dataset the 3D+time data originates from Kinect 2D+depth data, the biological motion patterns stem from 3D volumetric microscopy recordings. We will make the new motion anomaly dataset and the code publicly available at http://lmb.informatik.uni-freiburg.de/resources/opensource/AnomalyDetection/.

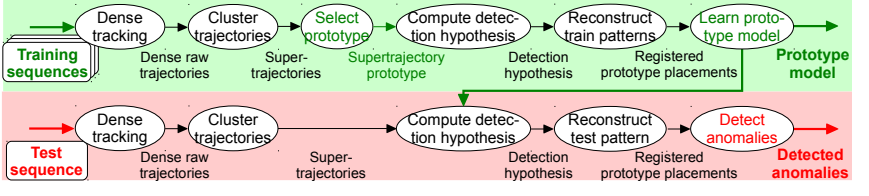


Figure 2: Overview of our approach.

2 Related Work

Many works in the anomaly detection literature consider the surveillance scenario in crowded scenes of people or traffic scenes. Usually a fixed scene and camera setting is assumed and absolute position information is the predominant feature. Commonly, a fixed spatial (and temporal) grid representation is used and local statistics of grid cells are learned. Anomalies are detected for cells in which the observed statistics deviate from the learned model, *e.g.* [12, 13, 14, 16]. However, in several scenarios the absolute position is not relevant for the characterization of anomalies. Consider cases where anomalies occur invariant to their

absolute position, or relative to a certain spatiotemporal context, or under spatio(-temporal) transformations, such as rotations or deformations. One type of approaches dealing with these cases relies on bag of invariant features. For instance, in [21] chaotic invariants are computed that are invariant to position and magnitude, and a global probabilistic scene model of normality is learned using Gaussian mixture models (GMMs). In [15] a social force model induces interaction forces that serve as input for a bag of words approach. With these approaches localization of anomalies is only possible indirectly by projecting anomalous features back to their spatiotemporal domain. Another type of approaches considers anomaly detection as a reconstruction task, in which anomalies remain as poorly reconstructed entities. Antić and Ommer [3] use the term "video parsing" for jointly explaining the foreground from normal training samples. In [9] a sparse reconstruction cost is used to perform sparse dictionary selection given an over-complete spatiotemporal basis. In [10] and [5, 6] the problem is formulated as matching against spatiotemporal segments in the training data, or composition from a database of patch ensembles, respectively. Reconstruction based approaches provide a more direct explanation of the test data by concrete instances from the training data. Furthermore, they can deal with few training samples, compared to approaches learning statistical models. The presented approach is most related to reconstruction based approaches. However, it is different in that, reconstruction is performed in a direct way by detection and elastic registration of a well-defined prototype. Furthermore, the presented approach deals with 3D+time data, while existing methods deal with 2D+time data.

3 Supertrajectory Representation

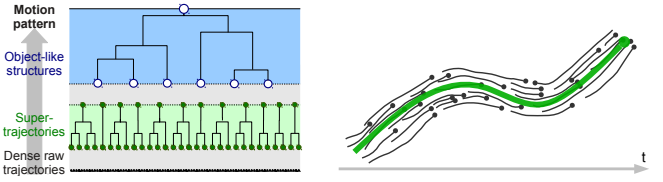
We suggest to represent motion patterns by "supertrajectories" describing the motion of local groups of similarly moving points, see Fig. 3(b). We initialize our representation by dense point trajectories [18]. These basic motion trajectories from tracked points constitute the lowest level of our representation and are denoted as raw trajectories. Trajectories are allowed to start and end at arbitrary points in time. We define motion patterns by two functions,

$$\mathbf{x}_{\text{raw}} : \Omega_{\text{raw}} \to \mathbb{R}^3 : (i_{\text{raw}}, t) \mapsto \mathbf{x}_{\text{raw}}(i_{\text{raw}}, t), \quad i_{\text{raw}} \in \{1, \ldots, N_{\text{raw}}\} \subset \mathbb{N}, t \in \mathbb{R}$$
$$w_{\text{raw}} : \Omega_{\text{raw}} \to \{0, 1\} : (i_{\text{raw}}, t) \mapsto w_{\text{raw}}(i_{\text{raw}}, t),$$

where $\Omega_{\text{raw}} \subset \mathbb{N} \times \mathbb{R}$ denotes the domain of raw trajectories i_{raw} and time t. The position is denoted by $\mathbf{x}_{\text{raw}}(i_{\text{raw}}, t)$ and the validity by $w_{\text{raw}}(i_{\text{raw}}, t)$. $w_{\text{raw}}(i_{\text{raw}}, t) = 1$ means that trajectory i_{raw} exists at time point t and the corresponding position is valid.

3.1 Hierarchical Clustering

Motion patterns often exhibit a natural hierarchical composition of subpatterns. We propose a clustering step to represent bundles of similarly moving points at different hierarchical levels by "supertrajectories", see Fig. 3. We build a hierarchical representation by agglomerative clustering [11], that iteratively groups trajectories in bottom-up manner. Low levels represent local trajectory bundles, while higher levels represent object-like structures, as shown in Fig. 3(a). Hierarchical clustering is defined by a distance metric between elements and a linkage criterion that defines distances between sets of elements. We define pairwise



(a) (b)

Figure 3: Supertrajectory representation. (a) Agglomerative clustering of motion trajectories yields a hierarchical representation of motion patterns. (b) A supertrajectory (*green*) provides a smooth, robust and efficient representation of a bundle of raw trajectories (*black*).

distances between trajectories (raw trajectories or supertrajectories) by their maximum Euclidean distance in the overlapping time window:

$$d(i,j) = \max_{t \in \mathbb{R}} (w(i,t) \cdot w(j,t) \cdot \|\mathbf{x}(i,t) - \mathbf{x}(j,t)\|). \tag{1}$$

The distance between temporally non-overlapping trajectories, i.e. $w(i,t) \cdot w(j,t) = 0, \forall t \in \mathbb{R}$, is defined as $d(i,j) = +\infty$. For bottom-up grouping we apply centroid linkage [11], that defines distances between sets of elements by their centroids. Here, the centroid of a set of raw trajectories is computed as the supertrajectory (described in the next section). Pairs of temporally non-overlapping trajectories are not evaluated during linkage. The resulting hierarchical cluster tree can be cut at arbitrary levels to obtain 1 to N_{raw} clusters. By splitting the hierarchy at a certain level, we obtain "supertrajectories" (the clusters at the split level).

3.2 Supertrajectories

As illustrated in Fig. 3(b), supertrajectories provide a smooth, robust and efficient representation of the dominant motion of a bundle of raw trajectories. We denote the set of raw trajectories that form one supertrajectory by $\mathcal{X}_i \subset \{1,\ldots,N_{\text{raw}}\}$, where $i \in \{1,\ldots,N_{\text{super}}\}$. A supertrajectory is computed by averaging the positions of all grouped raw trajectories at each time point. Analogous to raw trajectories, the position is denoted by a function $\mathbf{x}$ and the validity is denoted by a function w, on the domain $\Omega_{\text{super}} \subset \mathbb{N} \times \mathbb{R}$. Details are given in the supplementary material in Sec. 1.1.

4 Detection and Elastic Registration of Motion Patterns

We aim for reconstructing a whole test pattern by prototype placements. Based on that, we detect anomalies in the context of the prototype. Prototype detection shall be invariant to temporal shift and 3D spatial translation and rotation, as well as spatiotemporal deformations. Below, we describe how a motion pattern prototype, represented by supertrajectories, can be efficiently detected and elastically registered to the underlying test pattern. The prototype pattern is denoted by the supertrajectories $\mathbf{x}^a$. The test sequence is represented by the supertrajectories $\mathbf{x}^b$.

Detection Hypotheses To efficiently detect a prototype pattern in a new test sequence we modified the hashing approach [20] to deal with our spatiotemporal setting. For details we

point to Sec. 2.1 in the supplementary material. The output of the algorithm is a number of rigid transformation hypotheses. We parameterize them as a temporal shift t_{shift} and a spatial rigid transformation $\mathbf{T}$ with 3×3 rotation matrix R and translation $\mathbf{b}$.

Trajectory Association Function To perform registration, we have to estimate the correspondences and the transformation. We define trajectory correspondences by an association function $\sigma : \Omega_b \to \mathbb{N} : (i_b, t_b) \to i_a$, that assigns test trajectory i_b of pattern $\mathbf{x}^b$ at time point t_b to a prototype trajectory i_a of pattern $\mathbf{x}^a$. The assignment is undefined $\sigma(i_b, t_b) := 0$, if no temporally warped prototype trajectory i_a is valid at time t_b. Trajectory association is allowed to change in time, which is important, for instance, when a temporally long trajectory corresponds to two temporally consecutive trajectories.

4.1 Rigid Pre-Alignment

The transformation of each detection hypothesis is refined using a scheme similar to the iterative closest point algorithm [4]. The energy to be minimized is the sum of squared distances of all points of test pattern $\mathbf{x}^b$ to the associated points in pattern $\mathbf{x}^a$:

$$E_{\text{data}}(\mathbf{T}, t_{\text{shift}}, \sigma) = \sum_{\substack{(i_b,t_b)\in\Omega_b \\ w(i_b,t_b)=1 \\ \sigma(i_b,t_b)\neq 0}} \Psi\left(\left\| \mathbf{T}(\mathbf{x}^a(A[t_{\text{shift}}, \sigma](i_b, t_b))) - \mathbf{x}^b(i_b, t_b) \right\|^2 \right) + \sum_{\substack{(i_b,t_b)\in\Omega_b \\ w(i_b,t_b)=1 \\ \sigma(i_b,t_b)=0}} d^2_{\text{undef}}. \tag{2}$$

For easier notation, we define an association function $A[t_{\text{shift}}, \sigma] : \Omega_b \to \Omega_a : (i_b, t_b) \to (i_a, t_a)$, that assigns test trajectory i_b at time point t_b to prototype trajectory i_a at time point t_a. The function depends on the temporal shift t_{shift} and the trajectory association function σ. To be robust against outliers, we apply a truncated squared norm denoted by function $\Psi(d^2) = d^2$ for $d \leq d_{\max}$ and $\Psi(d^2) = d^2_{\max}$, otherwise. In addition, unassociated points, i.e. $\sigma(i_b, t_b) = 0$, are penalized with d_{undef}. The energy is minimized with respect to the rigid transformation parameters $\mathbf{T}$ and t_{shift}, and association function σ.

4.2 Elastic Registration

After rigid pre-alignment we perform a spatiotemporal elastic registration. The elastic transformation is parameterized by a spatial deformation function $\mathbf{u}(i,t) : \Omega_a \to \mathbb{R}^3$ and a temporal warping $\tau(i,t) : \Omega_a \to \mathbb{R}$, such that $\mathbf{x}'(i,t) = \mathbf{x}(i, t - \tau(i,t)) + \mathbf{u}(i, t - \tau(i,t))$ and $w'(i,t) = w(i, t - \tau(i,t))$ accordingly. The data term is defined analogous to Eq. 2, where the rigid transformations $(\mathbf{T}, t_{\text{shift}})$ get replaced by deformation and temporal warping $(\mathbf{u}, \tau)$.

Furthermore, we formulate the following smoothness assumptions: Firstly, the elastic transformation should be smooth both spatially (across trajectories) and temporally (along trajectories). Secondly, the assignment function should be temporally smooth as well. To this end, we formulate the total energy

$$E(\mathbf{u}, \tau, \sigma) = E_{\text{data}}(\mathbf{u}, \tau, \sigma) + \alpha_{\text{spatial}} E_{\text{spatial}}(\mathbf{u}, \tau) + \alpha_{\text{temp}} E_{\text{temp}}(\mathbf{u}, \tau) + \alpha_{\text{assign}} E_{\text{assign}}(\sigma). \tag{3}$$

The elastic coupling within the prototype pattern is described by the smoothness energy E_{spatial} on the spatiotemporal deformation functions for each pair of prototype trajectories. We define spatial smoothness across trajectories, such that the elastic coupling between trajectories $C(i,j) = \exp(-d(i,j)^2 / 2r^2)$ depends on the pairwise distances $d(i,j)$ (Eq. 1),

where r is a user selected distance (we used $r = 10$cm in our experiments). This results in strong coupling between similar trajectories (similar moving points) and weak coupling between dissimilar trajectories, that therefore can be transformed rather independently. A temporal smooth transformation along trajectories is enforced by the energy term E_{temp}. Moreover, a temporally smooth assignment is preferred by the smoothness term E_{assign} that penalizes temporal assignment changes from one to another trajectory by their pairwise distance $d(i, j)$. In this way, assignment changes between dissimilar trajectories get strong penalization. The formulation of the data term and the smoothness terms is described in detail in Sec. 2.2 in the supplementary material.

4.3 Energy Optimization

We found an approximate solution for minimizing the total energy, for both the rigid pre-alignment and the elastic registration, by alternating optimization of the transformation and the assignment σ. Given a fixed transformation the assignment function σ minimizing E_{assign} can be obtained by exact inference and computed efficiently by dynamic programming. The rigid transformation $\mathbf{T}$ and the temporal shift t_{shift} are found by a Procrustes algorithm [19] using all point correspondences with distance $d < d_{\max}$. The elastic transformation $(\mathbf{u}, \tau)$ minimizing the total energy in Eq. 3 is obtained by L-BFGS [8] optimization. Alternating optimization is repeated until the estimated transformation converges. Both parts of the optimization can be solved globally optimal, due to the convex energies.

5 Motion Anomaly Detection

5.1 Learning a Spatiotemporal Deformable Prototype Model

First a concrete motion pattern is selected as the spatiotemporal prototype which represents a clean and segmented instance of the "normal" motion pattern of interest, see Fig. 1(a) and suppl. Video 1. The "normal" variations of spatiotemporal deformation and remaining deviations observed in training sequences are learned and together with the prototype build the prototype model. The complete training pipeline is shown in Fig. 2(top row). For learning the prototype model, the prototype pattern is detected in training sequences and D detections are selected. For each detection d, the prototype pattern is rigidly pre-aligned and elastically registered. This results in rigid transformation parameters $((R_d, \mathbf{b}_d), t_{\text{shift},d})$ and elastic transformation parameters $(\mathbf{u}_d, \tau_d)$. Elastic spatial transformation parameters are transformed to the prototype coordinate system by $\mathbf{u}'_d = R_d^{-1}\mathbf{u}_d$. We build a statistical model that captures:

Global spatiotemporal deformations and data fitting costs after registration For both, we define bounds for validating prototype registrations.

Residual distances remaining after elastic registration For each prototype pattern point the obtained residual distances with associated training pattern points are locally aggregated and learned using a Gaussian residual model.

For more details see Sec. 3.1 in the supplementary material.

5.2 Reconstruction by Prototype Placements

For reconstructing a whole test pattern by prototype placements we apply a greedy search algorithm. It iteratively finds best placements of prototype patterns into the test pattern.

Candidate placements are obtained from prototype detections. The algorithm stops, if the test pattern is reconstructed completely, or if no candidates remain that can reconstruct parts of the unreconstructed test pattern. Details of the algorithm are given in Sec. 3.2 of the supplementary material.

Pointwise Anomaly Score We compute an anomaly score for each test pattern point expressing how much it deviates from the prototype model. The pointwise anomaly score is computed by taking the minimum residual distance to all registered prototype patterns and applying the locally learned residual model from the associated prototype pattern point.

Framewise Anomaly Score We map the anomaly scores of all trajectory points within one video frame to a framewise anomaly score by computing the maximum. The maximum measure is better suited for detecting local fine-grained anomalies, compared to the average measure, which is sufficient for detecting global anomalies. A sample anomaly profile is plotted in Fig. 4.

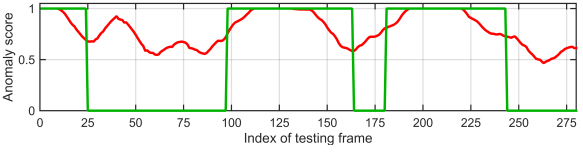
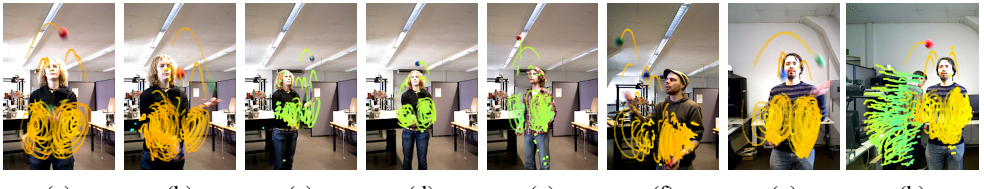


Figure 4: Anomaly profile (*red*). Ground truth profile (*green*).

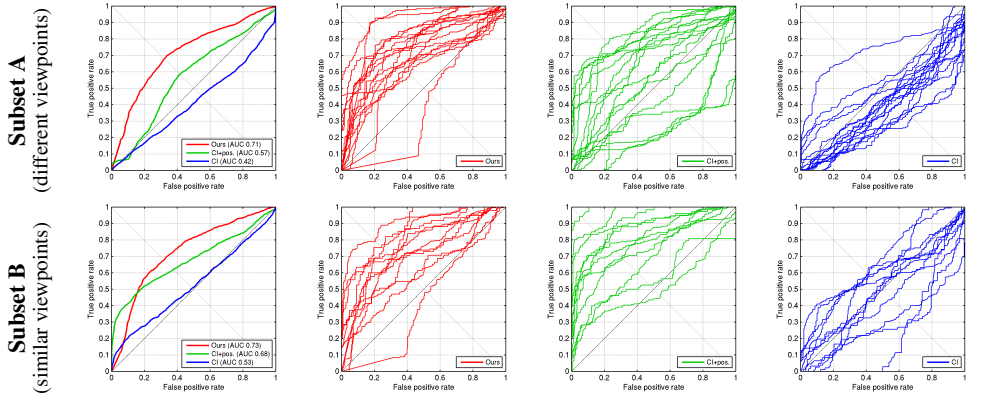
6 Experiments

Our anomaly detection approach is basically different from most existing approaches. Accordingly, we found existing benchmarks to be inappropriate for demonstrating our method. Popular datasets such as the UMN dataset [1], the UCSD dataset [14] or the Subway dataset [2] present surveillance scenarios with fixed scene and camera, where absolute position is relevant and fixed spatiotemporal grid representations are sufficient. Apart from dealing with 2D data only, particularly the task of detecting anomalies in context of a specific motion pattern of interest is not present. To demonstrate our method, we recorded a new motion anomaly dataset from persons juggling balls using a Kinect camera (see Fig. 5).



(a) (b) (c) (d) (e) (f) (g) (h)

Figure 5: Juggling pattern test datasets of different persons performing the standard 3-ball cascade pattern including various anomalies. (a-d) One person from four different viewpoints. (e-g) Three further persons performing the same juggling pattern. (h) Background motion. Supertrajectories are shown over a range of 45 frames ($\approx$ 1.5sec), the color corresponds to the depth obtained from Kinect camera.



(a) (b) Our method (3D) (c) *CI+pos.* (2D) (d) *CI* (2D)

Figure 6: Anomaly detection ROC curves. Comparison of our method and the chaotic invariants (CI) [21] for the two subsets A (top row) and B (bottom row). (a) Average performance of all methods. (b-d) Performance on single sequences for each method.

Motion trajectories are generated by large displacement optical flow tracking [18]. The tracking algorithm was adapted to include depth in addition to RGB data. Background motion is removed by a threshold on the maximum velocity. For training, three sequences (200 frames each) were selected, that contain juggling patterns from three different persons, recorded in frontal view with 1.5m distance to the camera. We chose a rather small training set with a single viewpoint, to demonstrate generalization capabilities of our method. For testing, we used 29 sequences[1]. See the supplementary Video 1 to get a better impression of the recorded training and test sequences.

We generated a framewise anomaly ground truth. Additionally, we provide a segmentation of juggling relevant motion patterns (hands and arms of persons, and juggling balls) for both, the training and test set. The segmentation is given as a pointwise labelling of supertrajectories from our representation.

6.1 Anomaly Detection in Juggling Patterns

We evaluate anomaly detection, i.e. classifying each video frame as normal or abnormal. Detection ROC curves are generated by thresholding the anomaly profile (Fig. 4) at different levels. We compare against an existing anomaly detection method, namely chaotic invariants (CI) for anomaly detection in crowded scenes [21] [2] Among the methods with code available, CI is the most related one as it provides invariance to position and magnitude. To the best of our knowledge, there is no previous method that deals with 3D+time data. However, to provide a comparison for CI on 3D+time data as well, we extended the approach to 3D.

The evaluation is based on a temporally short prototype pattern (25 frames), see Fig. 1(a). The prototype model is learned from the training sequences, and anomaly detection is performed on all 29 test sequences. For comparison, we evaluate two versions of [21]: 1) chaotic invariants only (position invariant), we denote by *CI* and 2) absolute position added to *CI*, we denote by *CI+pos*. We perform anomaly detection as described in [21]. For comparabil-

[1] An overview of the test set is given in the supplementary material in Table1.

[2] We thank the authors for providing essential code pieces to assemble an implementation of [21].

Method	Subset A	Subset B	All sequences
CI (2D) [21]	0.42	0.53	0.46
CI+pos. (2D) [21]	0.57	0.68	0.62
CI (3D)	0.42	0.41	0.42
CI+pos. (3D)	0.51	0.55	0.53
Our method (3D)	**0.71**	**0.73**	**0.72**

Table 1: Anomaly detection results of our method and the chaotic invariants (CI). The average performance is given by the area under curve.

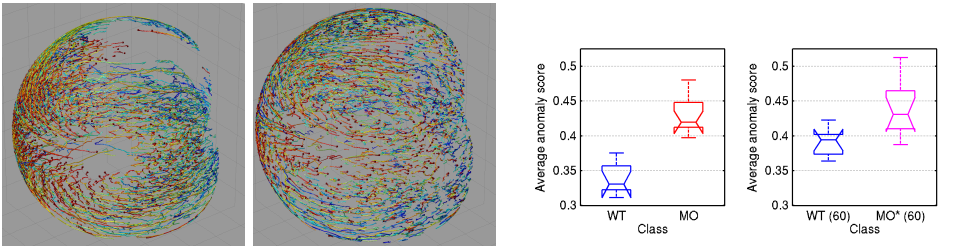
ity, we provide the same optical flow to both methods (LDOF [7]). Since the approach [21] does not include model detection, it performs anomaly detection on the whole frame. We use the segmentation of juggling relevant motion patterns to provide [21] with segmentation information. To obtain a framewise anomaly score we use the maximum measure, whereas chaotic invariants [21] use the mean measure. We split the test set into two subsets: Subset A contains 17 sequences with different viewpoints, i.e. the patterns are differently aligned compared to the training data. In contrast, subset B contains 12 sequences with similar viewpoints and well aligned patterns in 2D. ROC curves in Fig. 6 show the comparison against the original 2D version of CI. The average performance for each subset and the performance on single sequences is shown. Table1 lists the average performance for both subsets and all sequences, and includes results for the 3D version of CI.

The results in Fig. 6 and Table1 show that *CI* has problems with the strong variation between the training and test data caused by different jugglers. *CI+pos.*, which uses the absolute position, achieves a competitive performance to our approach for subset B, where the absolute position is a valuable feature. However, on subset A, where the test data includes different viewpoints, the absolute position is a rather weak feature and our method, which is invariant to changing viewpoint, clearly outperforms both *CI* and *CI+pos.* Also the use of 3D trajectories in CI does not improve results. Our method outperforms all tested variants of CI to its accurate modeling of the normal pattern variation.

We show additional results in supplementary Video 1. Anomaly detection results are rendered into the test sequences and give a good impression of the anomaly localization capability, complementary to the results presented in this section.

6.2 Anomaly Detection in Biological Motion Patterns

Furthermore, we evaluate the applicability to biological motion patterns. We use 3D+time trajectory data showing global endodermal cell dynamics in the early development of zebrafish embryos [17]. The data consists of two groups, 12 wild type (WT) embryos (normal patterns) and 12 *cxcr4a* morphant (MO) embryos (genetically modified). Fig. 7(a,b) shows an example from each group as 3D rendering (color indicates time, from blue to red). All data are resampled temporally to 100 time points. In our experiments, we select one WT pattern as the prototype and learn the prototype model using three further WT patterns. For testing we use the remaining 8 WT and 12 MO patterns. By computing overall anomaly scores for each pattern we are able to quantify a significant difference between the WT and MO motion patterns, as shown in the boxplot in Fig. 7(c). In addition, we tested our hypothesis that a time scaling of MO patterns (denoted MO*) might partially compensate this difference. To this end, we rescale MO patterns to 60 time points (increased speed by factor 1.66) and find



(a) Wild type (WT) (b) *cxcr4a* morphant (MO) (c) WT vs. MO (d) WT vs. MO*

Figure 7: Anomaly detection in biological motion patterns of cell dynamics in the early development of zebrafish embryos.

that the patterns overall can not be distinguished any more (at the 5% significance level), see the results in Fig. 7(d). The proposed approach allows to quantify significant differences between wild type and morphant motion patterns, and additionally reveals that a time scaling can partially explain and compensate for these differences.

7 Conclusion

We have presented a new approach to motion anomaly detection in complex motion patterns. Prototype patterns are allowed to appear anywhere in a video and in any orientation. Our approach therefore starts with a robust detection that is invariant to rigid transformations, followed by a spatiotemporal elastic registration of the prototype pattern to the test pattern. The precise alignment of the patterns has allowed us to detect and localize subtle anomalies, as demonstrated by experiments on a 3D motion anomaly dataset. An important application area for our approach is in biomedical image analysis, where complex developmental and growth patterns need to be compared.

Acknowledgements

We thank Nico Scherf (Institute for Medical Informatics and Biometry, TU Dresden) and Jan Huisken (Max Planck Institute of Molecular Cell Biology and Genetics (MPI-CBG), Dresden) for kindly providing the trajectory data from [17]. Also, we thank J. Koch, A. Krämer, T. Paxian and D. Mai who contributed their juggling expertise and agreed to perform diverse juggling patterns in front of our Kinect camera. This study was supported by the Excellence Initiative of the German Federal and State Governments (EXC 294).

References

[1] Unusual crowd activity dataset made available by the university of minnesota at:. URL http://mha.cs.umn.edu/Movies/Crowd-Activity-All.avi.

[2] A. Adam, E. Rivlin, I. Shimshoni, and D. Reinitz. Robust real-time unusual event detection using multiple fixed-location monitors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(3):555–560, March 2008. ISSN 0162-8828.

[3] B. Antic and B. Ommer. Video parsing for abnormality detection. In Dimitris N. Metaxas, Long Quan, Alberto Sanfeliu, and Luc J. Van Gool, editors, *IEEE International Conference on Computer Vision (ICCV)*, pages 2415–2422. IEEE, 2011. ISBN 978-1-4577-1101-5.

[4] P.J. Besl and Neil D. McKay. A method for registration of 3-d shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2):239–256, Feb 1992. ISSN 0162-8828. doi: 10.1109/34.121791.

[5] O. Boiman and M. Irani. Similarity by composition. In B. Schölkopf, J.C. Platt, and T. Hoffman, editors, *Advances in Neural Information Processing Systems 19 (NIPS)*, pages 177–184. MIT Press, 2007. ISBN 0-262-19568-2.

[6] O. Boiman and M. Irani. Detecting irregularities in images and in video. *International Journal of Computer Vision*, 74(1):17–31, 2007.

[7] T. Brox and J. Malik. Large displacement optical flow: descriptor matching in variational motion estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(3):500–513, 2011.

[8] R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu. A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing*, 16(5):1190–1208, September 1995. ISSN 1064-8275. doi: 10.1137/0916069.

[9] Y. Cong, J. Yuan, and J. Liu. Sparse reconstruction cost for abnormal event detection. In *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3449–3456. IEEE, 2011.

[10] Y. Cong, J. Yuan, and Y. Tang. Video anomaly search in crowded scenes via spatio-temporal motion context. *IEEE Transactions on Information Forensics and Security*, 8 (10):1590–1599, 2013.

[11] B.S. Everitt, S. Landau, M. Leese, and D. Stahl. *Cluster Analysis*. Wiley series in probability and statistics. Wiley, 2011. ISBN 9780470978443.

[12] J. Kim and K. Grauman. Observe locally, infer globally: A space-time mrf for detecting abnormal activities with incremental updates. In *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2921–2928, June 2009. doi: 10.1109/CVPR.2009.5206569.

[13] L. Kratz and K. Nishino. Anomaly detection in extremely crowded scenes using spatio-temporal motion pattern models. In *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1446–1453, June 2009. doi: 10.1109/CVPR.2009.5206771.

[14] V. Mahadevan, Weixin Li, V. Bhalodia, and N. Vasconcelos. Anomaly detection in crowded scenes. In *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1975–1981, June 2010. doi: 10.1109/CVPR.2010.5539872.

[15] R. Mehran, A. Oyama, and M. Shah. Abnormal crowd behavior detection using social force model. In *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 935–942, June 2009. doi: 10.1109/CVPR.2009.5206641.

[16] V. Saligrama and Zhu Chen. Video anomaly detection based on local statistical aggregates. In *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2112–2119, June 2012. doi: 10.1109/CVPR.2012.6247917.

[17] B. Schmid, G. Shah, N. Scherf, M. Weber, K. Thierbach, C. P. Campos, I. Roeder, P. Aanstad, and J. Huisken. High-speed panoramic light-sheet microscopy reveals global endodermal cell dynamics. *Nat Commun*, 4, July 2013.

[18] N. Sundaram, T. Brox, and K. Keutzer. Dense point trajectories by gpu-accelerated large displacement optical flow. In *European Conference on Computer Vision (ECCV)*, Lecture Notes in Computer Science, pages 438–451. Springer, Sept. 2010.

[19] S. Umeyama. Least-squares estimation of transformation parameters between two point patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(4):376–380, 1991. ISSN 0162-8828. doi: 10.1109/34.88573.

[20] S. Winkelbach, S. Molkenstruck, and F. M. Wahl. Low-cost laser range scanner and fast surface registration approach. In *Pattern Recognition (Proc. DAGM)*, pages 718–728. Springer, LNCS, 2006.

[21] S. Wu, B. E. Moore, and M. Shah. Chaotic invariants of lagrangian particle trajectories for anomaly detection in crowded scenes. In *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2054–2060, 2010.