Spatiotemporal Parsing of Motor Kinematics for Assessing Stroke Recovery

Borislav Antic[1,⋆], Uta Büchler[1,⋆], Anna-Sophia Wahl[2],
Martin E. Schwab[2], and Björn Ommer[1]

[1] HCI/IWR, Heidelberg University, Germany
[2] Department of HST, ETH Zurich, Switzerland

Abstract. Stroke is a major cause of adult disability and rehabilitative training is the prevailing approach to enhance motor recovery. However, the way rehabilitation helps to restore lost motor functions by continuous reshaping of kinematics is still an open research question. We follow the established setup of a rat model before/after stroke in the motor cortex to analyze the subtle changes in hand motor function solely based on video. Since nuances of paw articulation are crucial, mere tracking and trajectory analysis is insufficient. Thus, we propose an automatic spatiotemporal parsing of grasping kinematics based on a max-projection of randomized exemplar classifiers. A large ensemble of these discriminative predictors of hand posture is automatically learned and yields a measure of grasping similarity. This non-parametric distributed representation effectively captures the nuances of hand posture and its deformation over time. A max-margin projection then not only quantifies functional deficiencies, but also back-projects them accurately to specific defects in the grasping sequence to provide neuroscience with a better understanding of the precise effects of rehabilitation. Moreover, evaluation shows that our fully automatic approach is reliable and more efficient than the prevalent manual analysis of the day.

1 Introduction

Generation and control of movements are fundamental requirements to interact and respond to the environment. Complex hand functions such as grasping depend on well-orchestrated and precisely coordinated sequences of motor actions. When a stroke occurs, they are often impaired and the subject suffers from lost skilled motor functions. Studying motor impairment and establishing new therapeutic strategies, is a key challenge of neuroscience typically conducted in rodents to provide, as demonstrated in [1,2], information about the human model too. Often a tedious visual inspection of the grasping function is the only means to determine outcome levels after stroke. To tackle this challenge, we propose a fully automatic approach for analyzing the kinematics of grasping. These characteristic patterns of hand deformation are significantly more intricate than mere trajectories shown in Fig. 1. Our goal is not only to identify *if* motor function is impaired, but also *how* exactly the limb paresis affects grasping.

[⋆] Indicates equal contribution.

A. Frangi et al. (Eds.): MICCAI 2015, Part III, LNCS 9351, pp. 467–475, 2015.

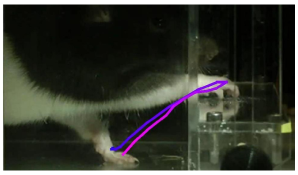
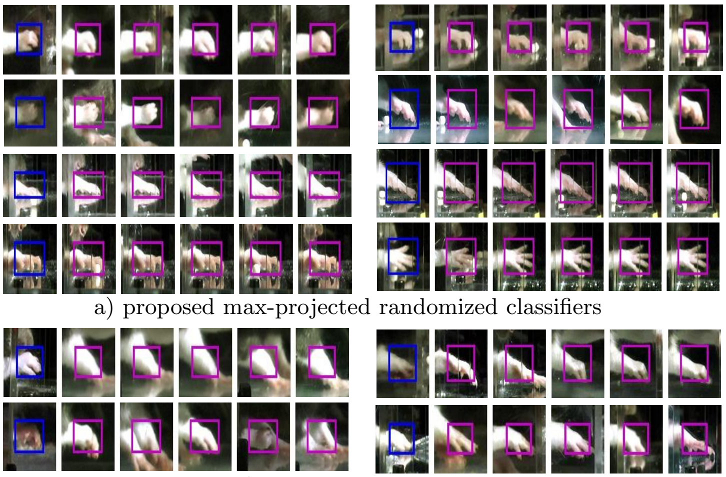


Fig. 1. Side view (cutout) of a rat performing single pellet grasping in a Plexiglas box with opening at the side. The paw trajectory of a successful grasp is superimposed (time running from blue to red).

To accurately represent and detect paws we extend the powerful exemplar-based paradigm by max-projection of randomized versions of an exemplar classifier. A large set of these paw classifiers is then aggregated to define similarities between hand postures and group them. Grasping is characterized by the deformation of the paw over time. Since we also need to represent the vastly different abnormal grasps of impaired animals, neither priors on the smoothness of kinematics, nor explicit models for a grasp are appropriate. Therefore, we propose a non-parametric representation that is based on robust matching of grasp sequences. Using this compact model, max-margin multiple-instance learning yields a classifier that not only recognizes impaired grasping but also identifies fragments of a grasp that are characteristically altered due to the injury.

This study analyzes the recovery of motor function in Long-Evans rats performing single pellet grasping [3] before and after a photothrombotic stroke destroys the corresponding sensorimotor cortex of the grasping paw. We compare recovery under *i)* a rehabilitative therapy (*Anti-Nogo* neuronal growth promoting immunotherapy followed by rehabilitation [4], denoted green group), *ii)* for a cohort without treatment (red), *iii)* and a control of sham operated rats (black).

Due to its importance for understanding and analyzing motor function, there has been considerable effort on analyzing grasping behavior. A main focus has been on studying hand trajectories [5,4], thus ignoring the crucial, intricate kinematics of hand motion [1], which involve the temporal deformation of fingers and the overall shape of the hand. Representing, detecting, and distinguishing the fine articulation of small, hairy, fast moving fingers of a rat paw under self-occlusion, noise, and variations between animals pose tremendous challenges for evaluation. Consequently, previous studies mainly rely on tedious manual identification of hand posture in individual video frames [5,1,2]. In large evaluations, motion analysis is often simplified by applying reflective motion capture markers [6] or even electromagnetic tracking gloves [7]. However, these techniques are typically used on human or primate subjects but are not applicable to rats, due to the small size of their paws and distraction that attached devises impose. In [8] human hand is tracked by relying on depth from structured light, which is too coarse and slow for the fast moving, small rat paws. Moreover, [8] discard appearance and solely depend on weak contour information and articulation models of normal hands, which are not appropriate for abnormal locomotion after stroke.


a) proposed max-projected randomized classifiers

b) exemplar classifiers

Fig. 2. a) Subset of our pooled classifiers proposed in Sect. 2.2. For each classifier (blue) we show the five best detections on query videos (magenta). Note the subtle difference that the model captures, such as pronation (1st row left), supination (2nd row left), or the different stages of hand closure (1st, 2nd row right). b) Baseline: Subset of classifiers trained by exemplar Support Vector Machine (SVM) [9] without our pooling. The matches are significantly less specific than the proposed approach.

2 Approach

2.1 Creating Candidate Foreground Regions

To focus our analysis on hand motor function we first extract a large set of candidate foreground regions. Following [10] we decompose frames of video into a low-rank background model and a sparse vector corresponding to foreground pixels. Candidate regions $x_i \in \mathcal{X}$ are then randomly sampled from the estimated foreground of a set of sample videos to initialize the subsequent learning of classifiers and compute their HOG features (size 10×10). K-Nearest Neighbors density estimation then reveals rare outliers, which are then removed from $\mathcal{X}$.

2.2 Robust Exemplar Classification

We seek a representation of rat paws that i) facilitates detection and tracking of hands in novel videos and ii) exposes the subtle differences in hand posture, while being robust to noise, illumination differences, and intra-class variability.

Max-projected Randomized Exemplar Classifiers: To measure the similarity to a sampled region x_i, we train a discriminative exemplar classifier w_i, using x_i as positive example. Since $\mathcal{X}$ likely contains other samples similar to x_i, we cannot use all remaining samples in $\mathcal{X}$ as negatives for training or perform hard negative mining [9]—too much overlap with the single positive makes

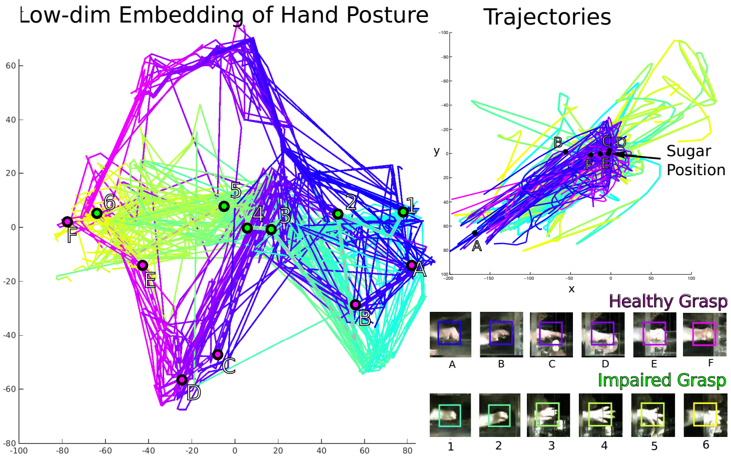


Fig. 3. *Left:* Paws are represented using the proposed non-parametric representation of hand posture. A distance preserving embedding then maps all detected hands onto 2D and straight lines connect successive frames within a single grasp. Time is encoded as blue to red for healthy animals and cyan to yellow for impaired 2 days post stroke. For one healthy and one impaired grasping sequence selected frames (marked as 1,2,.. and A,B,.. along the sequence) are shown *bottom right*. On *top right* the grasping trajectories of all grasps are depicted relative to the position of the sugar pellet at the origin. The non-parametric representation of postures helps to accurately distinguish impaired grasping patterns from healthy ones and emphasizes characteristic abnormal postures within individual grasps, such as D vs. 4.

learning unreliable. Thus we train in Eq. 1 an ensemble of K exemplar classifiers w_i^k with their offsets b_i^k, $k \in \{1, \ldots, K\}$, using randomly selected negative sets $\mathcal{X}_i^k \subset \mathcal{X} \backslash x_i$. The soft margin parameter $C = .01$ was chosen via cross-validation.

$$\min_{w_i^k, b_i^k} \|w_i^k\|^2 + C \max\big(0, 1 - \langle w_i^k, x_i\rangle - b_i^k\big) + \frac{C}{|X_i^k|} \sum_{j=1}^{|X_i^k|} \max\big(0, 1 + \langle w_i^k, x_j\rangle + b_i^k\big). \tag{1}$$

To compensate for the unreliability of individual classifiers, we aggregate them using a max-projection that selects for each feature dimension the most confident classifier $w_i(\bullet) := \max_k w_i^k(\bullet)$. The scores of these max-projected randomized exemplar classifiers w_i are then calibrated by a logistic regression [9].

Dictionary of Paw Classifiers: Now we reduce the large candidate set $\mathcal{X}$ (size 1000) and create a dictionary $\mathcal{D} \subseteq \mathcal{X}$ to canonically represent hand postures of previously unseen rats. The randomized exemplar classifiers provide a robust measure of pair-wise similarity $s(x_i, x_j) := \frac{1}{2}\big(\langle w_i, x_j\rangle + \langle w_j, x_i\rangle\big)$. Redundant candidates are merged by Normalized Cuts to obtain a dictionary $\mathcal{D}$ (size 100) that is sufficiently diverse and rich non-parametric representation of all viable hand-postures.[1] See Fig. 2 for examples and a comparison to standard exemplars.

[1] A coarser discretization is possible by discarding elements with small contribution.

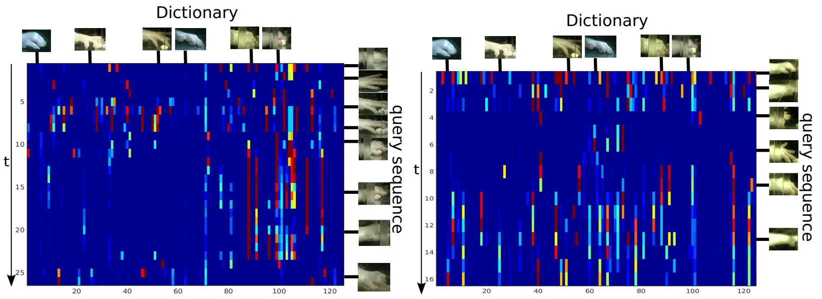


Fig. 4. Representation of hand posture for all frames of a successful (left) and a failed grasping sequence (right). On each query frame (rows) all classifiers (columns) from the dictionary of characteristic hand configurations of Sect. 2.2 are evaluated. In this representation successful grasps typically first exhibit activations of pronation, followed by a characteristic opening and closing (t=4..10, left) and retraction of the closed hand.

In a new video all classifiers w_i from dictionary $\mathcal{D}$ are applied densely. Averaging the scores of the top scoring k classifiers and taking the location with the highest score yields the final paw detection and its trajectories over time, Fig. 3.

Non-parametric Representation of Hand Posture: On a novel paw sample x the joint activation pattern of all $\{w_i\}_{i\in\mathcal{D}}$ yields an embedding $e := [\langle w_1, x\rangle, \ldots, \langle w_{|\mathcal{D}|}, x\rangle]$. Moreover, novel hand configurations can now be related by comparing their embedding vectors, since similar samples give rise to similar classifier activations. To visualize a hand posture, the high-dimensional embedding is mapped to a low-dimensional projection using the t-SNE method [11].

2.3 Spatiotemporal Parsing

While hand posture is a crucial factor, its deformation during a grasp is even more decisive. We represent an M frame long grasp j as a sequence in the embedding space $S_j := [e_1^j, \ldots, e_M^j]$. Measuring the similarity of grasps based on their embeddings requires a sequence matching since they are not temporally aligned. We thus seek a mapping $\pi : \{1, \ldots, M\} \mapsto \{0, \ldots, M'\}$ that aligns S_j of length M with $S_{j'}$ of length M' (0 denotes outliers), cf. Fig. 5. Their distance is then defined in the embedding space as $d(S_j, S_{j'}) := \sum_{i=1}^{M} \|e_i^j - e_{\pi(i)}^{j'}\|$. A matching $\pi(\bullet)$ should penalize outliers and variations in the temporal order,

$$\min_{\pi} \sum_{i=1}^{M} \|e_i^j - e_{\pi(i)}^{j'}\| + \lambda \sum_{i=1}^{M-1} \mathbf{1}\big(\pi(i) > \pi(i+1)\big), \text{ s.t. } |\pi(i) - i| \leq B, \forall i, \qquad (2)$$

where $\mathbf{1}(\cdot)$ denotes the identity function. The constraint in Eq. 2 prevents matched frames to be more than B frames apart from each other. The second sum in

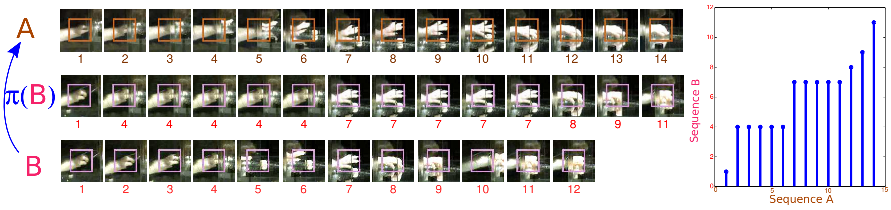
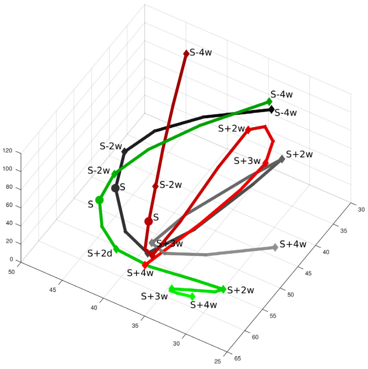


Fig. 5. Sequence matching: Sequence A is matched to B by permuting and warping this sequence into $\pi(B)$ using the approach of Sect. 2.3. Frames are numbered and the permutations $\pi(\bullet)$ are shown on the right.

Fig. 6. Novel grasps are represented by automatically explaining them with a compact set of prototypical grasping sequences (Sect. 2.3), yielding an embedding. For all animals of the black (sham control), green (therapy), and red (no treatment) cohort the mean representation of all graspings of a trial session is visualized by distance preserving projection to a 3D subspace. The mean representations of successive trial sessions are then connected to show recovery within each cohort (time runs from dark to light, from 4 weeks before stroke to 4 weeks after). Note that black and green cohort show similar final recovery, whereas red ends up close to its state 2 days post stroke, indicating only little recovery (the red loop at right is actually elevated and far away from the others).



Eq. 2 allows for more flexible matching than the standard string matching or dynamic time warping. Substitution $z_{i,i_1',i_2'} := \mathbf{1}\big(\pi(i) = i_1' \wedge \pi(i+1) = i_2'\big)$ in Eq. 2 transforms the sequence matching problem to an integer linear program (ILP),

$$\max_{z_\bullet \in \{0,1\}} \sum_{i=1}^{M-1} \sum_{i_1',i_2'=0}^{M'} z_{i,i_1',i_2'} \psi_{i,i_1',i_2'}, \text{ s.t. } \sum_{i_1',i_2'=0}^{M'} z_{i,i_1',i_2'} = 1 \wedge \sum_{i_1'=0}^{M'} z_{i,i_1',i_2'} = \sum_{i_3'=0}^{M'} z_{i+1,i_2',i_3'}, \quad (3)$$

where $\psi_{i,i_1',i_2'}$ is the sum of all terms in Eq. 2 that correspond to $z_{i,i_1',i_2'} = 1$. IBM ILOG CPLEX software is applied to solve Eq. 3 (tenth of a second for sequences of 30 frames).

Due to a large number of grasping actions in our dataset, it is computationally prohibitive to match a novel sequence to all others. Thus we reduce redundancy as in Sect. 2.2 and construct a dictionary $\mathcal{D}_{seq} := \{S_1, \ldots, S_Q\}$ with canonical grasping sequences that then explain novel grasps yielding a spatiotemporal parsing. Measuring the distances of a new sequence S' to all prototypical ones in $\mathcal{D}_{seq}$ yields the sequence-level embedding $E' := [d(S', S_1), ..., d(S', S_Q)]$, after aligning grasps by our sequence matching to compute distances.

2.4 Automatic Grasping Diagnostics

Multiple instance learning (MIL) [12] is utilized to train an SVM on few successful pre-stroke grasps against failed grasps from directly after stroke using the representation from Sect. 2.3. Since even before stroke there are failed attempts, the pre-stroke grasps are randomly aggregated in bags (20 attempts per bag) and the successful positive training samples are inferred using MIL. To diagnose a novel grasping sequence S', the MIL trained linear SVM is applied before transforming the classifier scores into the probability of a grasping sequence being abnormal. To automatically discover what made a grasp fail and how impaired motor function manifests, we back-project the kernel-based sequence representation onto frames that are indicative for grasping failure, i.e., the parts of the sequence that have highest responsibility for yielding a bad SVM score, Sect. 3.

3 Experimental Results

Recording Setup: For the three cohorts of 10 rats described in Sect. 1, grasping has been filmed from the side, cf. Fig. 1, with a Panasonic HDC-SD800 camcorder at 50fps, shutter 1/3000, ~8 hours of video in total. Recording sessions were -4, -2, 2, 3, 4 weeks before/after stroke, each with 100 grasp trials per animal per session, yielding in total 15000 individual grasps for evaluation of our analysis.

Grasping Diagnostics: In an initial experiment we computed paw trajectories during grasping, Fig. 3 (top right). However, since the unilateral stroke has most impact on the fine motor skills of the hand rather than the full arm, its implications for grasping are only truly revealed by the detailed analysis of hand posture, Fig. 3 (left). From a neurophysiological view, this unsupervised analysis reveals characteristics of impaired grasping such as incorrect supination (D vs. 4), evident by the large distances between healthy and impaired kinematics especially around the moment of pellet capture (C,D). The difference in hand posture around this moment is also evident around t=4..10 in Fig. 4.

Fig. 6 visualizes the recovery of all animals within one cohort by averaging over all their grasp trials conducted at the same time pre/post stroke. This unsupervised approach highlights the positive effect of the rehabilitation during the recovery process. Based on the same sequence representation, the classifier of Sect. 2.4 predicts if grasps are healthy or impaired in Fig. 8. This analysis not only underlines the positive outcome of the therapy, but also compare our prediction of fitness of a grasp against a manual labeling provided by 5 experts. They accurately annotated sequences as 1, .5, 0 if they were successful, partially successful (correct up to final supination), or failed. Statistical hypothesis testing, Tab. 1, shows that our automatic prediction compares favorably with manual annotations (p-value between .002 and .02, two-tailed t-test capturing deviations

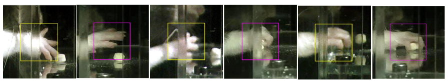


Fig. 7. Examples of hand postures that deteriorated grasping fitness (yellow) vs. those that improved it (magenta).

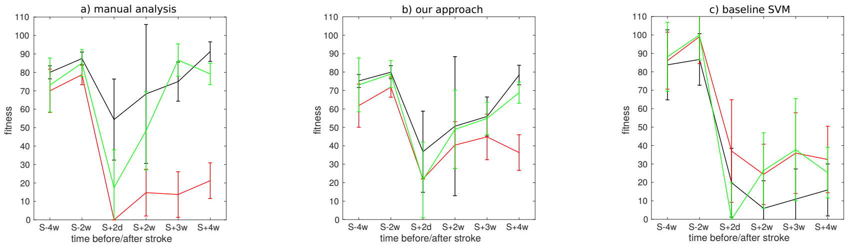


Fig. 8. Prediction of grasping fitness using a) manual analysis by neuroscientists, b) our approach, c) baseline SVM approach

Table 1. Comparing predictions of our approach Fig. 8b) and of the baseline approach (no randomized pooling and sequence matching) 8c) against the manual annotation of neuroscientists 8a). Overall our approach significantly improves upon the baseline.

	Our Approach			Baseline SVM		
Cohort	p-value	RMSE	R^2	p-value	RMSE	R^2
Sham control (black)	0.002	4.06	0.927	0.401	13.6	0.18
No treatment (red)	0.004	11.8	0.896	0.003	11.1	0.909
Therapy (green)	0.019	14.2	0.779	0.161	23	0.425

in both directions from the ground truth) and are significantly more robust than a baseline approach without the proposed randomized pooling and sequence matching. Finally we also retrieve frames of an impaired sequence that mostly deteriorated the grasp by causing bad alignments in sequence matching, Fig. 7.

4 Conclusion

To analyze grasping function and its restoration after stroke we have presented a fully automatic video-based approach. Details of hand posture are captured by a non-parametric representation based on a large set of max-projected randomized exemplar classifiers. The spatiotemporal kinematics of grasping is then explained by a robust sequence matching. With this representation an unsupervised and a supervised approach have been presented to automatically analyze the recovery after stroke and predict the impairment of individual grasps. This approach has the potential to open up new possibilities of studying motor restoration over time and evaluating the efficiency of therapeutic interventions after stroke.[2]

[2] This research has been funded in part by the Ministry for Science, Baden-Wuerttemberg and the Heidelberg Academy of Sciences, Heidelberg, Germany.

References

1. Alaverdashvili, M., Whishaw, I.Q.: A behavioral method for identifying recovery and compensation: hand use in a preclinical stroke model using the single pellet reaching task. Neurosci. Biobehav. Rev. 37, 950–967 (2013)
2. Sacrey, L.A., Alaverdashvili, M., Whishaw, I.Q.: Similar hand shaping in reaching-for-food (skilled reaching) in rats and humans provides evidence of homology in release, collection, and manipulation movements. Behav. Brain Res. (2009)
3. Metz, G.A., Whishaw, I.Q.: Skilled reaching an action pattern: stability in rat (rattus norvegicus) grasping movements as a function of changing food pellet size. Behav. Brain Res. 116(2), 111–122 (2000)
4. Wahl, A.S., Omlor, W., Rubio, J.C., Chen, J.L., Zheng, H., Schröter, A., Gullo, M., Weinmann, O., Kobayashi, K., Helmchen, F., Ommer, B., Schwab, M.E.: Asynchronous therapy restores motor control by rewiring of the rat corticospinal tract after stroke. Science 344, 1250–1255 (2014)
5. Whishaw, I., Pellis, S.: The structure of skilled forelimb reaching in rat: a proximally driven movement with single distal rotatory component. Behav. Brain Res. (1990)
6. Goebl, W., Palmer, C.: Temporal control and hand movement efficiency in skilled music performance. PLoS One 8 (2013)
7. Schaffelhofer, S., Agudelo-Toro, A., Scherberger, H.: Decoding wide range of hand configurations from macaque motor, premotor, and parietal cortices. J. N. Sci. (2015)
8. Hamer, H., Schindler, K., Koller-Meier, E., Gool, L.V.: Tracking a hand manipulating an object. In: ICCV, pp. 1475–1482 (2009)
9. Eigenstetter, A., Takami, M., Ommer, B.: Randomized max-margin compositions for visual recognition. In: CVPR, pp. 3590–3597 (2014)
10. Wright, J., Ganesh, A., Rao, S., Peng, Y., Ma, Y.: Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optim. In: NIPS (2009)
11. Van der Maaten, L., Hinton, G.: Visualizing data using t-sne. JMLR 9 (2008)
12. Andrews, S., Tsochantaridis, I., Hofmann, T.: Support vector machines for multiple-instance learning. In: NIPS, pp. 577–584 (2003)