Learning Motion Primitives using Spatio-Temporal NMF

Sven Hellbach, Christian Vollmer, Julian Eggert, Horst-Michael Groß

2011

Learning Motion Primitives using Spatio-Temporal NMF

Sven Hellbach[1], Christian Vollmer[1], Julian P. Eggert[2], and Horst-Michael Gross[1]

[1] Ilmenau University of Technology, Neuroinformatics and Cognitive Robotics Labs,
POB 10 05 65, 98684 Ilmenau, Germany
christian.vollmer@tu-ilmenau.de
[2] Honda Research Institute Europe GmbH, Carl-Legien-Strasse 30,
63073 Offenbach/Main, Germany
julian.eggert@honda-ri.de

1 Introduction

The understanding and interpretation of movement trajectories is a crucial component in dynamic visual scenes with multiple moving items. Nevertheless, this problem has been approached very sparsely by the research community. Most approaches for describing motion patterns, like [1], rely on a kinematic model for the observed human motion. This causes the drawback that the approaches are difficult to adapt to other objects. Here, we aim at a generic, model-independent framework for decomposition, classification and prediction.

Consider the simple task for a robot of grasping an object which is handed over by the human interaction partner. To avoid a purely reactive behaviour, which might lead to 'mechanical' movements of the robots, it is necessary to predict the further movement of the human's hand.

In [2] an interesting concept for a decomposition task is presented. Like playing a piano a basis alphabet – the different notes – are superimposed to reconstruct the observation (the piece of music). Regarding only the information, when a base primitive was active, gives rise to an instance of the so called 'piano model' which is a very low-dimensional and sparse representation and which can be exploited for further processing. While the so-called piano model relies on a set of given basis primitives, our approach is able to learn these primitives from the training data.

We use [3], a blind source separation approach in concept similar to PCA and ICA. The system of basis vectors which is generated by the NMF is not orthogonal. This is very useful for motion trajectories, since one basis primitive is allowed to share a common part of its trajectory with other primitives and to specialize later.

2 Non-negative Matrix Factorization

Like other approaches, e. g. PCA and ICA, non-negative matrix factorization (NMF) [3] is meant to solve the source separation problem. Hence, a set of training data is decomposed into basis primitives $\mathbf{W}$ and activations thereof $\mathbf{H}$:

$$\mathbf{V} \approx \mathbf{W} \cdot \mathbf{H} \tag{1}$$

Each training data sample is represented as a column vector $\mathbf{V}_i$ within the matrix $\mathbf{V}$. Each column of the matrix $\mathbf{W}$ stands for one of the basis primitives. In matrix $\mathbf{H}$ the element H_i^j determines how the basis primitive $\mathbf{W}_j$ is activated to reconstruct training sample $\mathbf{V}_i$.

For generating the decomposition, optimization-based methods are used. Hence, an energy function E has to be defined:

$$E(\mathbf{W}, \mathbf{H}) = \frac{1}{2} \|\mathbf{V} - \mathbf{T} \cdot \mathbf{W} \cdot \mathbf{H}\|^2 + \lambda \sum_{i,j} H_i^j \tag{2}$$

By minimizing the energy equation, it is now possible to achieve a reconstruction using the matrices $\mathbf{W}$ and $\mathbf{H}$. This reconstruction is aimed to be as close as possible to the training data $\mathbf{V}$. In

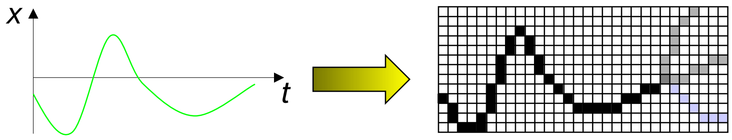

Fig. 1. Motion Trajectories are transferred into a grid representation. A grid cell is set to 1 if it is in the path of the trajectory and set to zero otherwise. Each dimension has to be regarded separately. During the prediction phase multiple hypotheses can be gained by superimposing several basis primitives. This is indicated with the grey trajectories on the right side of the grid.

addition the basis primitives are intended to be allowed to move, rotate and scale freely. This is achieved by adding a transformation matrix **T** to the decomposition formulation [4]. For each allowed transformation the corresponding activity has to be trained individually. To avoid trivial or redundant solutions a further sparsity constraint is necessary. Its influence can be controlled using the parameter λ [5].

The minimization of the energy function can be done by gradient descent. The factors H and W are updated alternately with a variant of exponentiated gradient descent until convergence.

3 Decomposing Motion Trajectories

For being able to decompose and to predict the trajectories of the surrounding dynamic objects, it is necessary to identify them and to follow their movements. For simplification, a tracker is assumed, which is able to provide such trajectories in real-time. A possible tracker to be used is presented in [6]. The given trajectory of the motion is now interpreted as a time series $\mathcal{T}$ with values $\boldsymbol{s}_i = (x_i, y_i, z_i)$ for time steps $i = 0, 1, \ldots, n-1$:

$$\mathcal{T} = (\mathbf{s}_0, \mathbf{s}_1, \ldots, \mathbf{s}_{n-1}). \tag{3}$$

It is now possible to present the vector $\mathcal{T}$ directly to the NMF approach. But this could result in an unwanted behaviour, while trying to reconstruct the motion by use of the basis primitives. Imagine two basis primitives, one representing a left turn and another representing a right turn. A superposition of those basis primitives would result in a straight movement.

The goal is to have a set of basis primitives, which can be concatenated one after the other. Furthermore, it is necessary for a prediction task to be able to formulate multiple hypotheses. For achieving these goals, the x-t-trajectory is transferred into a grid representation, as it is shown in figure 1. Then, each grid cell (x_i, t_j) represents a certain state (spatial coordinate) x_i at a certain time t_j. Since most of the state-of-the-art navigation techniques rely on grid maps, the prediction can be integrated easily. Grid Maps were first introduced in [7]. This 2D-grid is now presented as image-like input to the NMF algorithm. Using the grid representation of the trajectory also supports the non-negative character of the basis components and their activities.

It has to be mentioned, that the transformation to the grid representation is done for each of the dimensions individually. Hence, the spatio-temporal NMF has to be processed on each of these grids. Regarding each of the dimensions separately is often used to reduce the complexity of the analysis of trajectories (compare [8]). However, the algorithm's only limitations to handle multi-dimensional grid representation is the increase of computational effort.

While applying an algorithm for basis decomposition to motion trajectories it seems to be clear that the motion primitives can undergo certain transformations to be combined to the whole trajectory. For example, the same basis primitive standing for a straight move can be concatenated with another one standing for a left turn. Hence, the turning left primitive has to be moved to the end of the straight line, and transformation invariance is needed while decomposing motion data. For our purposes, we concentrate on translation. This makes it possible to reduce the complexity of the calculations and to achieve real time performance.

The sparse coding constraint helps to avoid trivial solutions. Since the input can be compared with a binary image, one possible solution would be a basis component with only a single grid cell filled. These can then be concatenated one directly after another. So, the trajectory would simply be copied into the activities.

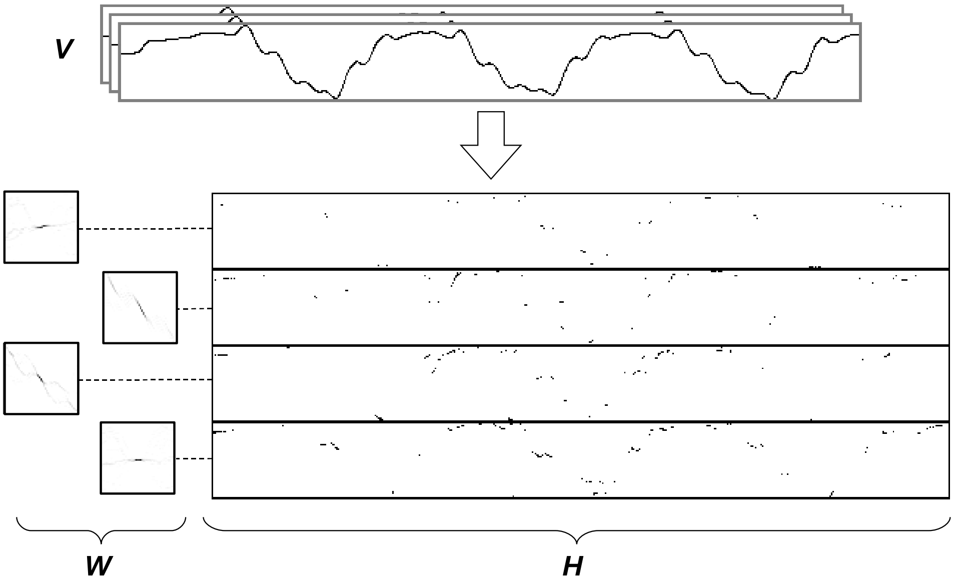

Fig. 2. Training with Spatio-Temporal NMF. Given is a set of training samples in matrix **V**. The described algorithm computes the weights **W** and the corresponding activities **H**. Only the weights are used as basis primitives for further processing.

Training Phase: The goal of the training phase is to gain a set of basis primitives which allow to decompose an observed and yet unknown trajectory (see Fig. 2). As it is discussed in section 3, the training samples are transferred into a grid representation. These grid representations are taken as input for the NMF approach and are therefore represented in matrix **V**. On this matrix **V** the standard NMF approach, extended by the sparsity constraint and by translation invariance, is applied. The algorithm is summarized in [9].

Beside the computed basis primitives, the NMF algorithm also provides the information of how each of the training samples can be decomposed by these basis primitives.

Application Phase: As it is indicated in Fig. 3, from the training phase a set of motion primitives is extracted. During the application phase, we assume that the motion of a dynamic object (e. g. a person) is tracked continuously. For getting the input for the NMF algorithm, a sliding window approach is taken. A certain frame in time is transferred into the already discussed grid like representation. For this grid the activation of the basis primitives is determined by trying to reconstruct the input.

The standard approach to NMF implies that each new observation at the next time step demands a new random initialization for the optimization problem. Since an increasing column number in the grid representation stands for an increase in time, the trajectory is shifted to the left while moving further in time. For identical initialization, the same shift is then reflected in the activities after the next convergence. To reduce the number of iterations until convergence, the shifted activities from the previous time step are used as initialization for the current one.

To fulfil the main goal discussed in this paper – the prediction of the observed trajectory into the future – the proposed algorithm had to be extended. Since the algorithm contains the transformation invariance constraint, the computed basis primitives can be translated to an arbitrary position on the grid. This means that they can also be moved in a way that they exceed the borders of the grid. Up to now, the size of reconstruction was chosen to be the same size as the input grid. Hence, using the standard approach means that the overlapping information has to be clipped. To be able to solve the prediction task, we simply extend the reconstruction grid to the right – or into the future (see Fig. 3). So, the previously clipped information is available for prediction.

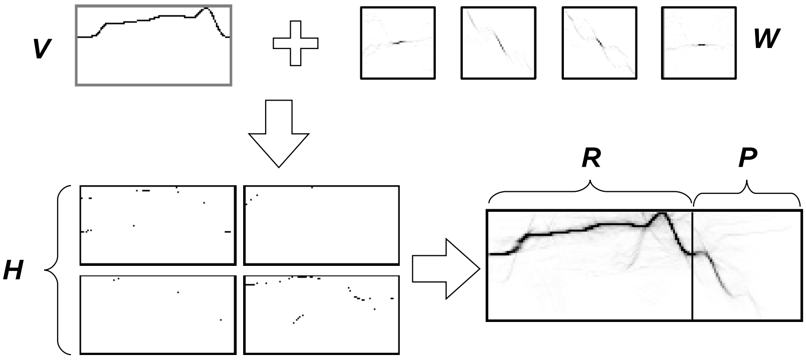

Fig. 3. The basis primitives **W**, which were computed during the training, are used to reconstruct (matrix **R**) the observed trajectory **V**. This results in a set of sparse activities – one for each basis primitive – which describe on which position in space and time a certain primitive is used. Beside the reconstruction of the observed trajectory (shown in Fig. 3), it is furthermore possible to predict a number of time steps into the future. Hence, the matrix **R** is extended by the prediction horizon **P**.

References

1. Hoffman, H., Schaal, S.: A computational model of human trajectory planning based on convergent flow fields. In: 37st Meeting of the Society of Neuroscience. (2007)
2. Cemgil, A., Kappen, B., Barber, D.: A generative model for music transcription. IEEE Transactions on Speech and Audio Processing **14** (2006) 679–694
3. Lee, D.D., Seung, H.S.: Algorithms for non-negative matrix factorization. Advances in Neural Information Processing **13** (2001) 556–562
4. Eggert, J., Wersing, H., Körner, E.: Transformation-invariant represenatation and NMF. In: IJCNN. (2004) 2535 – 2539
5. Eggert, J., Körner, E.: Sparse Coding and NMF. In: IJCNN. (2004) 2529 – 2533
6. Otero, N., Knoop, S., Nehaniv, C., Syrdal, D., Dautenhahn, K., Dillmann, R.: Distribution and Recognition of Gestures in Human-Robot Interaction. ROMAN (2006) 103–110
7. Elfes, A.: Using Occupancy Grids for Mobile Robot Perception and Navigation. Computer **12**(6) (June 1989) 46–57
8. Naftel, A., Khalid, S.: Classifying spatiotemporal object trajectories using unsupervised learning in the coefficient feature space. MM Syst. **12**(3) (2006) 227–238
9. Hellbach, S., Eggert, J., Koerner, E., Gross, H.M.: Basis decomposition of motion trajectories using spatio-temporal nmf. In: ICANN. (2009) 804–814