SCIENCE CHINA
Information Sciences

• LETTER •

January 2022, Vol. 65 119103:1–119103:2
https://doi.org/10.1007/s11432-020-2967-3

Human motion segmentation based on structure constraint matrix factorization

Hongbo GAO[1*], Fang GUO[2], Juping ZHU[1], Zhen KAN[1] & Xinyu ZHANG[3]

[1]*Department of Automation, School of Information Science and Technology, University of Science and Technology of China, Hefei* 230022*, China;*
[2]*Xi'an Insitute of Microeiectronics Technology, Xi'an* 710054*, China;*
[3]*School of Vehicle and Mobility, Tsinghua University, Beijing* 100084*, China*

Received 24 January 2020/Revised 27 April 2020/Accepted 12 June 2020/Published online 1 February 2021

Citation Gao H B, Guo F, Zhu J P, et al. Human motion segmentation based on structure constraint matrix factorization. Sci China Inf Sci, 2022, 65(1): 119103, https://doi.org/10.1007/s11432-020-2967-3

Dear editor,

The human motion recognition based on the segmented datasets is a hot multidisciplinary research topic in the field of computer vision. However, in reality, the collected data is without segmented. Therefore, the motion segmentation is crucial. Currently, common methods include principal component analysis (PCA), cluster algorithm and deep learning. However, the difficulty of human motion segmentation is summarized as follows: (1) The reasonable and easy-to-calculate similarity measurement criteria are designed for sequential motion data points. (2) In general, motion sequential data are a high-dimensional sequential data, which needs to be preprocessed by similarity measurement, data alignment as well as data dimensionality reduction.

The proposed human motion segmentation framework for human motion sequence is shown in Figure 1. First, the solution H is obtained by constructing an optimization objective and acquiring its local optimal solution by the method of alternating iterative. Second, the relation diagram G can be generated easily by utilizing the matrix of H. Lastly, the segmentation result is obtained by segmenting G by the method of normalized cut (Ncut).

Decomposition method of structural constraint matrix. In this study, we propose a human motion data representation method based on structure constraint and semi-NMF, i.e., structural constraint semi-NMF (SCseNMF).

Let input $\boldsymbol{X} \in \mathbb{R}^{d\times n}$, and the corresponding low-dimensional representation be recorded as $\boldsymbol{H} \in \mathbb{R}^{p\times n}$, where d is the original dimension of the data points in the sequence, p is the dimension of data points in the low-dimensional space, and n is the total number of data points in the sequence. Since the human motion is composed of the continuous sequence, there is a high similarity among neighboring sequential points. We hope that when constructing the sequential data representation, neighboring points are as similar as possible. Let the current data point be i, and the sum of neighbors before and after the comparison be q with q being a positive even number. We hope that the error between point i and the corresponding neighboring points is as small as possible, that is, $\sum_{l=1}^{q} \boldsymbol{H}_i - \boldsymbol{H}_l$ tends to zero as much as possible. Therefore, a special banded matrix is defined as $\boldsymbol{R} \in \mathbb{R}^{n\times n}$:

$$\boldsymbol{R}(i,j) = \begin{cases} -q, & i=j, \\ 1, & 1 \leqslant |i-j| \leqslant \frac{q}{2}, \\ 1, & j+\frac{q}{2} < i \leqslant q+1, 1 \leqslant j \leqslant \frac{q}{2}, \\ 1, & n-q \leqslant i < j-\frac{q}{2}, n-\frac{q}{2} < j \leqslant n, \\ 0, & \text{otherwise}. \end{cases} \tag{1}$$

Therefore, $\boldsymbol{HR}$ represents the similarity error between the sequential point and its q neighbors.

To better process the data with negative values, Semi-NMF is defined in this study as $\boldsymbol{X} \approx \boldsymbol{ZH}^{+}$, where $\boldsymbol{Z} \in \mathbb{R}^{d\times p}$ represents the feature space, and $\boldsymbol{H}^{+}$ represents the negative matrix of the input data in the feature space. By combining semi-NMF and structural structure regular term, the optimization problem can be constructed as

$$\min_{\boldsymbol{ZH}} = ||\boldsymbol{X} - \boldsymbol{ZH}||_F^2 + \alpha||\boldsymbol{HR}||_{2,1} \quad \text{s.t.} \ \ \boldsymbol{H} \geqslant 0, \tag{2}$$

where $||\cdot||_F^2$ denotes the Frobenius norm, $||\cdot||_{2,1}$ denotes the $L_{2,1}$ norm, and α is the weight parameter.

The basic idea of solving the above optimization problem is to obtain the local optimal solution by the alternating iterative methods. We can fix $\boldsymbol{H}$ to obtain the iterative rule of $\boldsymbol{Z}$. Thus, this problem can be transformed into solving $\boldsymbol{X} = \boldsymbol{ZH}$, that is, $\boldsymbol{Z} = \boldsymbol{XH}^{\mathrm{T}}(\boldsymbol{HH}^{\mathrm{T}})^{-1}$. Furthermore, we can fix $\boldsymbol{Z}$ to obtain the iteration rule of $\boldsymbol{H}$. As a result, the above problem can be rewritten as follows:

$$\begin{aligned} L(\boldsymbol{H}) = &\operatorname{tr}(\boldsymbol{XX}^{\mathrm{T}} - 2\boldsymbol{X}^{\mathrm{T}}\boldsymbol{ZH} + \boldsymbol{H}^{\mathrm{T}}\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{ZH}) \\ &+ \alpha \cdot \operatorname{tr}(\boldsymbol{HRDR}^{\mathrm{T}}\boldsymbol{H}^{\mathrm{T}}), \end{aligned} \tag{3}$$

where $\boldsymbol{D} \in \mathbb{R}^{n\times n}$ is the diagonal matrix. And the diagonal element of matrix $\boldsymbol{D}$ is defined in the form of $\boldsymbol{D}(j,j) = \frac{1}{||(\boldsymbol{HR})_j||}$, where $||(\boldsymbol{HR})_j||$ denotes the L_2 norm of vector

* Corresponding author (email: ghb48@ustc.edu.cn)

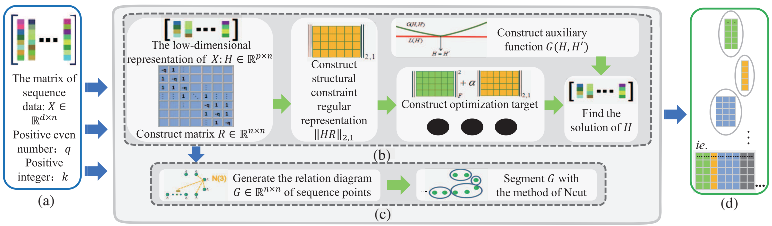

Figure 1 (Color online) The framework of the proposed method. (a) Input data and parameters; (b) find the low-dimensional representation of $X : H$; (c) clustering and video segmentation; (d) output segmentation results.

consisting of elements of column j of matrix $\boldsymbol{HR}$. Since $\boldsymbol{X}$, $\boldsymbol{Z}$ and $\boldsymbol{R}$ contain negative values, the data can be segmented into positive and negative parts. In the subsequent process, the positive and negative parts of the matrix can be represented by the symbols $[\boldsymbol{X}]^+ = (|\boldsymbol{X}| + \boldsymbol{X})/2$ and $[\boldsymbol{X}]^- = (|\boldsymbol{X}| - \boldsymbol{X})/2$, respectively.

To obtain the local minimum solution of $L(\boldsymbol{H})$, the constructed auxiliary function $G(\boldsymbol{H}, \boldsymbol{H}')$ is required to satisfy the following condition: $G(\boldsymbol{H}, \boldsymbol{H}') \geqslant L(\boldsymbol{H})$ and $G(\boldsymbol{H}, \boldsymbol{H}) = L(\boldsymbol{H})$, where $\boldsymbol{H}'$ represents a constant matrix. Thus, the solution of $G(\boldsymbol{H}, \boldsymbol{H}')$ can be regarded as a solution of $L(\boldsymbol{H})$.

Based on inequalities $z \geqslant 1 + \log z$ for any $z > 0$, $2ab \leqslant a^2 + b^2$ for $\forall a, b > 0$ and the correlation lemmas [1,2], clearly,

$$\begin{aligned}
&G(\boldsymbol{H}, \boldsymbol{H}') \\
&= \sum_{ik} \boldsymbol{X}_{ik}\boldsymbol{X}_{ik} - 2\sum_{ik} (\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{X})^+_{ik} \boldsymbol{H}'_{ik} \left(1 + \log \frac{\boldsymbol{H}_{ik}}{\boldsymbol{H}'_{ik}}\right) \\
&+ \sum_{ik} (\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{X})^-_{ik} \frac{\boldsymbol{H}^2_{ik} + \boldsymbol{H}'^2_{ik}}{\boldsymbol{H}'_{ik}} + \sum_{ik} ((\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{Z})^+\boldsymbol{H}')_{ik} \frac{\boldsymbol{H}^2_{ik}}{\boldsymbol{H}'_{ik}} \\
&- \sum_{ikl} \left((\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{Z})^-_{li} \boldsymbol{H}'_{ik}\boldsymbol{H}'_{lk} \left(1 + \log \frac{\boldsymbol{H}_{ik}\boldsymbol{H}_{lk}}{\boldsymbol{H}'_{ik}\boldsymbol{H}'_{lk}}\right)\right) \\
&+ \alpha \left(\sum_{ik} \left(\boldsymbol{H}' \left(\boldsymbol{RDR}^{\mathrm{T}}\right)^+ \right)_{ik} \frac{\boldsymbol{H}^2_{ik}}{\boldsymbol{H}'_{ik}} \right. \\
&\left. - \sum_{ikl} \left(\boldsymbol{H}'_{ik} \left(\boldsymbol{RDR}^{\mathrm{T}}\right)^-_{kl} \boldsymbol{H}'_{il} \right) \left(1 + \log \frac{\boldsymbol{H}_{ik}\boldsymbol{H}_{il}}{\boldsymbol{H}'_{ik}\boldsymbol{H}'_{il}}\right) \right) \qquad (4)
\end{aligned}$$

is an auxiliary function with regard to $L(\boldsymbol{H})$.

Let $\frac{\partial G(\boldsymbol{H},\boldsymbol{H}')}{\partial \boldsymbol{H}_{ik}} = 0$, and then the successive update of $\boldsymbol{H}$ with an initial value of $\boldsymbol{H}$ is expressed according to

$$\boldsymbol{H}_{ik} \leftarrow \boldsymbol{H}_{ik} \sqrt{\frac{(\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{X})^+_{ik} + ((\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{Z})^-\boldsymbol{H})_{ik} + \alpha(\boldsymbol{H}(\boldsymbol{RDR}^{\mathrm{T}})^-)_{ik}}{(\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{X})^-_{ik} + ((\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{Z})^+\boldsymbol{H})_{ik} + \alpha(\boldsymbol{H}(\boldsymbol{RDR}^{\mathrm{T}})^+)_{ik}}}. \qquad (5)$$

Segmentation of the relation graph. We use matrix $\boldsymbol{H}$ to generate the relation graph, and the graph cut technique [3] is then employed to classify the sequential points. Specifically, firstly, based on the spectral graph theory, matrix $\boldsymbol{H}$ is used to construct the relationship diagram $\boldsymbol{G} \in \mathbb{R}^{n\times n}$ of sequential point, where the vertex of the graph stands for the sequential point. The edge of the graph is obtained by calculating the similarity between the sequential points and other points on the matrix $\boldsymbol{H}$. The top k points with the highest similarity constitute the k-neighbor set $N(i)$ of the point i. The edge of $\boldsymbol{G}(i, j)$ is defined as

$$\boldsymbol{G}(i,j) = \begin{cases} 1, & \text{if } j \in N(i), \\ 0, & \text{otherwise.} \end{cases} \qquad (6)$$

Secondly, the Ncut method is adopted to classify $\boldsymbol{G}$. Relevant details can be found in [3, 4].

Experiments. In this study, four real datasets are used to carry out the experiment, namely, Weiz, Keck, and the no. 3 and no. 9 of CMU Mocap 86. The comparison algorithm includes many classical sparse subspace clustering algorithms, such as sparse subspace clustering (SSC), clustering with adaptive neighbors (CAN), semi-NMF, temporal subspace clustering (TSC), structured SSC (strSSC) and ordered robust NMF (ORNMF).

To quantitatively evaluate the performance of the clustering, we adopt the accuracy (AC) and normalized mutual information (NMI) metrics. The details about AC and NMI, please refer to [5].

By comparison these algorithms on 4 datasets, experimental results show that SCseNMF method is capable of extracting the feature information of the motion action sequence better, so that the neighbor sequential points are segmented into the same motion class accurately.

Conclusion and future work. This study develops a structure constraint regularization and proposes a structural constraint semi-non-negative matrix factorization method. The low-dimensional representation is used to generate the sequence similarity relation matrix. Finally, we apply the spectral clustering and graph segmentation methods to obtain the segmentation of the motion sequence. Experiments manifest that the proposed method can improve the accuracy of segmentation. For the purpose of segmenting the motion sequence more accurately, it is necessary to further investigate the segmentation method with respect to the starting point and ending point of motions, and the influence of data characteristics on the parameter setting in future.

Acknowledgements This work was supported by the Fundamental Research Funds for the Central Universities, the Science and Technology Innovation Planning Project of Ministry of Education of China, NVIDIA NVAIL Program, National Natural Science Foundation of China (Grant No. U1804161), and Key Laboratory of Advanced Perception and Intelligent Control of High-end Equipment of Ministry of Education (Anhui Polytechnic University, Wuhu, China, 241000) (Grant Nos. GDSC202001, GDSC202007). And experiments are conducted on NVIDIA DGX-2.

References

1 Li S, Li K, Fu Y. Temporal subspace clustering for human motion segmentation. In: Proceedings of the IEEE International Conference on Computer Vision, Santiago, 2015. 4453–4456
2 Liu H, Sun F. Material identification using tactile perception: a semantics-regularized dictionary learning method. IEEE/ASME Trans Mechatron, 2018, 23: 1050–1058
3 Shi J B, Malik J. Normalized cuts and image segmentation. IEEE Trans Pattern Anal Machine Intell, 2000, 22: 888–905
4 Liu H, Sun F, Zhang X. Robotic material perception using active multimodal fusion. IEEE Trans Ind Electron, 2019, 66: 9878–9886
5 Wang L, Ding Z, Fu Y. Low-rank transfer human motion segmentation. IEEE Trans Image Process, 2019, 28: 1023–1034