

Received November 27, 2017, accepted December 20, 2017, date of publication December 27, 2017, date of current version March 13, 2018.

Digital Object Identifier 10.1109/ACCESS.2017.2787675

Sensor Network Oriented Human Motion Segmentation With Motion Change Measurement

YANG LIU¹, LIN FENG², SHENGLAN LIU², AND MUXIN SUN³

¹Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology, Dalian 116024, China

²School of Innovation and Entrepreneurship, Dalian University of Technology, Dalian 116024, China

³State Key Laboratory of Software Architecture, Neusoft Corporation, Shenyang 110179, China

Corresponding author: Lin Feng (fenglin@dlut.edu.cn)

This work was supported in part by the National Natural Science Foundation, China, under Grant 61370200, Grant 61672130, and Grant 61602082, and in part by the Open Program of State Key Laboratory of Software Architecture under Grant SKLSAOP1701.

ABSTRACT Smart Internet of Things has greatly improved the quality of human life with increasingly intelligent sensor networks. Efficient and accurate human motion time series segmentation is the key issue in human motion analysis and understanding. To realize human motion sequence segmentation, a comprehensive human motion description and an intelligent segmentation algorithm are required. Hence, this paper proposes a sensor network-based human motion sequence segmentation framework. With the facilitation of sensor network and sensor network-based feature fusion method, human motions can be comprehensively described. Based on the comprehensive description of motion data, a new motion change variation-based segmentation method is proposed to realize human motion sequence segmentation. Moreover, to satisfy the time efficiency demand in the applications of large scale sensor networks, a hashing algorithm is introduced to compress the original captured sensor data, which can effectively represent the human motions with short binary codes and facilitate the motion change measurement. Experiments on real-world human motion data sets validate the effectiveness of our proposed sensor network-based human motion sequence segmentation framework compared with other state-of-the-art human motion segmentation methods.

INDEX TERMS Sensor networks, human motion sequence segmentation, hashing learning, motion change measurement.

I. INTRODUCTION

With the enormous development in the field of Internet of Things (IoT), smarter IoT systems have greatly promoted the development of society and raised the living standards of mankind [1]–[4], [8]. Different IoT systems have been applied to many walks of life [5]–[7]. In the fields of athletic training, medical diagnostics, and security monitoring, etc, smart IoT systems are utilized to analyze human motion time series. Based on the analysis, motion information which is of great significance to improve the quality of human life can be obtained [8]–[10]. For instance, in the process of medical diagnostics for heart diseases, monitoring the condition of heartbeat, breath, and motions through sensor networks can effectively facilitate the evaluation of the illness. Human activities can be effectively analyzed and deep mined through the facilitation of smart IoT systems, especially sensor networks [8], [12], [13].

Analysis on human motions by sensor networks not only focuses on the relations of different sensors, but also on the relations of the states of sensor networks in different moments [10], [11], [14]. The analysis on human motions and the derived technologies have improved the intelligence of the IoT [15], [16]. The fundamental issue in human motion research is Human Motion Sequence Segmentation (HMSS). The segmentation results significantly influence the analysis of human motion sequence. Hence, analyzing the time series generated by sensor networks and segmenting the human motion sequence are key in human motion analysis and understanding. HMSS has attracted more and more attention in research fields [17], [19], [22].

Various types of pattern recognition methods have been proposed to realize HMSS. For instance, clustering methods are employed by researchers to realize HMSS [18], [19], which are based on correlations of motions in the

same class. Aligned Cluster Analysis (ACA) combined kernel k-means with the generalized dynamic time alignment to realize HMSS by clustering the time series data [18]. Moreover, an unsupervised hierarchical bottom-up framework is adopted to embed the the low-dimensional human motion sequences. The bottom-up framework facilitates the efficiency of the clustering based HMSS methods [19]. However, clustering-based motion sequence segmentation methods require that the intra-cluster variance of time series should be low, which is quite difficult for real world applications. Furthermore, if the initial number of clusters is not perfectly set, the segmentation performance will be limited. To solve these problems, neighborhood similarity information of the time series is employed to construct a similarity graph to enhance the clustering performance [20]. Apart from clustering methods, classification methods are also considered in HMSS, which distinguish motion classes from each other [21]. Data Point Classification (DPC) labels data points as either segment points or non-segment points by using an online classifier. Based on the online classifier, data points labeled with segment points are detected as transition frames of the human motion time series. Inspired by DPC, detecting transition frame methods are adopted in HMSS. Kernelized Temporal Cut (KTC) is proposed to cut structured sequential data into different regimes sequentially [22]. KTC can detect the transition frames as well as repetitive frames in a human motion time series simultaneously.

Apart from the aforementioned segmentation methods, dimensional reduction methods are also adopted in HMSS process [17], [23]. For example, the Principle Component Analysis (PCA) [24] approach to human motion segmentation employs the intrinsic dimension of human motion to realize segmentation, and assigns a cut when the intrinsic dimension of a local neighborhood in the motion time series suddenly increases [23]. Locality Linear Projections (LPP) approach to HMSS utilizes the local neighborhood structure to detect segment points [17], [25]. The Probability Principle Component Analysis (PPCA) [26] to human motion segmentation analyzes the distributions of motions of the time series to detect the segmentation points. That is, the PPCA approach assumes that frames from the same class should obey the same probability distribution, and places a cut when the distribution of human motions changes [23]. Unlike the PCA and PPCA approaches which only focus on the data, a physical driven motion segmentation method is proposed to realize human motion segmentation [17]. Time series-Warp Metric Curvature Segmentation (TS-WMCS) constructs a curvature-like descriptor to evaluate the changes in human motions. TS-WMCS utilizes the physical property that motions from the same class are alike in a given time series, and detects the transition points when the local curvature of a frame changes greatly [17].

In general, the key issue in HMSS is to detect the segmentation points in the sequences. Note that, human motions will change significantly near the segmentation points [17], [23]. Based on this fact, this paper proposes

a sensor network based segmentation (SNBS) method for HMSS. With the facilitation of sensor networks, human motions can be effectively recorded and captured from different perspectives. Moreover, a motion variation description method is proposed to depict changes of human motions in a given time series, and the motion of each frame is regarded as a bucket in the hashing table. Based on the physical properties of human motions [17], [23], movements from the same class will not change dramatically in the time series. However, due to the appearance of motions from new classes, variations in a motion sequence will significantly increase near the segmentation points. Moreover, to promote the efficiency of SNBS, hashing method [33] is employed to describe the bucket of each frame in hamming space. With the facilitation of hashing methods, data similarities are represented effectively.

The main contributions of this paper can be summarized as follows.

Firstly, propose a data fusion method to organically combine the data collected from different kinds of sensors.

Secondly, employ a hashing method to represent the original sensor network data, which projects original data to hamming space to facilitate the description of motion change.

Thirdly, realize HMSS from the perspective of motion change degree.

The remainder of this paper is organized as follows: In section II, mainstream HMSS methods are introduced. In section III, the SNBS method for HMSS will be detailedly presented. Experimental results are provided in section IV, which is followed with the conclusion and future work in section V.

II. RELATED WORKS

This section mainly introduces the mainstream HMSS methods. Generally, HMSS can be divided as data driven methods and physical driven methods [17]. In section II-A, the PCA and locality preserving projections (LPP) approaches to HMSS will be introduced, which utilizes the low dimensional embedding of original data to realize HMSS. In section II-B, the PPCA approach to HMSS will be introduced, which evaluates motion changes through the motion distributions. Unlike data driven methods (PCA, LPP and PPCA approaches to HMSS), physical driven methods pay more attention to the physical meaning of the human motions and realize HMSS by researching the motion change process. As a typical physical driven segmentation method, TS-WMCS algorithm will be introduced in section II-C.

A. PCA AND LPP APPROACHES TO HMSS

Both the PCA and LPP approaches to HMSS focus on the data of the human motions in each frame, this section will briefly introduce the PCA and LPP approaches. Given human motion sequence $X = [x_1, \dots, x_n]^T \in \mathbb{R}^{n \times D}$ with n frames and D dimensions. For each frame x_i , utilizing $k - NV$ method to construct the neighborhood X_i for x_i , where $X_i = \{x_1, \dots, x_j, \dots, x_k\} \in \mathbb{R}^{k \times D}$ and $j \in N_i$. Therein, k is the size

of the neighborhood and N_i is the collection of the indices of X_i .

The PCA approach to HMSS assumes that the error between neighborhood X_i and its projection X'_i will not vary largely for motions from the same class. However, the error will increase largely for the frames near the transition clips [27], [28]. To evaluate the error e_i in each frame, original data X_i is reduced to its intrinsic dimension [23]. The error e_i between X_i and X'_i can be expressed as Eq.(1).

$$e_i = \sum_{i=1}^n \|X_i - X'_i\|^2 \tag{1}$$

The PCA approach to HMSS works well if motion subsequences of different classes have clear transition clips and motions from the same class will not change significantly. Nevertheless, since the complexity of real world human motions and the continuity of real world movements, the error between X_i and X'_i will not change significantly in time series. Under this circumstance, the PCA approach to HMSS can hardly detect all the transition clips in the human motion time series [17].

Inspired by PCA approach to HMSS, LPP can also be applied in HMSS [25]. Like PCA approach, LPP utilizes a linear projection to map original data to low dimensional embedding as well. However, LPP aims at preserving local similarities in the projection process. The similarity between sample x_i and x_j can be represented by $w_{ij} = e^{-\frac{\|x_i - x_j\|_2^2}{\sigma}}$ ($\sigma = 1$ in our experiment), where $e^{(\cdot)}$ is the exponential function.

LPP approach to HMSS also utilizes an error e_i of each frame to evaluate the motion change degree [17]. The error can be calculated by Eq.(2)

$$e_i = \sum_{j \in N_i} w_{ij} (y_i - y_j)^2 \tag{2}$$

where y_i and y_j represent the low dimensional embedding of x_i and x_j , N_i denotes the adjacent frames of the i th frame.

LPP utilizes the local structure on human motion sequence to detect transition clips. However, since the complexity of real-world motions, LPP approach to HMSS can not effectively detect the motion changes in human motion time series [17].

B. PPCA APPROACH TO HMSS

The PPCA approach to HMSS utilizes the probability principle component analysis method to evaluate the distribution of motions in time series. The PPCA approach to HMSS utilizes the Gaussian distribution to extend the traditional PCA approach, and represents the relations of different motions with a correlation covariance matrix C . The correlation covariance matrix C can be utilized to compute the average Mahalanobis distance H of the whole neighborhood of the i th frame, which can evaluate how likely are motions from $(i + 1)$ th frame to $(i + k)$ th frame belong to the Gaussian distribution [23]. The calculation of H can be expressed as

Eq. (3).

$$H = \frac{1}{T} \sum_{j=i+1}^{i+k} (x_j - \bar{x})^T C^{-1} (x_j - \bar{x}). \tag{3}$$

Thereinto, i is the index of the start frame of the neighborhood X_i . Unlike the PCA approach to HMSS, the PPCA approach to HMSS simply selects the next k motions for x_i to construct neighborhood instead of adopting $K - NN$ method. The PPCA approach to HMSS can effectively capture the correlation and variance of different joint angles respectively [23]. However, the Gaussian distribution assumption for the human motions is too strong for real world applications (Especially for complex human motions, such as Tai Chi), which will lower the segmentation accuracy of PPCA segmentation approach. Besides, Both the PCA and PPCA human motion sequence segmentation methods are based on angles, which will cause segmentation accuracy decrease when they are extended to real world applications [17].

C. TIME SERIES-WARP METRIC CURVATURE SEGMENTATION (TS-WMCS) ALGORITHM

This section mainly introduces the Time series-Warp Metric Curvature Segmentation (TS-WMCS) algorithm to HMSS [17]. TS-WMCS utilizes a curvature-like descriptor to depict the the changes of human motions in time series. For each frame $x_i \in X$, construct the neighborhood X_i for x_i with $k - NN$ method like the PCA approach to HMSS. Then, construct the angle between the samples and the tangent space in each neighborhood. The warp degree of the data can be described by the angle between x_j and its orthogonal projection, where x_j is the neighbor of x_i . Therein, $\alpha_{ij} = \alpha_j(Q_i)$, and $\alpha_{ij} \in [0, \pi/2]$. The local low dimensional space can be expanded by Q_i , and Q_i can be obtained by the optimization equation Eq.(4),

$$\arg \max_{Q_i} Tr(Q_i^T X_i Z X_i^T Q_i) \quad s.t. \quad Q_i^T Q_i = I \tag{4}$$

where Z is the normalization matrix of $X_i X_i^T$. Then, the curvature-like descriptor can be expressed as Eq.(5).

$$c_i = \sum_{j \in N_i} \cos \alpha_{ij} / \sum_{j \in N_i} \|x_j\| \tag{5}$$

The transition points in human motion sequences are those frames whose curvature-like descriptors are high. Apart from the curvature-like descriptor segmentation, TS-WMCS algorithm also reduces the dimension of original motion sequences. The low dimensional temporal feature curves are utilized to assist the segmentation. For each neighborhood, compute the low dimensional embedding Θ_{k_i} of X_i , and then add affinity projection L_i for each Θ_{k_i} of X_i to map the low dimensional embedding to a global embedding $T = [\tau_1, \tau_2, \dots, \tau_n]^T \in R^{n \times d}$. Hence, the low dimensional

embedding T can be obtained by Eq.(6).

$$\min_{\tau_i, L_i} \sum_{i=1}^n c_i \left\| \left(T_{k_i} - \tau_i e_{k_i}^T \right) - L_i \Theta_{k_i} \right\|^2 \quad s.t. \quad T T^T = I_d \quad (6)$$

where I_d is a $d - by - d$ identity matrix, and e_{k_i} is a vector with all ones.

TS-WMCS algorithm utilizes the curvature-like descriptor to realize HMSS, which can effectively depict the changes of human motion sequences. However, a certain move in complex motion sequences will contain motions which are quite different from each other (e.g. “white crane spreads its wing” in Tai Chi), TS-WMCS algorithm can not yield promising result on these kind of motion sequences [17]. Moreover, TS-WMCS algorithm employs the low dimensional temporal feature curves to assist the segmentation, which is lack of clear physical significance.

III. SENSOR NETWORK BASED SEGMENTATION FOR HMSS

In this section, we will introduce the proposed sensor network based segmentation (SNBS) approach for HMSS. SNBS utilizes hamming distance to describe the change degree of motions in time series, which is based on the binary representations of the motion data collected by sensor network. The sketch of SNBS is summarized as Fig. 2. In real-world applications, a sensor network may utilize different kinds of sensors to collect the motion data [29], [30]. Hence a feature fusion method is proposed to combine the data collected by different sensors (detailed introduction is in section III-A). Note that, motions of the same class change slightly while motions of different classes change significantly [17], [23]. Hence, we construct the local change degree to fully reveal the change of motions (details are in section III-C). To accurately describe the motion changes in time series, the binary representations should represent the similarities and differences of motions simultaneously. Considering the property of hashing methods in preserving data similarities with binary codes (hash buckets), we utilize hashing method to represent the motions. Here, we employ the widely utilized hashing method, Iterative Quantization Hashing (ITQ) [33], to represent the motions. Detailed introduction of ITQ is in section III-B.

A. FEATURE FUSION FOR MULTI-SOURCE DATA COLLECTED BY MULTI-SENSORS

Segmentation for human motion sequence not only requires effective segmentation methods but also requires comprehensive description of the human motions. In real-world applications, sensor network is always adopted to record the human motions [29], [30]. However, different kinds of sensors may adopted by a sensor network to fully describe the poses of human motions [23], [35], which causes the problem of representing human motions with features of different scales. To solve this problem, a Sensor Network Based Feature

Fusion (SNBFF) method is proposed in this paper to align the multi-source data collected by different kinds of sensors.

Different kinds of sensors reflect human motions from different perspectives. To effectively utilize the data obtained from different kinds of sensors, we employ a graph based feature fusion method named Feature Graph Fusion (FGF) to combine the data collected by different sensors [34]. To clearly clarify the process of the feature fusion method, we assume that three kinds of sensors are adopted in a sensor network (motion sensors, depth sensors and visual sensors). SNBFF can organically combine the collected data, and help describing the motion of each frame with a single descriptor.

We employ the graph fusion method in [34] to combine data collected by the aforementioned three kinds of sensors. For each kind of sensors, Jaccard coefficient [37] is adopted to measure the similarities of the collected data. Based on the Jaccard coefficient, the similarity graph of the data collected by each kind of sensors can be constructed. The similarity graph of motion movement data can be represented as $G^m = (V^m, E^m, w^m)$, the similarity graph of motion depth data can be denoted by $G^d = (V^d, E^d, w^d)$, and the similarity graph of motion data collected by visual sensors can be defined as $G^c = (V^c, E^c, w^c)$. All the similarity graphs are undirected graphs. Therein, V^m, V^d and V^c denote the frames of the motion sequence, E is the edge constructed by Jaccard similarity coefficient for V , and w represents the weight of E . The fused similarity graph $G = (V, E, w)$ can be obtained by fusing G^m, G^d and G^c with the following constraints. First, $V = V^m \cup V^d \cup V^c$. Second, $E = E^m \cup E^d \cup E^c$. Third, $w = w^m \cup w^d \cup w^c$. Hence, the similarity weight of the fused graph can be represented as $W = [W_{ij}]$, where $i, j \in V$ are two motions of the motion sequence. Then, we can get the normalized similarity description matrix $W = [W_{ij}]$, where $i, j \in V$. W_{ij} can be normalized by Gaussian kernel as Eq.(7).

$$W_{ij} = \begin{cases} \frac{\exp\left(-\frac{W_{ij}}{2\sigma_i^2}\right)}{\sum_{j' \in N_k(i)} \exp\left(-\frac{W_{ij'}}{2\sigma_i^2}\right)}, & j \in N_k(i) \\ 0, & else \end{cases} \quad (7)$$

where σ_i is the bandwidth parameter of Gaussian kernel [38], j is the indices of the neighborhood frame of the i th frame. In this paper, σ_i is assigned by the variance of similarities of the i th frame.

After obtaining the similarity weight $W = [W_{ij}]$ of the fused graph. The fused motion feature can be obtained by Eq.(8).

$$\max_F \prod_{i,j \in V} \left(\frac{\exp(f_i^T f_j)}{\sum_{j \in V} \exp(f_i^T f_j)} \right)^{W_{ij}} \quad (8)$$

where $\{f_1, f_2, \dots, f_n\} \in \mathbb{R}^{D \times 1}$, f_i is the fused feature of the i th motion in the motion sequence. Eq.(8) can be expressed

as the form of log function as Eq.(9).

$$\begin{aligned} \ell &= \sum_{i,j \in V} W_{ij} \cdot \log \left(\frac{\exp(f_i^T f_j)}{\sum_{j \in V} \exp(f_i^T f_j)} \right) \\ &\approx \frac{\sum_{i \in V} \sum_{j \in s(i)} \log \left(\frac{\exp(f_i^T f_j)}{\sum_{j \in V} \exp(f_i^T f_j)} \right)}{M} \end{aligned} \quad (9)$$

where $s(i)$ denotes sampling M times as the distribution of the neighborhood similarities of the i th frame. The solution of the fused motion feature $F = [f_1, f_2, \dots, f_n] \in \mathbb{R}^{D \times n}$ can be obtained by word embedding model [31], [32], [36]. The whole process of SNBFF is summarized as Fig.1.

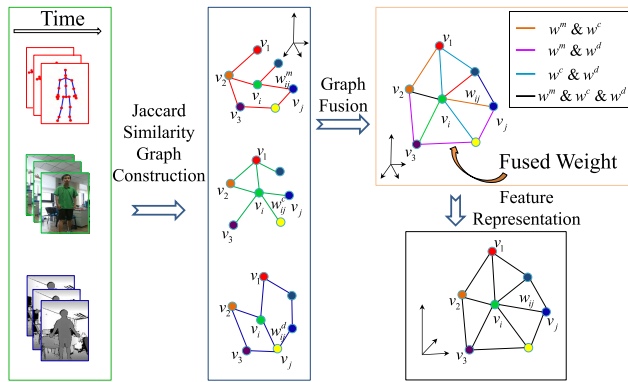


FIGURE 1. Summary of human motion feature capturing with various kinds of sensors.

B. ITERATIVE QUANTIZATION HASHING (ITQ)

To fully keep the information of movements in human motion time series, the corresponding binary representations of motions should preserve the action similarities. Considering the ability of hashing method in preserving data similarities to hamming space [33], [39]–[41] and compacting data storage. SNBS adopts a typical hashing method, Iterative Quantization hashing (ITQ), to realize data encoding. ITQ can effectively facilitates representing human motions with short binary codes. This section will detailedly introduce the process of ITQ.

For motion sequence data X , utilizing the PCA method to reduce the dimension of original data like many hashing methods [33], [40]. The affinity matrix P for reducing the dimension of original data can be obtained by Eq.(10).

$$\arg \min_P E(P) = \arg \min_P \|(X - \bar{X})P\|_F^2 \quad (10)$$

where \bar{X} is the centralization matrix of X . By utilizing the affinity matrix, the low-dimensional embedding Y can be obtained as $Y = XP$. Based on the low-dimensional data Y , a rotation matrix R will be calculated by ITQ to further adjust the low-dimensional embedding [33], which significantly improved the performance of ITQ. The rotation matrix can be calculated by Eq.(11).

$$\arg \min_{B,R} Q(B, R) = \|B - YR\|_F^2 \quad (11)$$

where $B \in \{-1, 1\}^{n \times c}$ is the binary representations of original data X , and c denotes the length of binary codes. Eq. (11) can be transformed as follows:

$$\begin{aligned} \arg \min_{B,R} Q(B, R) &= \|B\|_F^2 + \|Y\|_F^2 - 2\text{tr}(BR^T Y^T) \\ &= nc + \|Y\|_F^2 - 2\text{tr}(BR^T Y^T) \end{aligned} \quad (12)$$

EM algorithm is utilized to solve Eq.(12) [33].

1) FIX R , CALCULATE B

Since the determinant of a rotation matrix should be 1, R can be initialized with an orthogonal matrix. Then, Eq.(12) can be transformed as Eq.(13), and the optimal value B is the binary representation of original data.

$$\arg \max_B \text{tr}(BR^T Y^T) = \arg \max_B \sum_{i=1}^n \sum_{j=1}^c B_{ij} \tilde{Y}_{ij} \quad (13)$$

where $\tilde{Y} = R^T Y^T$.

2) FIX B , CALCULATE R

When B is calculated, rotation matrix R can be optimized by Eq.(14).

$$\arg \max_R \text{tr}(BR^T Y^T) = \arg \max_R (S^T B^T Y S) \quad (14)$$

where $R = SS^T$. Since Eq.(14) is a quadratic form problem, the optimal S can be calculated by operating SVD decomposition for $B^T Y$ [48]. Based on S , the rotation matrix can be expressed as $R = SS^T$.

With the facilitation of ITQ, original motion sequence can be represented with a series of binary codes, which explicitly represent the original data relations and effectively compact the data storage. In next section, the binary representation of original motion data are utilized to measure motion change degree in the time series. The change process of human motions is described with binary representations.

C. SNBS WITH HASHING BASED MOTION CHANGE MEASUREMENT

This section mainly introduces the way to construct binary representations of human motion sequence by ITQ, and the way to evaluate the motion change degree by SNBS. Considering the temporal locality of human motion sequence where adjacent motion frames are similar in the time series [17], [18]. Note that the similarities of motions from the same class will be higher than that of motions belong to different classes. Hence, a human motion sequence can be segmented by detecting the similarity change degree of each motion frame in time series.

Hashing learning aims to represent original data with binary codes and preserves the original data similarity to hamming space. With the excellent performance in compacting data storage and preserving data similarities, hashing learning have been widely applied in different kinds of applications [43]–[45]. Based on hashing methods, original data

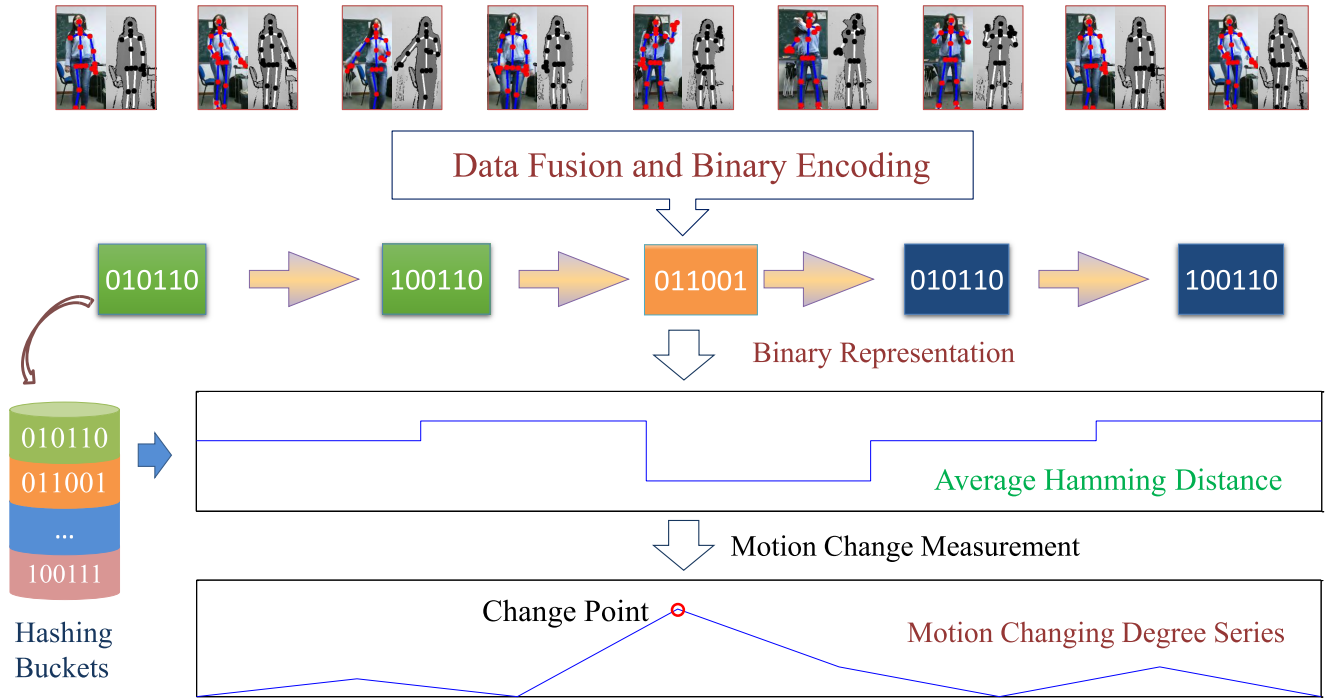


FIGURE 2. The sketch of SNBS in constructing the motion change measurement.

can be mapped to hashing buckets and similar data will be mapped to close hashing buckets. Therefore, motions of the same class will be mapped to buckets which are close measured by hamming distance. When motions of new classes appear, new buckets will also appear. Hence, we can utilize hashing method to project original data to hamming space to effectively measure the motion changes. With the facilitation of hashing learning, human motion sequence can be segmented into many motion clips, where each clip belongs to a hashing bucket and represents a motion class.

For human motion sequence X , each frame x_i can be represented with binary codes $b_i \in \{1, 0\}^c$ with the hashing method introduced in previous section. Therein, c is the length of binary representation. For each frame x_i , compute the average hamming distance of neighborhood X_i

$$D(x_i) = \frac{1}{k} \sum_j^k Dist(b_i, b_j), \quad x_j \in N(x_i) \quad (15)$$

where k is the number of neighborhood for x_i , $Dist(\cdot)$ represents the distance metric function (Here, hamming distance is employed to measure the similarity of two binary codes), $N(x_i)$ denotes the neighborhood of x_i . ITQ preserves the original data structure with binary representations, that is $Dist(x_i, x_j)$ is proportional to $Dist(b_i, b_j)$ [33], [40]. The maximum distance of b_i and b_j is related to the length of b_i and b_j . Since the distance of two binary representations is discrete and the number of total binary codes is limited, similar motion frames will be projected to nearby hashing buckets and the corresponding hamming distance will be

small. On the other hand, motions of different classes will be projected to hashing buckets which share large hamming distance. To sum up, the changes of average hamming distances of motions from the same class will be small. On contrast, when motions of a new classes appear, the average hamming distance will change dramatically. Therefore, we can describe the motion change by the change of average hamming distance of each frame.

To explicitly reveal the changes of average hamming distances of human motions and effectively detect the transition points in human motion sequence, the change degree of each frame can be represented by Eq.(16).

$$M_i = \frac{e^{|D(x_i) - D(x_{i-1})|}}{e^{D(x_i)}}, \quad i \geq 2 \quad (16)$$

where $e(\cdot)$ is the exponential function. With the change degree of motions at each moment, the changes of human motions can be quantitatively measured. Since motions from the same class have large probabilities in sharing same binary representations, the average hamming distances for frames located beyond transition clips will be zero with a high probability. As is shown in Fig. 3, most of the average hamming distances are 0 on CMU motion capture database (each frame is represented with 8 bits as an example). Hence, exponential function is utilized to avoid the value of denominator to be zero in Eq.(16). Based on the change of M_i , the change points can be effectively detected. Take the change degree of human motion sequence No. 15_01 in CMU motion capture database as an example. The change degree is large enough

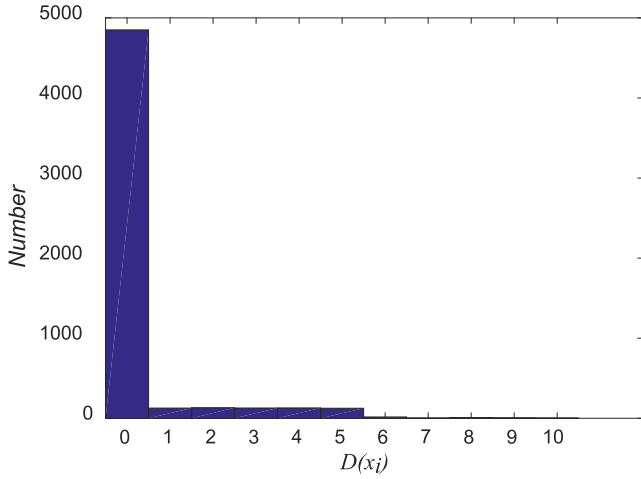


FIGURE 3. An example of the average hamming distance distribution on CMU motion capture database (8 bits).

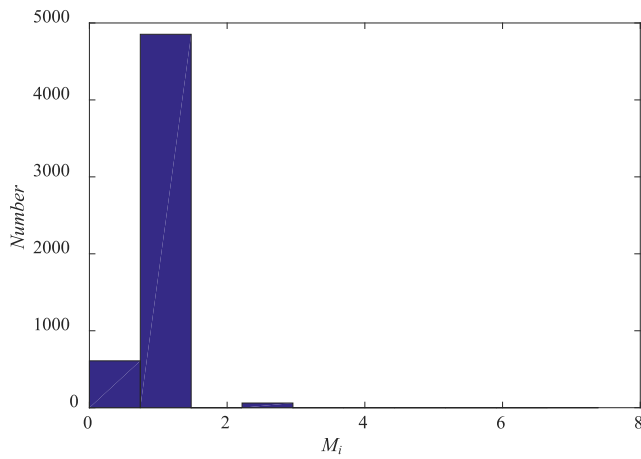


FIGURE 4. An example of the change degree distribution on CMU motion capture database (8 bits).

for detecting the segment points in time series (as is shown in Fig. 4).

Segment points in human motion sequence are those frames located in the motion transition clips, where motions of new classes appear and movements change greatly. Based on this physical property, the change degree of human motions in time series can be utilized to realize HMSS. In most cases, the motion changes are smooth and steady. However, the change degree will increase dramatically when motions of a new class appear. Hence, we can set a change degree threshold θ to select the segment points (In our experiment, θ is set as twice that of the average motion change degree of a sequence). When change degree M_i is larger than threshold θ , add the i th frame in the motion sequence into the segment points collection $C = [c_1, \dots, c_m] \in \mathbb{N}^{1 \times m}$. Note that, m is the number of evaluated cuts in human motion time series.

$$C = \{M_i | M_i \geq \theta\} \in \mathbb{N}^{1 \times m} \quad (17)$$

The process of SNBS can be summarized as Algorithm 1.

Algorithm 1 SNBS for HMSS

Input: Human motion data $X_1, X_2 \dots X_L$ captured by different sensor networks

Output: Segment cuts C

1. Construct Jaccard similarity graph G^i for data X_i collected from different sensor networks.
2. Fuse the Jaccard similarity graphs according to section III-A, and construct the fused graph weight by Eq.(7).
3. Compute the fused human motion sequence representation F according to section III-A.
3. Compute low dimensional embedding Y of fused human motion sequence F with PCA.
4. Obtain the binary Representation of Y with ITQ method.
5. Compute the average hamming distance series by Eq.(15).
6. Generate the change degree series M of human motion sequence by Eq.(16).
7. Detect the change degree series.
 - if $M_i > \theta$
 - Add M_i to C
 - end
8. Output the segmentations C .

IV. EXPERIMENT RESULTS

To validate the effectiveness of SNBS in segmenting human motion sequences, we conduct experiments on CMU motion capture database. CMU motion capture database has been adopted by many HMSS methods to evaluate the segmentation performance [17]–[19], [22], [23]. Moreover, human motion segmentation experiments are also conducted on DUT human motion dataset [34]. Data of the DUT human motion dataset is collected by our human motion capture system, which contains posture sensors, depth sensors and visual sensors. The introduced PCA, LPP, PPCA and TS-WMCS segmentation algorithms are adopted as comparison methods to detailedly evaluate the performance of SNBS. In our experiments, PCA and LPP reduce the dimension of original data to the intrinsic dimension which can preserve 90% information of original data [17], [23].

To evaluate the performances of SNBS and the comparison methods more precisely, three widely used protocols are employed in our experiments [17], [22], [23]. The protocols measure the performances of HMSS methods from different perspectives, which are precision Pre , recall Rec and F-Measure F . Therein, $Pre = N_{right} / N_{seg}$, $Rec = N_{right} / N_{total}$ and $F = Pre * Rec * 2 / (Pre + Rec)$. Note that, N_{right} is the number of right segmentation points calculated by the segmentation methods, N_{seg} represents the total number of segmentation points generated by the segmentation methods and N_{total} denotes the number of transition clips in a human motion sequence. F-Measure evaluates the performance of segmentation methods from the perspectives of



FIGURE 5. The sensors on human body of CMU motion capture database. (a) Sensor Locations from front view. (b) Sensor Locations from back view.

segmentation precision and recall simultaneously, which can provide more comprehensive evaluation [49].

A. EXPERIMENTS ON CMU MOTION CAPTURE DATABASE

CMU motion capture database contains common human motions of 144 subjects, which are represented with skeletons of 31 joints. 31 motion sensors are located in the joints to capture the motion process of humans (Details in Fig. 5). However, the CMU Motion Capture Database only utilizes the motion sensors to collect the motion data [23], [50], [51]. In our experiment, we randomly select the motion sequences of trail 1 to trail 5 in subject 15 of CMU motion capture database to evaluate the segmentation performances of HMSS methods. The selected sequences range from 5524 frames to 22948 frames, and are mainly composed of every-day behaviors (e.g. wandering, waving and cleaning windows, etc.).

SNBS adopts the motion change measurement to detect the transition clips in human motion time series, which provides a precise description of the human motion change process. From Table 1, we can find that SNBS achieves better performances both in precision and recall compared to the data driven HMSS methods (PCA, LPP and PPCA approaches. From Fig. 6, we can find that SNBS provides more precise segment points compared to the PCA, LPP, PPCA and TS-WMCS HMSS methods.

In the process of detecting segment points, SNBS takes the current motion change and history motion changes into consideration simultaneously. This mechanism can effectively avoid the invalid segmentations in the same motion class. As is shown in Table 1, the segmentation recall of TS-WMCS

TABLE 1. Comparisons of precisions and recalls of 4 segmentation methods on CMU motion database.

	N_{total}	N_{seg}	N_{right}	Pre (%)	Rec (%)	F (%)
SNBS	21	87	13	14.94	61.90	24.08
TS-WMCS	21	153	13	8.55	61.90	15.02
PCA	21	59	1	1.69	4.76	2.49
LPP	21	5	0	0	0	0
PPCA	21	91	7	7.69	33.33	12.50

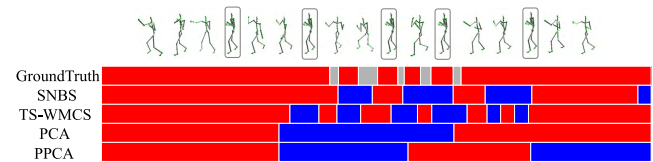


FIGURE 6. An example of segmentation results on CMU database.

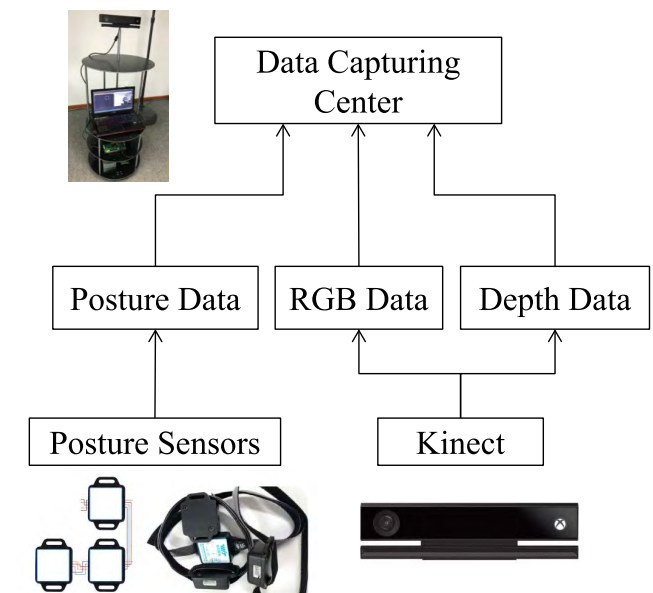


FIGURE 7. The structure of motion capture system for selecting human motion data for DUT human motion dataset.

is the same as SNBS. However, the segmentation accuracy of TS-WMCS is about 40% lower than SNBS. The main reason is that, TS-WMCS generates too many invalid segment points. Note that, TS-WMCS generates nearly 2 times segment cuts than that of SNBS. If evaluating the performances of these segmentation methods from the perspectives of precision and recall simultaneously, we can find that SNBS outperforms the other three segmentation methods obviously (over 35%).

B. EXPERIMENTS ON DUT HUMAN MOTION DATASET

Compared to the CMU motion capture dataset, the DUT human motion dataset is composed of more complex human motions [34]. The motion data is captured by our multi-sensors based human motion capture system. The structure of the motion capture system is shown in Fig. 7. The motion capture system utilizes three kinds of sensors to collect motion

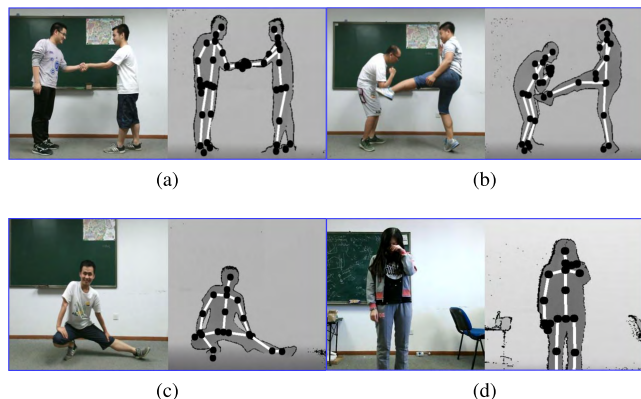


FIGURE 8. Samples of the DUT human motion dataset. (a) Greeting. (b) Wrestling. (c) Daily exercise. (d) Motions under different moods.

data, which are posture sensors, RGB camera and depth sensor. In the motion capture system, kinect is utilized to provide the RGB camera and depth sensor. The multi-sensors based human motion capture system can capture human posture information, motion scene, and motion depth information simultaneously. After capturing the motion data through different sensors, the SNBFF method is utilized to fuse the multi-source data and to generate a uniform representation for the human motion sequence.

The DUT human motion dataset not only contains single human motion sequences but also contains human-human interaction motion sequences. In this experiment, we simply select 2 human-human interaction motion sequences, and 2 single human motion sequences. The 2 human-human interaction motion sequences contain different kinds of interactions, including greeting and martial art. On the other hand, the 2 single human motion sequences contain human exercise motions and human motions under different moods. An example of DUT human motion dataset is shown in Fig. 8.

The DUT human motion dataset contains more complex and shorter human motions compared to the CMU motion capture database, which requires the segmentation methods to be more sensitive to the motion changes. The PCA, LPP and PPCA approaches only focus on the data changes in the time series [17], [18], and cannot achieve satisfying segmentation results when the human motions are complex. As is shown in Table 2, PCA, LPP and PPCA approaches cannot accurately detect the motion changes in the experiments. On the other hand, two physical driven segmentation methods, SNBS and TS-WMCS approaches, can effectively capture the motion changes in the time series. The main reason is that physical driven segmentation methods are more sensitive to the changes in human motion sequences.

SNBS employs motion change degree to evaluate the changes of human motion sequences, which not only focuses on the real-time motion change but on the history change process of human motions in time series. Hence, SNBS can effectively detect the motion changes in the time series. According to Table 2, SNBS achieves better performance in detecting the segment points compared to the other

TABLE 2. Comparisons of precisions and recalls of 4 segmentation methods on DUT human motion dataset.

	N_{total}	N_{seg}	N_{right}	Pre (%)	Rec (%)	F (%)
SNBS	14	42	9	21.43	64.29	32.14
TS-WMCS	14	24	4	16.67	28.57	21.05
PCA	14	2	0	0	0	0
LPP	14	3	1	7.14	33.33	11.76
PPCA	14	6	0	0	0	0

HMSS methods both in segmentation accuracy and recall. TS-WMCS adopts the curvature-like descriptor to describe the changes of human motion sequence, which only focuses on the real-time changes in a human motion sequence. Therefore, TS-WMCS can not accurately detect the transition clips in human motion sequences, especially when the human motions of a class are complex and change greatly. As is shown in Table 2, the segmentation accuracy of TS-WMCS is poorer than SNBS. The main reason is that SNBS can evaluate human motion precess more comprehensively by utilizing the hashing based data similarity to evaluate the change degree of human motions.

V. CONCLUSION

This paper proposes a new sensor network oriented method for HMSS named sensor network based segmentation (SNBS). SNBS analyzes human motion sequences from the perspective of motion similarities, and employs hashing method to construct motion change degree measurement to reveal the motion changes in time series. Based on the constructed motion change measurement, the change process of a human motion sequence is described with change degree time series. The change degree considers the real-time change of human motions and the history change process in the human motion sequences simultaneously. To precisely evaluate the change degree of human motions, a motion change evaluation criterion is constructed in this paper. The constructed evaluation criterion can effectively reveal the change points in the human motion sequences and avoid meaningless cuts for HMSS tasks. Experimental results validate the effectiveness of the proposed method in detecting the change points compared to several state-of-the-art methods.

Our future work will mainly focus on the following issues.

Firstly, how to employ supervised information to improve the segmentation accuracy of HMSS.

Secondly, how to extend the HMSS methods to real-time analysis of time-series.

REFERENCES

- [1] A. Mainwaring, D. Culler, J. Polastre, R. Szewczyk, and J. Anderson, "Wireless sensor networks for habitat monitoring," in *Proc. 1st ACM Int. Workshop Wireless Sensor Netw. Appl.*, 2002, pp. 88–97.
- [2] Z. He, Z. Cai, J. Yu, X. Wang, Y. Sun, and Y. Li, "Cost-efficient strategies for restraining rumor spreading in mobile social networks," *IEEE Trans. Veh. Technol.*, vol. 66, no. 3, pp. 2789–2800, Mar. 2017.
- [3] C.-W. Tsai, "An effective WSN deployment algorithm via search economics," *Comput. Netw.*, vol. 101, no. 4, pp. 178–191, 2016.
- [4] T. Qiu, A. Zhao, F. Xia, W. Si, and D. O. Wu, "ROSE: Robustness strategy for scale-free wireless sensor networks," *IEEE/ACM Trans. Netw.*, vol. 25, no. 5, pp. 2944–2959, Oct. 2017.

- [5] T. Qiu, K. Zheng, M. Han, C. L. P. Chen, and M. Xu, "A data-emergency-aware scheduling scheme for Internet of Things in smart cities," *IEEE Trans. Ind. Informat.*, to be published, doi: [10.1109/TII.2017.2763971](https://doi.org/10.1109/TII.2017.2763971).
- [6] F. Xiao, Z. Wang, N. Ye, R. Wang, and X.-Y. Li, "One more tag enables fine-grained RFID localization and tracking," *IEEE/ACM Trans. Netw.*, to be published, doi: [10.1109/TNET.2017.2766526](https://doi.org/10.1109/TNET.2017.2766526).
- [7] F. Xiao, W. Liu, Z. Li, L. Chen, and R. Wang, "Noise-tolerant wireless sensor networks localization via multi-norms regularized matrix completion," *IEEE Trans. Veh. Technol.*, to be published, doi: [10.1109/TVT.2017.2771805](https://doi.org/10.1109/TVT.2017.2771805).
- [8] C. Perera, A. Zaslavsky, P. Christen, and D. Georgakopoulos, "Sensing as a service model for smart cities supported by Internet of Things," *Trans. Emerg. Telecommun. Technol.*, vol. 25, no. 1, pp. 81–93, 2014.
- [9] C. M. Medaglia and A. Serbanati, "An overview of privacy and security issues in the Internet of Things," in *The Internet Things*. New York, NY, USA: Springer, 2010, pp. 389–395.
- [10] T. Qiu, R. Qiao, and D. O. Wu, "EABS: An event-aware backpressure scheduling scheme for emergency Internet of Things," *IEEE Trans. Mobile Comput.*, vol. 17, no. 1, pp. 72–84, Jan. 2018, doi: [10.1109/TMC.2017.2702670](https://doi.org/10.1109/TMC.2017.2702670).
- [11] S. Cheng, Z. Cai, J. Li, and H. Gao, "Extracting kernel dataset from big sensory data in wireless sensor networks," *IEEE Trans. Knowl. Data Eng.*, vol. 29, no. 4, pp. 813–827, Apr. 2017.
- [12] Y.-K. Chen, "Challenges and opportunities of Internet of Things," in *Proc. 17th Asia South Pacific. Design Autom. Conf. (ASP-DAC)*, Jan./Feb. 2012, pp. 383–388.
- [13] L. Tan and N. Wang, "Future Internet: The Internet of Things," in *Proc. 3rd Int. Conf. IEEE Adv. Comput. Theory Eng. (ICACTE)*, Aug. 2010, pp. V5-376–V5-380.
- [14] T. Qiu, R. Qiao, M. Han, A. K. Sangaiah, and I. Lee, "A lifetime-enhanced data collecting scheme for the Internet of Things," *IEEE Commun. Mag.*, vol. 55, no. 11, pp. 132–137, Nov. 2017.
- [15] I. F. Akyildiz, T. Melodia, and K. R. Chowdhury, "A survey on wireless multimedia sensor networks," *Comput. Netw.*, vol. 51, no. 4, pp. 921–960, Mar. 2007.
- [16] D. J. Baker and A. Ephremides, "The architectural organization of a mobile radio network via a distributed algorithm," *IEEE Trans. Commun.*, vol. COM-29, no. 11, pp. 1694–1701, Nov. 1981.
- [17] S. Liu, L. Feng, Y. Liu, H. Qiao, J. Wu, and W. Wang, "Manifold warp segmentation of human action," *IEEE Trans. Neural Netw. Learn. Syst.*, to be published, doi: [10.1109/TNNLS.2017.2672971](https://doi.org/10.1109/TNNLS.2017.2672971).
- [18] F. Zhou, F. De la Torre, and J. K. Hodgins, "Aligned Cluster analysis for temporal segmentation of human motion," in *Proc. IEEE Int. Conf. Autom. Face Gesture Recognit.*, Sep. 2008, pp. 1–7.
- [19] F. Zhou, F. D. L. Torre, and J. K. Hodgins, "Hierarchical aligned cluster analysis for temporal clustering of human motion," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 3, pp. 582–596, Mar. 2013.
- [20] A. Vögele, B. Krüger, and R. Klein, "Efficient unsupervised temporal segmentation of human motion," in *Proc. ACM SCA*, 2014, pp. 167–176.
- [21] J. F.-S. Lin, V. Joukov, and D. Kulic, "Human motion segmentation by data point classification," in *Proc. 36th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Aug. 2014, pp. 9–13.
- [22] D. Gong, G. Medioni, and X. Zhao, "Structured time series analysis for human action segmentation and recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 7, pp. 1414–1427, Jul. 2014.
- [23] J. Barbič, A. Safonova, J.-Y. Pan, C. Faloutsos, J. K. Hodgins, and N. S. Pollard, "Segmenting motion capture data into distinct behaviors," in *Proc. Graph. Interface*, 2004, pp. 185–194.
- [24] S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," *Chemometrics Intell. Lab. Syst.*, vol. 2, nos. 1–3, pp. 37–52, 1987.
- [25] X. He and P. Niyogi, "Locality preserving projections," in *Proc. Adv. Neural Inf. Process. Syst.*, 2004, pp. 153–160.
- [26] M. E. Tipping and C. M. Bishop, "Probabilistic principal component analysis," *J. Roy. Statist. Soc. B*, vol. 61, no. 3, pp. 611–622, 1999.
- [27] H. Qiao, P. Zhang, D. Wang, and B. Zhang, "An explicit nonlinear mapping for manifold learning," *IEEE Trans. Cybern.*, vol. 43, no. 1, pp. 51–63, Feb. 2013.
- [28] H. Qiao, P. Zhang, B. Zhang, and S. Zheng, "Learning an intrinsic-variable preserving manifold for dynamic visual tracking," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 40, no. 3, pp. 868–880, Jun. 2010.
- [29] Z. Yu, L. Yuan, W. Luo, L. Feng, and G. Lv, "Spatio-temporal constrained human trajectory generation from the pir motion detector sensor network data: A geometric algebra approach," *Sensors*, vol. 16, no. 1, p. 43, 2015.
- [30] P.-Z. Chen, J. Li, M. Luo, and N.-H. Zhu, "Real-time human motion capture driven by a wireless sensor network," *Int. J. Comput. Games Technol.*, vol. 8, no. 4, pp. 1–14, 2015.
- [31] K. Jarrett, K. Kavukcuoglu, M. Ranzato, and Y. LeCun, "What is the best multi-stage architecture for object recognition?" in *Proc. IEEE 12th Int. Conf. Comput. Vis.*, Sep./Oct. 2009, pp. 2146–2153.
- [32] N. Pinto, D. D. Cox, and J. J. DiCarlo, "Why is real-world visual object recognition hard?" *PLoS Comput. Biol.*, vol. 4, no. 1, p. e27, 2008.
- [33] Y. Gong, S. Lazebnik, A. Gordo, and F. Perronnin, "Iterative quantization: A procrustean approach to learning binary codes for large-scale image retrieval," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 12, pp. 2916–2929, Dec. 2013.
- [34] S. Liu, M. Sun, X. Huang, W. Wang, and F. Wang, "Feature fusion using extended jaccard graph and word embedding for robot," *Assembly Autom.*, vol. 37, no. 3, pp. 278–284, 2017.
- [35] H. Qiao, C. Li, P. Yin, W. Wu, and Z.-Y. Liu, "Human-inspired motion model of upper-limb with fast response and learning ability—A promising direction for robot system and control," *Assembly Autom.*, vol. 36, no. 1, pp. 97–107, 2016.
- [36] J.-F. Yu, W.-B. Tang, Y. Li, and J. Zhang, "Dimensional variation propagation modeling and analysis for single-station assembly based on multiple constraints graph," *Assembly Autom.*, vol. 36, no. 3, pp. 308–317, 2016.
- [37] S. Niwattanakul, J. Singthongchai, E. Naenudorn, and S. Wanapu, "Using of Jaccard coefficient for keywords similarity," in *Proc. Int. Multiconf. Eng. Comput. Sci.*, vol. 1, no. 6, pp. 1–5, 2013.
- [38] K. Q. Weinberger, F. Sha, and L. K. Saul, "Learning a kernel matrix for nonlinear dimensionality reduction," in *Proc. 21st Int. Conf. Mach. Learn.*, 2004, p. 106.
- [39] M. Datar, N. Immorlica, P. Indyk, and V. S. Mirrokni, "Locality-sensitive hashing scheme based on p-stable distributions," in *Proc. 20th Annu. Symp. Comput. Geometry*, 2004, pp. 253–262.
- [40] Y. Weiss, A. Torralba, and R. Fergus, "Spectral hashing," in *Proc. Adv. Neural Inf. Process. Syst.*, 2009, pp. 1753–1760.
- [41] G. Lin, C. Shen, Q. Shi, A. van den Hengel, and D. Suter, "Fast supervised hashing with decision trees for high-dimensional data," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 1963–1970.
- [42] Y. Liu, L. Feng, S. Liu, and M. Sun, "Global similarity preserving hashing," *Soft Computing*. Berlin, Germany: Springer, 2017, pp. 1–16, doi: [10.1007/s00500-017-2683-7](https://doi.org/10.1007/s00500-017-2683-7).
- [43] Z. Babar, P. Botsinis, D. Alanis, S. X. Ng, and L. Hanzo, "The road from classical to quantum codes: A hashing bound approaching design procedure," *IEEE Access*, vol. 3, pp. 146–176, Mar. 2015.
- [44] J. Qi, X. Hu, Y. Ma, and Y. Sun, "A hybrid security and compressive sensing-based sensor data gathering scheme," *IEEE Access*, vol. 3, pp. 718–724, 2015.
- [45] K. Christidis and M. Devetsikiotis, "Blockchains and smart contracts for the Internet of Things," *IEEE Access*, vol. 4, pp. 2292–2303, 2016.
- [46] S. Liu, L. Feng, and H. Qiao, "Scatter balance: An angle-based supervised dimensionality reduction," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 2, pp. 277–289, Feb. 2015.
- [47] H. Wang, L. Feng, J. Zhang, and Y. Liu, "Semantic discriminative metric learning for image similarity measurement," *IEEE Trans. Multimedia*, vol. 18, no. 8, pp. 1579–1589, Aug. 2016.
- [48] L. De Lathauwer, B. De Moor, and J. Vandewalle, "A multilinear singular value decomposition," *SIAM J. Matrix Anal. Appl.*, vol. 21, no. 4, pp. 1253–1278, 2000.
- [49] G. Hripcsak and A. S. Rothschild, "Agreement, the F-measure, and reliability in information retrieval," *J. Amer. Med. Informat. Assoc.*, vol. 12, no. 3, pp. 296–298, 2005.
- [50] J. Barbič, M. da Silva, and J. Popović, "Deformable object animation using reduced optimal control," *ACM Trans. Graph.*, vol. 28, no. 3, 2009, Art. no. 53.
- [51] J. K. T. Tang and H. Leung, "Retrieval of logically relevant 3D human motions by adaptive feature selection with graded relevance feedback," *Pattern Recognit. Lett.*, vol. 33, no. 4, pp. 420–430, 2012.



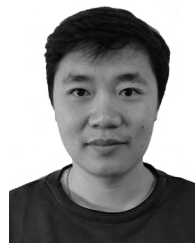
YANG LIU received the B.S. degree from the School of Computer Sciences, Dalian University of Technology, China, in 2012, where he is currently pursuing the Ph.D. degree with the School of Computer Science and Technology. His research interests include video analysis, Internet of Things, and machine learning.



SHENGLAN LIU received the Ph.D. degree from the School of Computer Science and Technology, Dalian University of Technology, China, where he is currently pursuing the Ph.D. degree. He has authored papers for the *IEEE TRANSACTIONS ON NEURAL NETWORKS* and the *Neurcomputing*. His current research interests include pattern recognition, image retrieval, and machine learning.



LIN FENG received the B.S. degree in electronic technology, the M.S. degree in power engineering, and the Ph.D. degree in mechanical design and theory from the Dalian University of Technology, China, in 1992, 1995, and 2004, respectively. He is currently a Professor and a Doctoral Supervisor with the School of Innovation Experiment, Dalian University of Technology. His research interests include intelligent image processing, robotics, data mining, and embedded systems.



MUXIN SUN received the B.S. and M.S. degrees from the School of Computer Sciences, Dalian University of Technology, China, in 2014 and 2017, respectively. His research interests include video analysis and machine learning.

...