Abstract—Multi-sensor fusion strategies have been widely applied in Human Activity Recognition (HAR) in Body Sensor Networks (BSNs). However, the sensory data collected by BSNs systems are often uncertain or even incomplete. Thus, designing a robust and intelligent sensor fusion strategy is necessary for high-quality activity recognition. In this paper, Dezert-Smarandache Theory (DSmT) is used to develop a novel sensor fusion strategy for HAR in BSNs, which can effectively improve the accuracy of recognition. Specifically, in the training stage, the Kernel Density Estimation (KDE) based models are first built and then precisely selected for each specific activity according to the proposed discriminative functions. After that, a structure of Basic Belief Assignment (BBA) can be constructed, using the relationship between the test data of unknown class and the selected KDE models of all considered types of activities. In order to deal with the conflict between the obtained BBAs, Proportional Conflict Redistribution-6 (PCR6) is applied to fuse the acquired BBAs. Moreover, the missing data of the involved sensors are addressed as ignorance in the framework of the DSmT without manual interpolation or intervention. Experimental studies on two real-world activity recognition datasets (The OPPORTUNITY dataset; Daily and Sports Activity Dataset (DSAD)) were conducted, and the results showed the superiority of our proposed method over some state-of-the-art approaches proposed in the literature.

Index Terms—HAR, Multi-sensor fusion, Belief function theory, KDE, DSmT.

I. INTRODUCTION

This work was supported in part by the National Natural Science Foundation of China under Grant 61573097 and Grant 91748106, in part by the Key Laboratory of Integrated Automation of Process Industry under Grant PALN201704, in part by the Advanced Research Project of the 13th Five-Year Plan under Grant 31511040301, in part by Science and Technology on Information System Engineering Laboratory, No:05201905, in part by the Fundamental Research Funds for the Central Universities under Grant 3208008401, in part by the Qing Lan Project and Six Major Top-Talent Plan, and in part by the Priority Academic Program Development of Jiangsu Higher Education Institutions. (Corresponding author: Xinde Li.)

Yilin Dong is with Key Laboratory of Measurement and Control of CSE, School of Automation, Southeast University, Nanjing, 210096, China (Email:dy@seu.edu.cn).

Xinde Li is with Key Laboratory of Measurement and Control of CSE, School of Automation and also with the School of Cyber Science and Engineering, Southeast University, Nanjing, 210096, China (Email:xindeli@seu.edu.cn).

Jean Dezert is with The French Aerospace Lab, ONERA, DTIS, F-91123 Palaiseau, France (Email:jean.dezert@onera.fr).

M. O. Khyam is with School of Electrical Engineering, Central Queensland University, Melbourne, Australia (Email:m.khyam@cqu.edu.au).

Md. Noor-A-Rahim is with School of Computer Science and IT, University College Cork, Ireland (Email:m.rahim@cs.ucc.ie).

S. S. Ge is with the Social Robotics Laboratory, Department of Electrical and Computer Engineering, Interactive Digital Media Institute, National University of Singapore, Singapore 117576 and also with Institute for Future (IFF), Qingdao University, Qingdao 266071, China (Email:smge@nus.edu.sg).

HUMAN Activity Recognition (HAR) has spawned intense researches in the past decades and continues to be an active research area [1], [2], [3], [4]. These HAR systems have enabled several practical applications, such as health monitoring [5], physical activity [6] and gesture detection. Recently, multi-sensor fusion for activity recognition is playing an increasing role in HAR field and many strategies have been proposed (see [7] for more references). Generally speaking, multi-sensor fusion strategies can be mainly categorized into three level categories depending on the abstraction level used for data processing: data-fusion level [8], feature-fusion level [9] and decision-level fusion [10]. Among all these three fusion levels, decision-level fusion output is a unique decision obtained from local decision of multiple (homogeneous or heterogeneous) sensors. The fusion in this level has many advantages: communication bandwidth saving, allowing the combination of the heterogeneous sensors. In this paper, the main topic thus focus on decision-level fusion area. Two most common used approaches for this level of fusion are majority voting [11] and naive bayes [12]. However, complex sensory data, especially when these data are uncertain or even incomplete, make these two methods unsuitable for HAR. Two classical scenarios are described as follows:

1) Uncertain sensory data in HAR problem. In order to intuitively discuss the uncertainty of sensory data, one of the involved sensor in UCI OPPORTUNITY dataset [13], [14] was randomly selected and parts of the original data of three activities derived from the chosen sensor were drawn in Fig.1(a). As we can see from Fig.1(a), some objects that are very close can sometimes truly originate from different classes. Such objects are really difficult to classify correctly into a particular class using the given information. In this case, we
call this data uncertain when it can belong to different specific classes with probability mass assignments to estimate;

2) Incomplete sensory data in HAR problem. Missing data frequently occur during the measurement of wearable-based activity recognition. As we can see in Fig.1(b), sensory data with incomplete pattern occupy an important proportion which cannot be easily neglected in OPPORTUNITY dataset. The traditional ways to cope with these feature vectors, which include missing data, are to interpolate or delete the whole vector. However, interpolation or deletion is not the wise choice which may bring noise and information loss to the recognition system.

The aforementioned discussions motivate our study, where HAR in Body Sensor Networks (BSNs) is implemented based on belief function theory [15]. Belief function allows to model uncertainty and to fuse Basic Belief Assignments (BBAs) built from sensors’ measurements. Within this theory, information fusion relies on the use of a combination rule allowing the pieces of evidences (drawn from sensor readings) expressed in a common frame of discernment to be combined. Among all available combination rules, Dempster’s rule proposed by Shafer in Dempster-Shafer theory is the most well-known rule still used in many applications even if it remains very controversial. Recently, Chen et al. [16] proposed a new method based on Dempster-Shafer theory to improve human action recognition by using the fusion of depth camera and inertial sensors. Although the recognition results mentioned in [16] is good, two key issues are ignored by authors: 1) In Dempster-Shafer theory, there exists an assumption that hypotheses considered should be exclusive. However, in HAR, activities to be identified often fail to satisfy the characteristics of mutual exclusion. For example, the intersection between “Walking” and “Running” can be defined as “Standing” or intermediate transition state “Walking to Running” [17]; 2) Dempster’s rule cannot solve high conflict issues and even very low conflict issues in specific cases, which have been widely discussed in [18], [19].

To solve those mentioned drawbacks in Dempster-Shafer theory, Dezert and Smarandache proposed Dezert-Smarandache Theory (DSmT) [18] to solve multi-sensor fusion problems, with more reasonable assumptions and better combination rules, which is more appropriate to handle HAR problems. In this paper, a new use of DSmT is proposed to solve HAR issues thanks to a novel decision-level fusion strategy based on DSmT. Such DSmT-based HAR can be used for online activity recognition system because of its higher recognition accuracy and lower recognition delay, which can meet the required response speed in real-time recognition systems (less than 200ms)[2]. Specifically, the main contributions of this work are summarized as follows:

- A novel DSmT-based fusion strategy for HAR in BSNs is proposed;
- Kernel Density Estimation (KDE) models are constructed based on the sensor readings, and those selected KDE models of all considered classes are applied to calculate BBAs in DSmT;
- The missing data in original sensor readings are also modeled by vacuous BBA (i.e. the total ignorance source of evidence) in DSmT without any manual interpolation;
- The efficiency of our fusion system with two activities recognition open datasets is demonstrated.

This paper is organized as follows: Section II provides an inventory of the basic concepts of DSmT. Section III provides a description of the new proposed fusion method. Section IV includes the experimental results and discussions. The final section V contains a brief conclusion.

II. BASICS OF DSMT

In DSmT framework, the BBAs are defined on the so-called hyper-power set (or Dedekind’s lattice) denoted $D^\Theta = (\Theta, \cup, \cap)$ whose cardinalities follows Dedekind’s numbers sequence, see [18], Vol.1 for details and examples. A (generalized) BBA, called a mass function, $m(\cdot)$ is defined by the mapping: $D^\Theta \mapsto [0,1]$, verifying $m(\emptyset) = 0$ and $\sum_{A \in D^\Theta} m(A) = 1$.

To palliate the drawbacks of Demspter’s rule, Martin et.al [20] proposed a very interesting combination rule: PCR6. Due to its good performance, it is widely applied in recent applications. We recall that the PCR6 formula for the combination of two BBAs coincides with PCR5 formula originally developed by Smarandache and Dezert in [18]. The combination of two BBAs $m_1(\cdot)$ and $m_2(\cdot)$ by the PCR5 rule is given as follows: for $m_{PCR5}(B) = 0$ and $\forall A \in D^\Theta$

$$m_{PCR6}(A) = m_{PCR5}(A) = m_{12}(A) = \sum_{B \in D^\Theta \setminus \{A\} | A \cap B = \emptyset} \left[ \frac{m_1(A)^2 m_2(B) + m_2(A)^2 m_1(B)}{m_2(A) + m_1(B)} \right],$$

(1)

where $m_{12}(A) = \sum_{B \subseteq C \subseteq D^\Theta | B \cup C = A} m_1(B) m_2(C)$. The combinations of more than two BBAs altogether with PCR5 and with PCR6 fusion rule in general provide different results. The choice of PCR6 with respect to PCR5 was justified at first by Martin and Osswald in [20] from a specific application, and then theoretically by Smarandache and Dezert in [21]. The general formula of PCR6 for combining more than two BBAs was given in details in [20] with examples.

III. DSmT-BASED FUSION STRATEGY FOR HAR IN BSNs

A. The Flow Chart of Our Proposed Method

Before entering in the detailed presentation of our DSmT-based fusion strategy, we briefly introduce it through the flowchart of Fig.2 for convenience. Specifically, in the training stage, multiple KDE models are derived from the raw sensor readings so as to build the model pool. Then, the representative model is selected for a particular activity based on our proposed discriminative functions. After that, when the test sample comes, the corresponding BBA is calculated through each activity representative model. Finally, these BBAs are combined with PCR6 rule, from which we make the final decisions.
B. Mathematical Definitions of Daily Activities in DSmT

The goal of our work is to recognize human daily activities thanks to DSmT-based framework. Thus, the basic mathematical definitions of the interested activities need to be given. We assume that the finite frame of discernment considered in our activity recognition problem is $\Theta = \{\theta_1, \theta_2, \ldots, \theta_n\}$. The corresponding hyper-power set of $\Theta$ is denoted $D^\Theta$. Singletons in $D^\Theta$ are used to represent the simple daily activity such as $\theta_1 \triangleq$ Standing, $\theta_2 \triangleq$ Sitting, $\theta_3 \triangleq$ Lying and so on. Disjunctive focal elements in $D^\Theta$ represent the coarse-grained activities. For example, $\theta_1 \cup \theta_2 \cup \theta_3 \triangleq$ Static Activity. Also, if $\theta_4 \triangleq$ Walking, $\theta_5 \triangleq$ Running, then $\theta_4 \cup \theta_5$ is regarded as Dynamic Activity. Following the definition line of disjunctive focal elements, $\theta_1 \cup \theta_2 \cdots \cup \theta_n$ represents the whole unknown activity. Besides, the conjunctive focal elements in $D^\Theta$ can be used to stand for the transition activity like $\theta_1 \cap \theta_2 \triangleq$ Standing to Sitting or $\theta_1 \cap \theta_2 \triangleq$ Sitting to Standing because $\theta_1 \cap \theta_2 = \theta_3 \cap \theta_1$ and $\theta_2 \cap \theta_3 \triangleq$ Sitting to Lying or Lying to Sitting. In this paper, we only consider a restricted hyper-power set, which is denoted as $D^\Theta_{\text{restricted}} = \{\theta_1, \theta_2, \ldots, \theta_n, \theta_1 \cup \theta_2 \cdots \cup \theta_n\}$. In $D^\Theta_{\text{restricted}}$, only two types of focal elements exist: one is the singleton, which represents the simple activity and another is $\theta_1 \cup \theta_2 \cdots \cup \theta_n$, which represents the unknown activity. More complicated situations involving less restricted hyper-power sets will be discussed in our future work.

C. Training Model Stage

In the training stage, the KDE model is employed to fit the sensor readings. The most suitable KDE model to distinguish a certain activity is then selected to be regarded as the specific activity representative model. Among the process of this training stage, two main steps are involved:

1) Construction of KDE Models: We assume that there are $M$ kinds of activities that need to be classified and the original dataset collected from the wearable sensors are denoted as $x_{ij}$, $i = 1, \cdots, M$ and $j = 1, \cdots, N$. Here, $M$ represents the types of activities to be classified and $N$ is the number of sensors. Thus, based on the Eq.(2), the KDE model of the specific activity is derived from the sensor readings by

$$f_{ij}(x_{ij}) = \frac{1}{Q} \sum_{q=1}^{Q} K_h(x - x_{ij}^q) = \frac{1}{Qh} \sum_{q=1}^{Q} K\left(\frac{x - x_{ij}^q}{h}\right),$$

where $f(x_{ij})$ is the KDE model of $x_{ij}$ which represents the model of the $j$ sensor for the $i$ activity; $K(\cdot)$ is the kernel function which can be ‘normal’, ‘epanechnikov’, ‘box’ and ‘triangle’; $h$ is the smoothing parameter (the bandwidth) of the KDE model. In this paper, the value of $h$ is the adaptative bandwidth selected by the method presented in [22]; The parameter $Q$ is the dimension of $x_{ij}$.

2) Selection of the Best Discriminative KDE Model for the Specific Activity: As we can see from Eq.(2), each activity can have $N$ KDE models and we need to select the most discriminative KDE model in order to reduce the computational complexity and the interference model. Once the unique KDE model for each activity is selected, one can easily determine a specific sensor to identify activity because there is one-to-one correspondence between the KDE models and the wearable sensors. We propose two novel discriminant evaluation functions as follows:

Definition 1: For the specific activity $\theta_s$, $s \in \{1, \cdots, M\}$, the value of Sum of Statistical Difference (SSD) of the $j$, $j = 1, \cdots, N$ the KDE model is calculated as follows:

$$SSD_{\theta_s}(j) = [\Psi(f_{\theta_s,j}) - \Psi(f_{ij})] + \cdots + [\Psi(f_{\theta_s,j}) - \Psi(f_{\theta_M,j})] = (M - 1) \cdot \Psi(f_{\theta_s,j}) - \sum_{i=1, i \neq s}^{M} \Psi(f_{\theta_i,j}).$$

In Eq.(3), $\theta_s$ is one of the specific activity among the $M$ considered activities; $j$ is the sensor readings of the $j$ sensor; $\Psi(\cdot)$ calculates the statistical characteristic value of the derived distribution of the KDE model $f_{\theta_s,j}$. In this paper, $\Psi(\cdot) = Mean(\cdot)$, that is the average value of sensor readings.

The principle of selecting KDE model based on SSD is quite simple: for the specific activity $\theta_s$, if the SSD value of the $j$, $j = 1, \cdots, N$ sensor is large, it means that...
this j KDE model of \( \theta_s \) has a better discriminative ability. Here, a simple illustrative example was extracted from the OPPORTUNITY dataset in Section IV to show the principle of SSD. As we can see in Fig.3, for the specific activity \( \theta_1 \), the value of \( SSD_{\theta_1}(g_1) = \text{Mean}(f_{\theta_1 g_1}) - \text{Mean}(f_{\theta_2 g_1}) \) (Fig.3(a)) is larger that \( SSD_{\theta_1}(g_2) = \text{Mean}(f_{\theta_1 g_2}) - \text{Mean}(f_{\theta_2 g_2}) \) (Fig.3(b)). Here, \( g_1 \) and \( g_2 \) represent the \( g_1 \) sensor and the \( g_2 \) sensor. It can be clearly seen in Fig.3 that KDE model \( (f_{\theta_1 g_1}) \) has the higher discriminative ability than KDE model \( (f_{\theta_1 g_2}) \) for activity \( \theta_1 \).

In order to measure the distances between probability density functions of each pair of KDEs models, another well-known choice for such measurement is Kullback-Leibler (KL) divergence defined by, see [23]:

\[
Div_{KL}(f_{g_1}||f_{g_2}) = \sum_i f_{g_1}(i) \log \frac{f_{g_1}(i)}{f_{g_2}(i)},
\]

(4)

Here \( f_{g_1} \) and \( f_{g_2} \) are two discrete probability density functions. Similar to \( Div_{KL} \), another well-known divergence is Jensen-Shannon (JS) divergence defined by:

\[
Div_{JS}(f_{g_1}||f_{g_2}) = \frac{1}{2} [Div_{KL}(f_{g_1}||f_{g_2}) + Div_{KL}(f_{g_2}||f_{g_1})].
\]

(5)

Based on Eq.(4) and Eq.(5), another discriminative evaluation function is given to measure the discriminative ability between different KDE models, which is named as Sum of Divergence Difference (SDD).

**Definition 2:** For the specific activity \( \theta_s, s \in \{1, \cdots, M\} \), the value of SDD of the \( j, j = 1, \cdots, N \) KDE model is calculated as follows:

\[
SDD_{\theta_s}(j) = \sum_{i \neq s}^{M-1} \Upsilon(f_{\theta_{s,j}, f_{\theta_{i,j}}}).
\]

(6)

In Eq.(6), \( \theta_s \) is the specific class of daily activity; \( \Upsilon(\cdot) \) represents the divergence function. In this paper, \( \Upsilon(\cdot) \) is defined as \( KL \) (Eq.(4)) or \( JS \) (Eq.(5)). It is worth noting that in order to make the statements more clear in the following sections, we will directly use the \( \text{Mean}(\cdot) \) to represent that SSD criterion is applied for selecting KDE models in the process of activity recognition. Similarly, \( Div_{KL}(\cdot) \) or \( Div_{JS}(\cdot) \) mean that SDD is applied and \( Div_{KL}(f_{g_1}||f_{g_2}) \) or \( Div_{JS}(f_{g_1}||f_{g_2}) \) is used in SDD criterion to measure the difference between two distributions. For each activity \( \theta_1, \theta_2, \cdots, \theta_M \), the best discriminative M KDE models \( f_{\theta_i, i = 1, \cdots, M} \) can be selected and denoted as follows:

\[
\begin{bmatrix}
  f_{\theta_{1g_1}} & f_{\theta_{2g_1}} & \cdots & f_{\theta_{Mg_1}} \\
  f_{\theta_{1g_2}} & f_{\theta_{2g_2}} & \cdots & f_{\theta_{Mg_2}} \\
  \vdots & \vdots & \ddots & \vdots \\
  f_{\theta_{1g_M}} & f_{\theta_{2g_M}} & \cdots & f_{\theta_{Mg_M}}
\end{bmatrix}
\]

(7)

and \( g_1, g_2, \cdots, g_M \in [1, N] \). Each of \( g_i, i \in \{1, \cdots, M\} \) represents the selected wearable sensor number.

D. Testing Stage

When the test sample becomes available, the corresponding BBA is calculated through each KDE model of each activity. Finally, we combine all related BBAs with PCR6 rule and we make the final decisions from the combined BBAs.

1) **BBAs Calculation:** In this paper, the considered frame of discernment is \( \Theta = \{\theta_1, \theta_2, \cdots, \theta_M\} \). Each focal element in \( \Theta \) represents one kind of activity and here we just consider a simplified \( D_{\text{restricted}}^{\Theta} = \{\theta_1, \theta_2, \cdots, \theta_M\} \). We consider a testing vector \( \mathbf{x} \) with unknown class and we want to identify the label of \( \mathbf{x} \) corresponding to the activity it belongs to. Next, we use the following equations to calculate the BBAs \( (m_1(\cdot), m_2(\cdot), \cdots, m_M(\cdot)) \):

\[
\begin{align*}
m_1(\theta_1) &= f_{\theta_1 g_1}(x(g_1)), \cdots, m_1(\theta_M) = f_{\theta_M g_1}(x(g_1)); \\
m_2(\theta_1) &= f_{\theta_1 g_2}(x(g_2)), \cdots, m_2(\theta_M) = f_{\theta_M g_2}(x(g_2)); \\
\vdots \\
m_M(\theta_1) &= f_{\theta_1 g_M}(x(g_M)), \cdots, m_M(\theta_M) = f_{\theta_M g_M}(x(g_M)).
\end{align*}
\]

It is worth noting that when the value of one feature is missing, we directly assign “1” to \( m(\theta_1 \cup \theta_2 \cup \cdots \cup \theta_M) \) which means in this case, we cannot obtain the valuable
decision information. Besides, in order to make sure that the derived BBAs satisfy the normalization condition, the following normalization applies:

- If \( m_i(\theta_1) + \cdots + m_i(\theta_M) \leq 1 \), then \( m_i(\theta_1 \cup \cdots \cup \theta_M) = 1 - (m_i(\theta_1) + \cdots + m_i(\theta_M)) \);
- If \( m_i(\theta_1) + \cdots + m_i(\theta_M) > 1 \), then \( m_i(\theta_k) = \frac{m_i(\theta_k)}{\sum_{k=1}^{M} m_i(\theta_j)} \) for \( k = 1, \ldots, M \), and \( m_i(\theta_1 \cup \cdots \cup \theta_M) = 0 \).

2) Global Fusion with PCR6 and Decision Making: After obtaining the \( M \) BBAs, the PCR6 fusion rule is used to fuse all these BBAs which is denoted symbolically by

\[
m_{\text{fusion}} = \text{PCR6}(m_1, m_2, \ldots, m_M). \tag{8}
\]

Then the final decision of the predicted class of \( x \) can be made as \( \hat{\theta}^*_i = \text{argmax}_\theta m_{\text{fusion}}(\theta_i) \), where \( \theta_i \) is a focal element of the \( D'M \) restricted based on the max of belief mass.

The DSmT-Based Activity Recognition technique is described in Algorithm 1 for convenience.

**Algorithm 1: DSmT-Based HAR**

**Input:** Sequential original data \( x_{ij}, i = 1, \ldots, N, j = 1, \ldots, M \).

**Output:** The Predicted Class of Unknown data \( x^* \).

1. **Initialize:** Cross Validation \((x_{ij}) \rightarrow x_{\text{training}}, x_{\text{testing}}\);
2. **Training Stage:**
   3. for \( i = 1, \ldots, M \) do
      4. for \( j = 1, \ldots, N \) do
         5. \( \phi(x_{ij}) = \frac{1}{Q^k} \sum_{q=1}^{Q} K(\frac{x_{ij}}{\hat{\kappa}}) \);
      6. end
   7. end
8. for \( i = 1, \ldots, M \) do
   9. for \( j = 1, \ldots, N \) do
      10. \( SDD_{\theta_i}(j) = (M - 1) \cdot \Psi(f_{ij}) = \sum_{k=1}^{M} \Psi(f_{ij}) \); or
      11. \( SDD_{\theta_i}(j) = \sum_{k=1}^{M} T(f_{ij}, f_{ij}) \);
   12. end
   13. \( g_i = \max(SDD_{\theta_i}) \) or \( g_i = \max(SDD_{\theta_i}) \);
14. end
15. \( f_{\text{matrix}} = \phi_{\theta_1}, \ldots, f_{\theta_M, \theta_1}, \ldots, f_{\theta_M, \theta_2}, \ldots, f_{\theta_M, \theta_M} \);
16. **Testing Stage:**
17. \( D_{\text{restricted}}^{\theta} = \{\theta_1, \theta_2, \ldots, \theta_M, \theta_1 \cup \theta_2 \cup \cdots \cup \theta_M\} \);
18. for \( i = 1, \ldots, M \) do
19. \( m_i(\theta_j) = f_{\theta_j}(x^*(g_i)) \), \( m_i(\theta_M) = f_{\theta_M}(x^*(g_i)) \);
20. end
21. if \( m_i(\theta_1) + \cdots + m_i(\theta_M) \leq 1 \) then
22. \( m_i(\theta_1 \cup \cdots \cup \theta_M) = m_i(\theta_1) + \cdots + m_i(\theta_M) \);
23. else if \( m_i(\theta_1) + \cdots + m_i(\theta_M) > 1 \) then
24. Normalization of BBAs \( m_i(\theta_1), \ldots, m_i(\theta_M) \);
25. end
26. **Fusion Step:** \( m_{\text{fusion}} =\text{PCR6}(m_1, \cdots, m_M) \);
27. **Decision Step:** Take as decision the maximum of belief mass of focal elements \( \hat{\theta}^*_i = \text{argmax}_\theta m_{\text{fusion}}(\theta_i) \);
28. **final:**
29. **return** Predicted Class of \( x^* \);

IV. PERFORMANCE EVALUATION

A. Datasets

The performance of the proposed DSmT-Based HAR was evaluated on the following two open HAR datasets. The first one is **UCI OPPORTUNITY dataset** [13], [14]. The details of this dataset can be found in OPPORTUNITY UCI dataset1. Three basic activities were classified: Walking, Sitting and Lying; The other one is **UCI DSAD**2. The details of the DSAD can be found in [24]. In this dataset, five common daily activities including Sitting, Standing, Lying, Walking and Running were classified to prove the effectiveness of our proposed method.

B. Measures of Performance

As measures of the performance of our activity recognition system, the classical **Accuracy**, **Precision**, **Recall**, and **F1-score** [7] have been used. They are defined by

\[
\text{Accuracy} = \frac{1}{n} \sum_{k=1}^{n} \frac{TP_k + TN_k}{TP_k + FN_k + FP_k + TN_k} . \tag{9}
\]

\[
\text{Precision} = \frac{1}{n} \sum_{k=1}^{n} \frac{TP_k}{TP_k + FP_k} . \tag{10}
\]

\[
\text{Recall} = \frac{1}{n} \sum_{k=1}^{n} \frac{TP_k}{TP_k + FN_k} . \tag{11}
\]

\[
\text{F1-Score} = \frac{1}{n} \sum_{k=1}^{n} \frac{2 \cdot \text{precision}_k \cdot \text{recall}_k}{\text{precision}_k + \text{recall}_k} . \tag{12}
\]

where \( k \) denotes class index and \( n \) is the number of classes. True Negatives (\( TP_k \)): the number of correctly recognized class examples; True Negatives (\( TN_k \)): the number of correctly recognized examples that do not belong to the class; False Positives (\( FP_k \)): examples that were either incorrectly assigned to the class; False Negatives (\( FN_k \)): not recognized as class examples.

C. Results on UCI OPPORTUNITY dataset

1) Effectiveness of the Selection of \( \Psi(\cdot) \) and \( \Upsilon(\cdot) \) in Eq.(3) and Eq.(6): The selections of \( \Psi(\cdot) \) in SSD and \( \Upsilon(\cdot) \) in SDD were quite crucial to the representative KDE models for all involved activities. Thus, the relevant comparisons about the recognition rates were given in Fig.4 when \( \Psi(\cdot) \) and \( \Upsilon(\cdot) \) were set to (1) \( \Psi(\cdot) = \text{Mean}(\cdot) \), (2) \( \Psi(\cdot) = \text{Div}_{KL}(\cdot) \), (3) \( \Upsilon(\cdot) = \text{Div}_{JS}(\cdot) \), respectively. As we can see in Fig.4, our proposed method based on these three discriminative functions3 distinguished three mentioned activities in Opportunity dataset (four subjects) very well, which indirectly proved the effectiveness of Mean(\cdot), Div_{KL}(\cdot), Div_{JS} in measuring the difference between the distributions of activities. Besides, all the three generated models had the highest recognition accuracy on Subject 1. However, the sensors selected by each function were quite different, and the corresponding involved sensors were listed in Table I. It can be found that the sensitivity of sensors to different daily activities varied, and was influenced by their locations of deployment. Sensors


3As we introduced in Definition 1 and Definition 2, \( \Psi(\cdot) \) means that SSD (Eq.(3)) is used to choose the best KDE models and \( \Upsilon(\cdot) \) means that SDD (Eq.(6)) is applied in our activity recognition model.
TABLE I: The Selected Sensors in OPPORTUNITY dataset Based on $\Psi(\cdot)=\text{Mean}(\cdot), \text{Div}_K\text{L}(\cdot), \text{Div}_J\text{S}(\cdot)$.

<table>
<thead>
<tr>
<th>Subject</th>
<th>$SSD: \Psi(\cdot)=\text{Mean}(\cdot)$</th>
<th>$SDD: \Upsilon_1(\cdot)=\text{Div}_K\text{L}(\cdot)$</th>
<th>$SDD: \Upsilon_2(\cdot)=\text{Div}_J\text{S}(\cdot)$</th>
</tr>
</thead>
</table>

*According to [26], each triaxial (x,y,z) sensor unit has 3-degree of freedom. And in this Table, all the meanings of the involved sensors are: Left Lower Arm (LLA); Right Lower Arm (RLA); Right Knee (RKN); Left Wrist (LWR); Left Shone (LShone); Hips (HIP); Right Hand (RH); Right Shoe (RShoe); Left Upper Arm (LUA); Accelerator x axis (accX); Magnetic Z-axis (magZ). More details about OPPORTUNITY Dataset can be referred to [26].

located on the arm such as left lower arm, right hand, left wrist were more likely to identify “Walking” but sensors located on the back or shoes had higher recognition rates of “Lying” than other sensors. This directly indicates that it is not feasible or wise to rely on a single sensor deployed in a single location to identify various kinds of activities [25]. This is also our motivation to use multi-sensor fusion strategy based on DSmT to solve activity recognition problems.

![Fig. 4: Effectiveness of the selection of $SSD(\Psi(\cdot))$ and $SDD(\Upsilon(\cdot))$ in OPPORTUNITY dataset.](image)

2) Recognition Rate versus Training Percentage: In this experiment, we did modify the percentage of training set and investigated the relationship between the training percentage and the classification accuracy of our proposed method on OPPORTUNITY dataset. It is worth mentioning that the discriminative function chosen here was $SSD$ (Eq.(3)) and $\Psi(\cdot)=\text{Mean}(\cdot)$. Since our experiments were conducted based on ten-fold cross validation method, it is convenient for us to test the relationship between recognition rate and training percentage. According to the principle of ten-fold cross validation, the original datasets were first randomly divided into ten equal parts. And then, in the first experiment, we first treated 10% data as training dataset and the remaining 90% data were used as testing dataset; And in the second experiment, 20% datasets were used for training and the remaining 80% for testing, and so on, until the last experiment which we used 90% datasets for training and the last 10% datasets for testing. Besides, in order to further observe the performance of the proposed method, we divided the original data into 100 equal parts on the basis of one hundred cross-validation. And then one of the equal parts was randomly selected as the training datasets (1%) and the remaining (99%) were regarded as testing datasets. The average accuracy rates of all these ten experiments was shown in Fig.5, which showed that even if there were few training samples, the model proposed in this paper still gave higher recognition accuracy.

![Fig. 5: Classification accuracy vs. training percentage for the OPPORTUNITY dataset.](image)

3) Comparison Between Base Classifiers and Fused Classifiers in OPPORTUNITY dataset: In order to deeply analyze the relationship between base classifiers and fused classifier in our proposed model, the detailed comparisons were given in Fig.6. Based on the results presented in Fig.4, the discriminative function chosen here was $SSD$ (Eq.(3)) and $\Psi(\cdot)=\text{Mean}(\cdot)$. In Fig.6, the x-axis represents the KDE model corresponding to the selected sensor, the y-axis represents the number of correctly classified test samples, the value above each histogram represents the classification recognition rate corresponding to each KDE model, and the solid line at the top of the histogram represents the total number of test samples. As we can see from Fig.6: (1) the recognition accuracy of the fused model was significantly improved compared with that of the base classifier; (2) the performance of based classifiers were obviously different.
Among these mentioned base classifiers, RH-accY in subject 4 had the lowest rate: 56.9885% and LWR-accY also in subject 4 had the highest rate: 88.7390%. The main reason for the performance difference of the base classifiers is that we looked for the relative best KDE model for the specific activity based on our proposed SSD or SDD, not the absolute best KDE model for all activities. More concretely, in subject 1, the specific KDE model corresponding to LLA-accX had the best classification only for Walking; the specific KDE model corresponding to RLA-accX had the best classification only for Sitting and the specific KDE model corresponding to BackmagX had the best classification only for Lying. In this way, we could effectively guarantee the degree of diversity among base classifiers, which is really important for ensemble fusion [11].

4) Comparisons with State-of-the-art Approaches Based on Monte-Carlo Simulation: In this part, we further gave the confusion matrix (Fig.7) of the four subjects in OPPORTUNITY dataset based on our proposed method. It is worth noting that in the confusion matrix of subject 2-4, there existed a special label "UNKNOWN" which was quite different from the three mentioned activities: Walking, Sitting and Lying. This "UNKNOWN" label occurred in our DSmT-Based method because of the missing value in original sensor readings. When the current sensor reading was NULL or missing value, the maximum belief mass (’1’) was assigned to the focal element (Θ) which meant at current time, we really did not know the actual class. Modeling missing or NULL information is the feature of our proposed method in this paper, which is quite different from the traditional supplementation of NULL or missing information by interpolation. In this way, our proposed method can reduce the risk of misjudgment without guaranteeing any changes to the original data. Besides, we repeated 50 experiments and recorded the recognition rates of all four subjects in Table II. Among the mentioned classical approaches, the performance of k-Nearest Neighbours and Nearest Centroid Classifier were heavily affected by the number of ‘k’-closest samples and the centroid of each class. These two principles of classification were difficult to work very well when there existed uncertain data in HAR problem. Linear discriminative analysis and quadratic discriminant analysis based on the assumption that the features are normally distributed are obviously unsuitable in HAR problems. Extreme learning machine has been successfully applied for the task of HAR. And for extreme learning machine, sigmoid activation function was utilized and the number of hidden nodes was set to 100. However, due to the randomness of the algorithm, the results of extreme learning machine were unstable and had a wide variability. As we can observe in Table II, our method gave the highest activity recognition accuracy in subject-1, subject-2 and subject-4, and Ensemble-Extreme Learning Machine (Majority Voting) gave the highest recognition accuracy in subject 3. In addition to the comparison of classification accuracy, we also showed the testing time for each individual sample of our proposed method in Table II. Our method was running in MATLAB R2018b with a hardware of Intel Quad Core i5-4670 CPU at 3.4GHz and 16G RAM. As shown in Table II, our proposed method was significantly more efficient than other general listed methods. The low recognition delay of our method was mainly because in the testing phase, only the data of selected sensors in the testing sample participates in the BBA calculation. The low-recognition delay also showed its potential for the application in online activity recognition systems, because such real-time activity recognition often requires the predictions are updated 1-5 times/s [2].
TABLE II: Comparison with state-of-the-art results on UCI OPPORTUNITY dataset.

<table>
<thead>
<tr>
<th>Reported Methods</th>
<th>Accuracy</th>
<th>Average Computational Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Subject1</td>
<td>Subject2</td>
</tr>
<tr>
<td>Extreme Learning Machine [27]</td>
<td>0.7056±0.1123</td>
<td>0.7126±0.0687</td>
</tr>
<tr>
<td>Linear Discriminant Analysis [28]</td>
<td>0.7859±0.0246</td>
<td>0.8147±0.0274</td>
</tr>
<tr>
<td>Nearest Centroid Classifier [14]</td>
<td>0.8305±0.0312</td>
<td>0.8718±0.0289</td>
</tr>
<tr>
<td>K-Nearest Neighbours (k = 5) [14]</td>
<td>0.8995±0.0015</td>
<td>0.8516±0.0101</td>
</tr>
<tr>
<td>Quadratic Discriminant Analysis [14]</td>
<td>0.9143±0.0076</td>
<td>0.8517±0.0078</td>
</tr>
<tr>
<td>Naive Bayes [12]</td>
<td>0.8742±0.0015</td>
<td>0.8401±0.0053</td>
</tr>
<tr>
<td>Ensemble-Extreme Learning Machine(Majority Voting) [11]</td>
<td>0.9142±0.0098</td>
<td>0.8843±0.0144</td>
</tr>
<tr>
<td>New Method (HAR DSMT-based)</td>
<td>0.9714±0.0014</td>
<td>0.8869±0.0026</td>
</tr>
<tr>
<td>Computational Testing Time For Each Individual Sample</td>
<td>8.6545 ms</td>
<td>14.2733 ms</td>
</tr>
</tbody>
</table>

TABLE III: The Selected Sensors in DSAD Based on \(\text{Div}_{JS}(\cdot)\).

<table>
<thead>
<tr>
<th>Subject</th>
<th>(\text{SDD}: \text{Y}<em>2(\cdot) = \text{Div}</em>{JS}(\cdot))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sitting</td>
<td>(RA_{zyg} \quad LA_{zmag} \quad LA_{zacc} \quad LA_{zmag} \quad RA_{zacc})</td>
</tr>
<tr>
<td>Standing</td>
<td>(RL_{zacc} \quad RA_{zmag} \quad RL_{zacc} \quad LA_{zmag} \quad T_{zyg})</td>
</tr>
<tr>
<td>Lying</td>
<td>(T_{zacc} \quad T_{zmag} \quad RA_{zacc} \quad RA_{zmag} \quad LA_{zmag})</td>
</tr>
<tr>
<td>Walking</td>
<td>(RL_{zmag} \quad RL_{zacc} \quad RL_{zacc} \quad RL_{zacc} \quad T_{zyg})</td>
</tr>
<tr>
<td>Running</td>
<td>(RL_{zmag} \quad LL_{zmag} \quad LL_{zmag} \quad RL_{zacc} \quad T_{zmag})</td>
</tr>
</tbody>
</table>

D. Results on UCI DSAD

1) Effectiveness of the Selection of \(\Psi(\cdot)\) and \(\text{Y}(\cdot)\) in Eq.(3) and Eq.(6): Similar to the discussions in OPPORTUNITY dataset, we also gave the performance comparisons between the selections of \(\Psi(\cdot)\) and \(\text{Y}(\cdot)\) in DSAD. First, the comparisons of recognition accuracy with different evaluation criterion was shown in Fig.8 when \(\Psi(\cdot)\) and \(\text{Y}(\cdot)\) were set to (1) \(\Psi(\cdot) = \text{Mean}(\cdot)\), (2) \(\text{Y}_1(\cdot) = \text{Div}_{KL}(\cdot)\), (3) \(\text{Y}_2(\cdot) = \text{Div}_{JS}(\cdot)\), respectively. Different from the phenomenon in Fig.4, our proposed method based on \(\text{Div}_{KL}(\cdot)\) and \(\text{Div}_{JS}(\cdot)\) could give higher recognition accuracy in DSAD. Due to the robust performance of our proposed method based on \(\text{Y}(\cdot) = \text{Div}_{JS}(\cdot)\) in DSAD, in the following experiments, the discriminative function \(\text{Div}_{JS}\) was applied in Eq.(6). Besides, the sensors selected by \(\text{Div}_{JS}\) were also listed in Table III. It can be found that the sensitivity of sensors to different daily activities varied, and was influenced by their locations of deployment and the types of sensors. In Table III, \(T: \text{Torso}; RA: \text{Right Arm}; LA: \text{Left Arm}; RL: \text{Right Leg}; LL: \text{Left Leg}; x, y, z \text{ accerometers}; x, y, z \text{ magnetometers}; x, y, z \text{ gyroscopes}."

Fig. 8: Effectiveness of the selection of SSD(\(\Psi(\cdot)\)) and SDD(\(\text{Y}(\cdot)\)) in DSAD.

Fig. 9: Classification accuracy vs. training percentage for DSAD (SDD: \(\text{Y}_2(\cdot) = \text{Div}_{JS}(\cdot)\)).

2) Recognition Rate versus Training Percentage: In this part, we also varied the percentage of training set and investigated the relationship between the training percentage and the classification accuracy of our proposed method on DSAD. Similar to the experiments in OPPORTUNITY dataset, here we also conducted ten independent experiments. The average accuracy rates of all ten experiments can be seen in Fig.9.
In this part, we further gave the was only selected for the specific activity, which did guaranty modeling strategy proposed in this paper: base KDE model two groups of phenomena further verified the rationality of the performance of the final fused model was 92.4498%. These 86.6024%, 91.9036%, 87.2459%, 91.9036%, 91.9036% and the performance of final fused model was LL-zacc:93.6386%] and the final rate of fused model was accuracy of all base classifiers was [RA-yacc: 74.0080%, consistency with those mentioned in the literature, which repeated 50 experiments and compared DSmT-based method our method had a higher recognition rate in identifying the characteristics of multiple sensors; conversely, DSmT-based approaches. Although SVM and dRNN were powerful models strategy could achieve even higher accuracy than traditional codes for dRNN could be downloaded from [29]). As shown into the second class. The overall accuracy rate of support versus-the-rest method, each type of activity was assumed as the first class and the remaining 4 activity types were grouped in Table IV. For support vector machine, following the one-versus-the-rest method, each type of activity was assumed as the first class and the remaining 4 activity types were grouped into the second class. The overall accuracy rate of support vector machine was calculated as 87.6%. Besides, we also conducted performance comparison between our technique and differential Recurrent Neural Networks [29] 0.8956 50.9993 ms New Method (HAR DSmT-based) 0.9515 17.0964 ms 0.8956 50.9993 ms 0.9515 17.0964 ms 0.876 25.9724 ms 0.8956 50.9993 ms 0.9515 17.0964 ms 0.758 27.4170 ms 0.9018 19.4653 ms 0.876 25.9724 ms.

### TABLE IV: Comparison with state-of-the-art results on UCI DSAD.

<table>
<thead>
<tr>
<th>Reported Methods</th>
<th>Accuracy</th>
<th>Computational Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Artificial Neural Networks [24]</td>
<td>0.743</td>
<td>23.2442 ms</td>
</tr>
<tr>
<td>Bayesian Decision Making [24]</td>
<td>0.758</td>
<td>27.4170 ms</td>
</tr>
<tr>
<td>K-Nearest Neighbours [24]</td>
<td>0.860</td>
<td>20.2664 ms</td>
</tr>
<tr>
<td>Support Vector Machines [24]</td>
<td>0.876</td>
<td>25.9724 ms</td>
</tr>
<tr>
<td>differential Recurrent Neural Networks</td>
<td>0.8956</td>
<td>50.9993 ms</td>
</tr>
</tbody>
</table>

**New Method (HAR DSmT-based)** 0.9515 17.0964 ms
In this paper, we addressed the challenge of HAR problem in BSNs from the perspective of multi-sensor fusion strategy and exploited the unique DSmT-Based fusion strategy. In this novel fusion strategy, there were two points worth mentioning: 1) unlike traditional fusion strategy, not all sensor readings were used for modeling and fusing; only the selected representative sensors were finally fused; 2) BBA of each test sample was constructed according to KDE models. Besides, the vacuous BBA was directly given when test sample had incomplete pattern. Extensive performance evaluations on two wearable sensor-based HAR datasets (OPPORTUNITY dataset and DSAD) demonstrated that the proposed approach outperformed start-of-the-art methods in accuracy. In our future work, we will explore the performance of the proposed method in complex activity recognition. In this work, our proposed DSmT-based model was currently trained and tested offline. In our future research works, we will investigate and test how such new model can be applied to an online activity recognition system in real-time.

ACKNOWLEDGMENT

The authors thank the reviewers and editors for giving valuable comments, which are very helpful for improving this manuscript.

REFERENCES


