Learning from Less for Better: Semi-Supervised Activity Recognition via Shared Structure Discovery

Lina Yao[1], Feiping Nie[2], Quan Z. Sheng[3], Tao Gu[4], Xue Li[5], and Sen Wang[6]
[1]UNSW Australia, NSW 2052, Australia, lina.yao@unsw.edu.au
[2]Northwestern Polytechnical University, Shaanxi 710072, China, feipingnie@gmail.com
[3]The University of Adelaide, Adelaide, SA 5005, Australia, michael.sheng@adelaide.edu.au
[4]RMIT University, Melbourne, VIC 3000, Australia, tao.gu@rmit.edu.au
[5]The University of Queensland, QLD 4072, Australia, xueli@itee.uq.edu.au
[6]Griffith University, QLD 4111, Australia, sen.wang1982@gmail.com

ABSTRACT
Despite the active research into, and the development of, human activity recognition over the decades, existing techniques still have several limitations, in particular, poor performance due to insufficient ground-truth data and little support of intra-class variability of activities (i.e., the same activity may be performed in different ways by different individuals, or even by the same individuals with different time frames). Aiming to tackle these two issues, in this paper, we present a robust activity recognition approach by extracting the intrinsic shared structures from activities to handle intra-class variability, and the approach is embedded into a semi-supervised learning framework by utilizing the learned correlations from both labeled and easily-obtained unlabeled data simultaneously. We use $\ell_{2,1}$ minimization on both loss function and regularizations to effectively resist outliers in noisy sensor data and improve recognition accuracy by discerning underlying commonalities from activities. Extensive experimental evaluations on four community-contributed public datasets indicate that with little training samples, our proposed approach outperforms a set of classical supervised learning methods as well as those recently proposed semi-supervised approaches.

ACM Classification Keywords
I.5 Pattern Recognition

Author Keywords
Activity recognition; shared structure analysis; semi-supervised learning; optimization

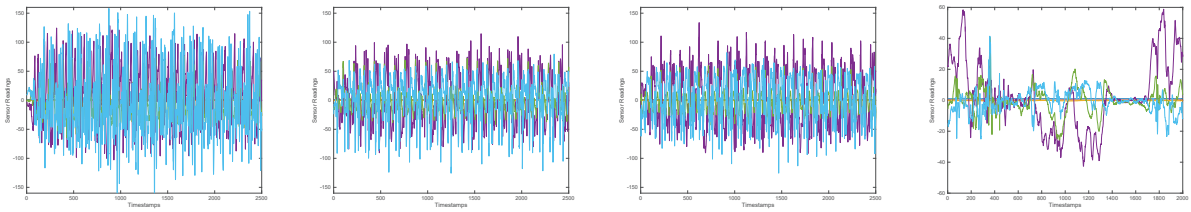
INTRODUCTION
Recent advances in sensor technologies and the growing interest in many ubiquitous applications (e.g., sleep quality monitoring [19], rehabilitation [32], intelligent assisted living [1, 31]) and abnormal behavior identification [30] have been stimulating the demand for learning human activities and behaviors. For example, an assistant service can track how completely and consistently an elderly person's daily routines are performed, and determine when assistance is needed (e.g., a fall occurs) [43].

Despite much efforts on different aspects of human activity recognition [11] over the past few years, challenges still remain in the community. One challenge is how to effectively deal with intra-class variability [10]. Such variability occurs due to the fact that the same activity may be performed differently by different individuals. If an activity recognition model is trained for a single person, so-called *person-dependent training*, the robustness to the intra-person variability in performing a specific activity can be increased by using a large amount of training data that capture as much of activity variability as possible. Clearly, the design of such a system is subject to a trade-off between the use of highly specific and discriminative information, and the information that is more generic and therefore potentially less discriminative but more robust across different people (e.g., *person-independent*).

In many cases, even a same individual may perform the same activity in different ways. Several factors can affect the performance of an activity, such as stress, fatigue or emotional or environmental state in which the activity is performed. For example, the walking style of a person may be more dynamic in the morning after a good sleep than in the evening after a full day of hectic working schedule. To illustrate, Figure 1 shows two subjects performing different activities from a public dataset named USC-HAD [44]. We draw the key observations as follows:

- The difference of different people performing same activities are significant and discriminative (e.g., subject 2 and subject 5 both are walking forward as shown in Figures 1 (a) and (b)).
- The difference of the same person performing different activities is also obvious, (e.g., subject 2 is standing shown in Figure 1 (d) and walking forward shown in Figure 1 (a)). We can discern the different patterns between these

UbiComp '16, September 12-16, 2016, Heidelberg, Germany
©2016 ACM. ISBN 978-1-4503-4461-6/16/09…$15.00
DOI: http://dx.doi.org/10.1145/2971648.2971701



(a) (b) (c) (d)

Figure 1. An illustrative example of intra-class variability: (a) Subject 2: Walking Forward, (b) Subject 5: Walking Forward, (c) Subject 5: Walking Forward in another time frame, and (d) Subject 2: Standing

two types of activities, which can be easily distinguished by any discriminative classifier.

- Same activities performed by a same person in different time frames are inconsistent, e.g., subject 5 is walking forward in trial 1 shown in Figure 1 (b) and walking forward in trial 2 shown in Figure 1 (c), respectively. Even though the data variations are slightly different (e.g., frequency or amplitude), we still observe the common fluctuation patterns, e.g., periodical pattern.

Intuitively, it is reasonable to assume that there are certain underlying local commonalities under the intra-class variabilities shared by all the activities. In other words, there exists a shared subspace for the original data space. For example, the same activities performed by different people share some commonality (e.g., *walking forward* and *standing* both share a torso perpendicular-like motion) or the same activities performed by the same person but with different time frames (e.g., in the morning or at night), while they hold differences inherited from different people. These intrinsic relationships have invariant properties and are less sensitive and variant with different subjects, which can be used as a signature to profile each activity, and to be leveraged further in recognition. However, very few work has been done to capture this underlying correlations to differentiate different activities. To discover such invariant commonalities, we design to uncover a shared structure from original data space using subspace learning to obtain a robust interpretation of intra-class variability, along with discriminative information of original feature space from inter-class variability together. In this way, the activity label of a sequence of sensor data is predicted by its vector representation in the original feature space and the embedding in the shared subspace.

Another challenge is the need of properly labeled training data as many activity recognition approaches rely on machine learning techniques. The majority of the existing work is based on supervised-learning (e.g., Bayesian derivative models) [4, 17, 22, 7, 38]. Such approach relies heavily on sufficient annotated recordings of activity data in order to train a machine learning algorithm. Obtaining such data, especially with sufficiently annotations or the ground truth, is tedious, time-consuming, error-prone, and may even be impossible in some cases (e.g., older people with dementia or collecting data from doing dangerous activities), thus posing a significant barrier to progress in the field. In addition, this approach can only recognize pre-selected activities and does not handle behavior changes.

To reduce the labeling cost, researchers have been investigating different techniques such as weakly supervised learning [33], transfer learning [37] and daily self-recall methods [39]. Semi-supervised based approach has been drawn much attention since it effectively leverages on unlabeled data during the training phase to save tedious labeling task [34, 14, 35]. However, these approaches treat label propagation as a separate process, for example, labeling the unlabeled data using graph label propagation [34] or a high precision classifier [35] first, and then put the labeled data in the training dataset for performing recognition. Therefore, correlations between labeled data and unlabeled data cannot be effectively exploited and utilized during the recognition stage.

To address the aforementioned limitations, in this paper, we propose a novel approach by integrating the shared structure analysis into a $\ell_{2,1}$ based semi-supervised learning, we present a new joint learning framework with embedded shared knowledge discovery, wherein we discern commonality between activities by learning a low dimensional subspace from sensor data, which is robust to intra-class variability and yet still preserve the local discriminant information. In the meanwhile, the manifold structure hidden in a large volume of easily-obtained unlabeled or sparsely labeled data can be effectively exploited as well. The proposed unified framework is robust and can achieve good performance with little training dataset. The main contributions are as follows:

- We propose to improve activity recognition by learning shared knowledge of activities from heterogeneous low-level features extracted from sensor data. The extracted shared structure is robust to intra-class variability in the activity recognition task.

- We aggregate the shared structure into a unified semi-supervised learning framework, which incorporates both labeled and unlabeled data simultaneously to improve the performance of activity recognition and relax the demand on data annotation. Specially, we impose $\ell_{2,1}$ minimization on both loss function and regularization term within this framework to better handle noisy sensor data.

- We extensively evaluate our proposed method using four public datasets. Our approach outperforms the state-of-the-art techniques in terms of accuracy (e.g., over 70%) even when only very few labeled training data are used (e.g., two samples per activity class). The result is promising, and may have practical implication for real-world applications.

RELATED WORK

With the advancement in sensor and wireless communication technologies over the past few years, much attentions have been drawn to sensor-based activity recognition. Significant efforts have been devoted on how to model and predict human activity patterns from ubiquitous sensor data.

Researchers have developed several machine learning based techniques to recognize low-level human activities (e.g., sitting, or walking) from various types of sensory data, such as Naive Bayes based model [36, 4], decision-tree based methods [20, 24], *k* nearest neighbors model [39], SVM based solutions [16], and boosting [29]. The authors in [12] propose an approach to decompose on-body sensor data into events and scenes by using hierarchical HMM models. In [24], the authors adopt LDA transformation to detect activity patterns.

While these methods typically achieve high recognition performance, they require significant amounts of well labeled activity data for training classifiers. It is well known in the ubiquitous community that annotation is particularly time-consuming and tedious. Semi-supervised learning (SSL) approach has become more appealing in recent years. Semi-supervised learning allows to use both labeled and unlabeled data to train a recognition system. They typically require only a small part of labeled training data in addition to a large amount of unlabeled data. With this approach, it is possible to record long-term activity data without the need of detailed continuous activity annotations. It is sufficient to ask users to provide occasional labels about their current activities and use labeled data together with the remaining unlabeled data for learning activity models.

Several semi-supervised recognition systems have been proposed recently. Bhattacharya et al. [6] propose a sparse coding framework that exploits unlabeled data to handle the annotation burden. The approach only requires small amount of labeled data for bootstrapping very effective recognition systems. The authors in [23] propose a semi-supervised kernel logistic regression, by extending kernel logistic regression to semi-supervised learning approach, and the model is solved by using EM algorithm. The work in [25] proposes a Bayesian model to integrate the outputs of a very low number of sensors, and models human activities as a first order Markov chain. Both supervised and semi-supervised learning are developed. Longstaff et al. [21] explore the feasibility of using various semi-supervised and active learning methods to improve activity classification on mobile phones after application deployment. Nguyen-Dinh et al. [26] propose a semi-supervised Gaussian Mixture model by leveraging both labeled and unlabeled samples with multiple mixture components per context class for activity recognition.

Our framework is particularly building upon graph-based semi-supervised learning, which focuses on exploit the relationships between labeled and unlabeled data samples by exploring the manifold structure over similarity graph during the training stage, then a graph Laplacian is implemented for semi-supervised learning problem [45]. Our approach is closed to Stikic et al's work [34], which proposes a graph-based semi-supervised recognition framework to annotate unlabeled samples with activity label, where multiple graphs connecting labeled and unlabeled samples are constructed to propagate the labels based on similarities in terms of time and spatial between features. Our work is also built upon a graph regularization based semi-supervised learning framework, wherein we explore the low-dimensional graph manifold between activities, which is robust to activity variations under the assumption that the class label of input data can be predicted by a linear classifier. The main differences of our work lie in: 1) we adopt $\ell_{2,1}$-norm other than ℓ_2 minimization as the regularization and loss function, which makes our method more robust to data noise; 2) we particularly incorporate discriminative analysis into a joint graph-based semi-supervised learning framework, other than only using the graph to propagate the labels and classify them separately.

On the other hand, shared structure learning has been explored in areas of multi-tasking learning [2] and multi-media processing [41]. Inspired by such success, we apply the shared structure learning to uncover the underlying commonality across different activities to cope with intra-class variability. To the best of our knowledge, our work is the very first to investigate the impact of shared structure learning in sensor-based human activity recognition.

THE PROPOSED METHODOLOGY

Figure 2 illustrates the main workflow of our proposed framework. First, we segment raw sensing data using a common sliding window technique, and extract a set of features from each segment. We use a Canonical Correlation Analysis (CCA) [15] based forward searching algorithm to find a subset of features. Our proposed model consists of three main phases: 1) we first develop $\ell_{2,1}$ based shared structured learning [2] to decode the underlying commonality among features (① in Figure 2); 2) we then construct an adjacent graph using selected features of data points based on cosine similarity. We therefore can propagate the labels of unlabeled data in training stage via learning the graph manifold [45] (② in Figure 2); and 3) we combine both shared structure learning and graph-based semi-supervised learning as a joint framework (③ in Figure 2), which is boiled down to an optimization problem, and the model parameters are learned by an iterative process during training stage. In the rest of this section, we give the technical details.

In the following sections, we will describe our proposed framework formulated by shared structure learning (Section 3.1) and $\ell_{2,1}$ based semi-supervised learning integration (Section 3.2), respectively, followed by an optimization algorithm for solving the human activity recognition in Section 3.3.

Shared Structure Learning

Given a datum $x \in X = \{x_1, x_2, ...x_n\} \in \mathbb{R}^d$ represented by a feature vector as training dataset, we aim to learn a set of discriminant functions $f_j(x_i)$, where $j \in [1, c]$ and c is the number of classes, which can predict an output $y \in Y = \{y_1, y_2, ..., y_n\} \in \mathbb{R}^c$ for input x_i. To achieve this goal, we first collect a training dataset $\{(x_i, y_i)\}_{i=1}^n$, and then use a learning algorithm to learn a prediction function that correlates x with y. A common approach to obtain f_j is to minimize the

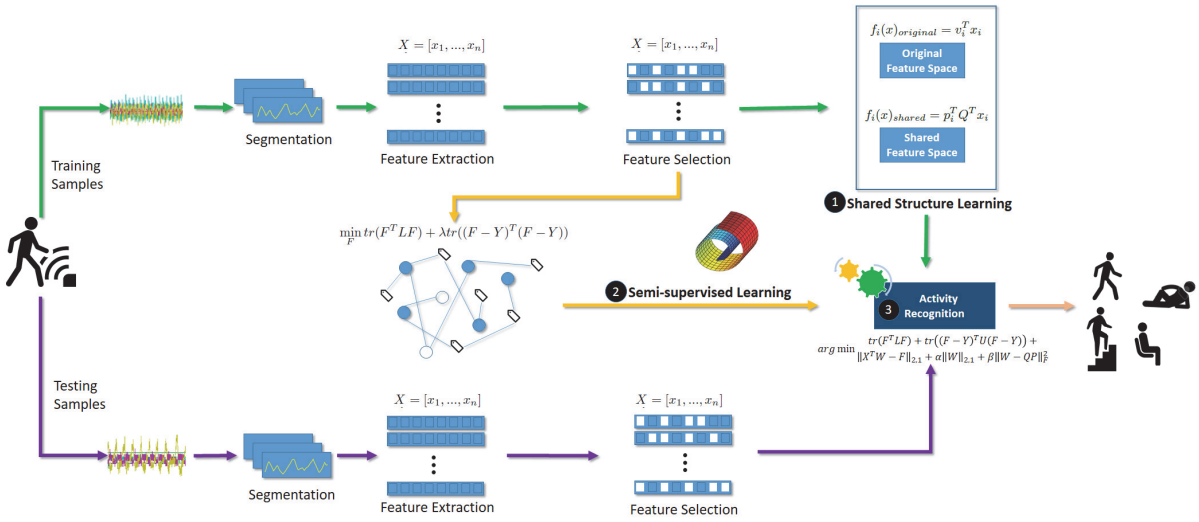


Figure 2. The workflow of our proposed human activity recognition model ③, which is combined with ① shared structure learning (Section 3.1) and ② graph-based semi-supervised learning framework (Section 3.2)

following regularized empirical error [2]:

$$\min_{f_j}[\sum_{i}^{n} loss(f_j(x_i), y_i) + \mu\Omega(f_j)] \tag{1}$$

where $loss(\cdot)$ is the loss function, $\Omega(\cdot)$ is the regularization, and μ is a regularization parameter.

As discussed in the introduction, we intuitively assume that there is a shared underlying structure among the activities, where function $f_j(x_i)$ can be formulated as: $f_j(x_i) = v_j^\top x_i + p_j^\top Q^\top x_i = w_j^\top x_i$, where $w_j = v_j + Qp_j$, v_j and p_j are weight vectors, and $Q \in \mathbb{R}^{d\times sd}$ is the transformation matrix to map the original feature space to the low-dimensional shared subspace by all the features. Thus, the label of activities can be learned by its original feature space along with the shared subspace. By incorporating the shared structure, Equation 1 can be reformulated as:

$$\min_{f} \sum_{j=1}^{c} \left(\sum_{i}^{n} loss((v_j^\top + p_j^\top Q^\top)x_i, y_i) + \mu\Omega(f_j) \right) \quad s.t. Q^\top Q = I \tag{2}$$

where constraints $Q^\top Q = I$ is imposed to make the problem tractable.

Let $X = [x_1, ..., x_n]$ denote the training sensor data matrix, where $x_i \in \mathbb{R}^d (1 \le i \le n)$ is the i-th datum and n is the total number of the training data.

Based on previous analysis, we assume that there is a low-dimensional shared subspace between different activities, $V = [v_1, ..., v_c] \in \mathbb{R}^{d\times c}$ and $P = [p_1, ..., p_c] \in \mathbb{R}^{sd\times c}$ where d and sd are the dimensions of the feature space and the shared subspace. The prediction function can be written as:

$$f(X) = (V + QP)^\top X \tag{3}$$

The objective function in Equation 2 can be formulated as:

$$\begin{aligned} &\min_{f(X),Q,Q^\top Q=I} loss((V+QP)^\top X, Y) + \mu\Omega(f(X)) \\ \Rightarrow &\min_{V,P,Q,Q^\top Q=I} loss((V+QP)^\top X, Y) + \mu(\alpha||V||^2_{Frob} \\ &+\beta||V+QP||_{2,1}) \end{aligned} \tag{4}$$

where the first regularization term $\alpha||V||$ controls the information of shared subspace, and second regularization $\beta||V + QP||$ controls the model complexity.

We further define $W = V + QP$, where $W \in \mathbb{R}^{d\times c}$, and the above function equivalently becomes with replacing the variables $V + QP$ and V in Equation 4:

$$\min_{W,P,Q,Q^\top Q=I} ||X^\top W - Y||_{2,1} + \mu(\alpha||W||_{2,1} + \beta||W - QP||^2_{Frob}) \tag{5}$$

where $||\cdot||_{Frob}$ is the Frobenius norm, $||W||_{2,1} = \sum_{i=1}^{d}\sqrt{\sum_{j=1}^{c} W_{ij}^2}$, and $||X^\top W - Y|| = \sum_{i=1}^{n}\sqrt{\sum_{j=1}^{c}(X^\top W - Y)_{ij}^2}$. Other than the widely used least square loss ℓ_2, which is easily deteriorated by noisy data, we impose a $\ell_{2,1}$ norm on both loss function and regularization, which is robust to outliers and leads to discern discriminant structure [27, 40].

By using the $\ell_{2,1}$ loss function, we can keep the objective function away from noisy sensor data, which often appear in real data set collected with a high-sampling rate. We can also obtain the sparsity of W to naturally retain the most discriminative features associated with the non-zero elements in W. Another contribution is that our approach imposes the $\ell_{2,1}$ regularization on the objective function for exploiting the sparsity characteristics in the sensor data volume.

Embedding into Semi-supervised Learning

Semi-supervised learning aims to leverage both labeled and unlabeled data, with the motivations lying in that labeled data are expensive while abundant unlabeled data are easy to obtain and helpful to improve the performance of learning tasks. A major paradigm for semi-supervised learning is to construct graph to utilize manifold structure for learning [45, 46]. Graph-based semi-supervised models treat all the data points including labeled and unlabeled ones on a graph, which can be approximated to the density and manifold information. Manifold regularization [5] is the most well-known approach based on the graph Laplacian to extend many algorithms to a semi-supervised manner. Specially, data points which can be connected via a path through high density regions on the data manifold are likely to have the same label.

Let $Y = [Y_l^\top, Y_u^\top]^\top = [y_1, ..., y_n]^\top$ be the label matrix for n training samples $X = \{x_1, ..., x_n\}$ from c classes where $x_i|_{i=1}^n \in \mathbb{R}^d$ is the vector representation of the i-th instance. If x_i belongs to j-th class, $y_{ij} = 1$, otherwise $y_{ij} = 0$. Y_l is the label matrix with $Y_l = [y_1, ..., y_m]^\top \in \{0,1\}^{m\times c}$, and $Y_u = [y_{m+1}, ..., y_n]^\top \in \mathbb{R}^{(n-m)\times c}$ with all elements equal to 0. All the data points are assumed to form a graph based on two assumptions: i) nearby data points are highly possible to be assigned in a same class, and ii) nearby data points on the same manifold are consistent with the ground truth labels of the labeled training data. $F = [F_1, ..., F_n]^\top \in \mathbb{R}^{n\times c}$ is the predicted matrix, where $F_i \in \mathbb{R}^{c\times 1}$ is the predicted label vector of the i-th data x_i by the classifier. The idea of manifold and label consistency can be integrated into the unified framework:

$$\begin{aligned} &\min_F \sum_{l=1}^{c}\left[\frac{1}{2}\sum_{i,j=1}^{n}(F_{il}-F_{jl})^2 A_{ij} + \sum_{i=1}^{n} U_{ii}(F_{il}-y_{il})^2\right] \\ &\Rightarrow \min_F \underbrace{tr(F^\top LF)}_{①} + \underbrace{tr((F-Y)^T U(F-Y))}_{②} \end{aligned} \tag{6}$$

where $tr(\cdot)$ is the trace of matrix, A_{ij} is an element of the adjacent graph A constructed from labeled and unlabeled data, L is the Laplacian of Graph A, $U \in \mathbb{R}^{n\times n}$ is a diagonal matrix, whose diagonal element $U_{ii} = \eta$ if x_i is labeled, and $U_{ii} = 0$ otherwise. η is a large constant [41]. The first term ① is the smoothness constraint. Minimizing this part means that all data points should have similar ranking scores if they are contained in many common manifolds. The second term ② measures the difference between the ground-truth labels obtained ranking scores and the pre-given scores.

To utilize the information of unlabeled data, we apply the semi-supervised manifold learning framework illustrated in Equation 6. The basic idea of graph regularization based semi-supervised framework is to define good functional structures using unlabeled data. Since it does not bootstrap labels, there is no label noise which can potentially corrupt the learning procedure. An example of this approach is to use unlabeled data to create a data-manifold (graph structure), on which proper smooth function classes can be defined. If such smooth functions can characterize the underlying classifier well, the graph regularization based semi-supervised approach is able to improve the classification performance.

The local structure graph $A \in \mathbb{R}^{n\times n}$, whose element $A_{ij} = 1$ if we use the cosine similarity to measure the affinity between two data points represented by features. When building such a graph, we only connect each feature vector to its k-nearest neighbors ($k = 3$ in this work) by undirected edges expressing symmetric neighborhoods.

At this point, by combining shared structure in Equation 5 and the semi-supervised framework in Equation 6 under a unified framework, the proposed activity recognition problem is formulated as:

$$\begin{aligned} \arg \min_{F,W,P,Q,Q^\top Q=I} & tr(F^\top LF) + tr((F-Y)^\top U(F-Y)) + \\ & \mu(||X^\top W - F||_{2,1} + \alpha||W||_{2,1} + \beta||W-QP||_F^2) \end{aligned} \tag{7}$$

The Activity Recognition Algorithm

We formulate the human activity recognition problem as objective function in Equation 7, which is not a convex problem and can be solved by converting into an efficient iterative solution [27, 28]. For each iteration, F is calculated with the current W, which is updated based on P. The value of P is calculated based on the value of current Q. The iteration procedure is repeated until the algorithm converges. Equation 7 can be rewritten as:

$$\begin{aligned} \mathscr{L} = \arg \min_{F,W,Q,P,Q^\top Q=I} & \mathrm{tr}F^\top LF + \mathrm{tr}(F-Y)^\top U(F-Y) \\ & + \mu(\mathrm{tr}(X^\top W - F)D_1(X^\top W - F) \\ & + \alpha \mathrm{tr}W^\top D_2 W + \beta \mathrm{tr}(W-QP)^\top (W-QP)) \end{aligned} \tag{8}$$

where D_1 and D_2 are the diagonal matrices with $D_1(ii) = \frac{1}{||z_1^i||_2}$, $Z_1 = X^\top W - F$ and $Z_1 = [z_1^1, ..., z_1^n]^\top \in \mathbb{R}^{n\times c}$, and $D_2(ii) = \frac{1}{||z_2^i||_2}$, $Z_2 = W^\top Q - P$ and $Z_2 = [z_2^1, ..., z_2^n]^\top \in \mathbb{R}^{d\times sd}$.

The optimization process on objective function Equation 8 can be briefly obtained as follows.

Taking derivation w.r.t. P. By setting the derivative of Equation 8 w.r.t. P to zero, we have:

$$\frac{\partial \mathscr{L}}{\partial P} = \beta(2Q^\top QP - 2Q^\top W) = 0 \Rightarrow P = Q^\top W \tag{9}$$

By substituting Equation 9 into Equation 8, the objective function becomes:

$$\begin{aligned} \mathscr{L}' = \min_{F,W,Q} & tr(F^\top LF) + tr(F-Y)^\top U(F-Y) \\ & + \mu(tr(X^\top W - F)^\top D_1(X^\top W - F) + trW^\top(\alpha D_2 \\ & + \beta I - \beta QQ^\top)W) \end{aligned} \tag{10}$$

Taking derivation w.r.t. W. By setting the derivative w.r.t. W to zero, we have:

$$\begin{aligned} \frac{\partial \mathscr{L}'}{\partial W} &= 2X(D_1X^\top W - D_2Y) + 2(\alpha D + \beta I - \beta QQ^\top)W = 0 \\ &\Rightarrow W = (M - \beta QQ^\top)^{-1} X D_1 F \end{aligned} \tag{11}$$

where $M = XD_1X^\top + \alpha D_2 + \beta I$. Then, the objective function becomes:

$$\begin{aligned}\mathscr{L}'' = &\min_{F,Q} tr(F^\top LF) + tr(F-Y)^\top U(F-Y) \\ &+ \mu(trF^\top D_1 F - trF^\top D_1 X^\top (M-\beta QQ^\top)^{-1} XD_1 F)\end{aligned} \tag{12}$$

Taking derivation w.r.t. F. By setting the derivative of Equation 12 w.r.t. F to zero, we have:

$$\begin{aligned}\frac{\partial \mathscr{L}''}{\partial F} &= LF + UF - UY + \mu D_1 F \\ &\quad - \mu D_1 X^\top (M-\beta QQ^\top)^{-1} XD_1 F = 0 \\ &\Rightarrow F = (B - \mu(D_1 X^\top A^{-1} XD_1)^{-1} UY)\end{aligned} \tag{13}$$

where $A = M - \beta QQ^\top$, $B = L + U + \mu D_1$. By substituting F, and the objective function becomes:

$$\mathscr{L}''' = \max_{Q, Q^\top Q} Y^\top U^\top (B - \mu D_1 X^\top A^{-1} XD_1)^{-1} UY \tag{14}$$

According to Woodbury matrix identity [13], the objective function can be written as:

$$\max_{Q, Q^\top Q} YUB^{-1} D_1 X^\top \underbrace{(A - \mu XD_1 B^{-1} D_1 X^\top)^{-1}}_{①} XD_1 B^{-1} UY \tag{15}$$

where $(B - \mu D_1 X^\top A^{-1} XD_1)^{-1} = B^{-1} + \mu B^{-1} D_1 X^\top (A - \mu XD_1 B^{-1} D_1 X^\top)^{-1} XD_1 B^{-1}$. The component ① can be further inferred as follows:

$$\begin{aligned}&(A - \mu XD_1 B^{-1} D_1 X^\top)^{-1} \\ &\Rightarrow (M - \beta QQ^\top - \mu XD_1 B^{-1} D_1 X^\top)^{-1} \\ &\Rightarrow (N - \beta QQ^\top)^{-1} \\ &\Rightarrow N^{-1} + \beta N^{-1} Q(I - \beta Q^\top N^{-1} Q)^{-1} Q^\top N^{-1} \\ &\Rightarrow N^{-1} + \beta N^{-1} Q(Q^\top (I - \beta N^{-1}) Q)^{-1} Q^\top N^{-1}\end{aligned} \tag{16}$$

where $N = M - \mu XD_1 B^{-1} D_1 X^\top$. Then the objective function of Equation 15 becomes:

$$\begin{aligned}&\max_{Q, Q^\top Q = I} tr[Y^\top UB^{-1} D_1 X^\top N^{-1} Q \\ &(Q^\top (I - \beta N^{-1}) Q)^{-1} Q^\top N^{-1} XD_1 B^{-1} UY] \\ &\Rightarrow \max_{Q, Q^\top Q = I} tr(Q^\top (I - \beta N^{-1}) Q)^{-1} \\ &Q^\top N^{-1} XD_1 B^{-1} UY Y^\top UB^{-1} D_1 X^T N^{-1} Q \\ &\Rightarrow \max_{Q, Q^\top Q = I} tr(Q^\top CQ)^{-1} Q^\top HQ\end{aligned} \tag{17}$$

where

$$\begin{aligned}C &= I - \beta N^{-1} = I - \beta(M - \mu XD_1 B^{-1} D_1 X^\top)^{-1} \\ &= I - \beta(XD_1 X^\top + \alpha D_2 + \beta I \\ &\quad - \mu XD_1 (L + U + \mu D_1)^{-1} D_1 X^\top)^{-1}\end{aligned} \tag{18}$$

and

$$H = N^{-1} XD_1 B^{-1} UY Y^\top UB^{-1} D_1 X^\top N^{-1} \tag{19}$$

Algorithm 1: Proposed Activity Recognition Algorithm

Input: Training sample matrix $X \in R^{d \times n}$;
Training label matrix $Y \in \mathbb{R}^{n \times c}$;
Model parameters α, β and μ
Output: $W \in \mathbb{R}^{d \times c}$ matrix

```
Compute the graph Laplacian matrix L ∈ R^{n×n};
Compute the selection matrix U ∈ R^{n×n};
Initialize W ∈ R^{d×c} randomly;
Initialize F ∈ R^{n×c} randomly;
while not convergence do
    Compute the diagonal matrix D1 and D2;
    Compute C = I − β(XD1X^⊤ + αD2 + βI − μXD1(L+U+μD1)^{-1}D1X^⊤)^{-1}
    (Equation 18);
    Compute H = N^{-1}XD1B^{-1}UYY^⊤UB^{-1}D1X^⊤N^{-1} (Equation 19);
    Compute optimal Q (Equation 17);
    Compute F (Equation 13);
    Compute W = (M − βQQ^⊤)^{-1}XF (Equation 20).
end
Output W.
```

where $N = M - \mu XD_1 B^{-1} D_1 X^\top$, $B = L + U + \mu D_1$, and $M = XD_1 X^T + \alpha D_2 + \beta I$.

Up to this point, we can easily calculate Q from Equation 17, then with the obtained Q, we can have:

$$W = (M - \beta QQ^\top)^{-1} XD_1 F \tag{20}$$

where $F = (B - \mu D_1 X^\top A^{-1} XD_1)^{-1} UY$, and $A = M - \beta QQ^\top$, and $B = L + U + \mu D_1$. As we have obtained W, given a testing sensor sample, its predicted activity label can be computed using Equation 3. Algorithm 1 gives a summary of this iterative algorithm.

THEOREM 1. *The global optimal Q^* of Equation 8 can be obtained by solving the following ratio trace maximization problem:*

$$\max_{Q, Q^\top Q = I} tr(Q^\top CQ)^{-1} Q^\top HQ \tag{21}$$

where

$$C = I - \beta N^{-1} = I - \beta(M - \mu XD_1 B^{-1} D_1 X^\top)^{-1} \tag{22}$$

and

$$H = N^{-1} XD_1 B^{-1} UY Y^\top UB^{-1} D_1 X^\top N^{-1} \tag{23}$$

EXPERIMENTS

In this section, we first briefly describe our experimental settings, and then report the evaluation results by comparing with the state-of-the-art activity recognition approaches using four publicly available real-world datasets.

Experimental Setup

In this section, we introduce the four datasets, and also describe the feature extraction and the validation strategy used in our evaluation.

Dataset Description

We validate our model using four public datasets, which are collected from heterogeneous sensors including wireless wearable sensors, embedded mobile phone sensors, RFID passive tags, respectively:

- *USC-HAD* [44]. The dataset is collected in an open environment. There are 14 participants recruited for this study, with diversity in gender (7 males, 7 females), age, height, and weight. Each participant is asked to perform 12 activities (i.e., walking forward, walking left, walking right, going upstairs, going downstairs, jumping up, running, standing, and sitting, sleeping, elevator up, elevator down).

- *Actitracker* [18]. The dataset is collected with 29 subjects performing six activities. All subjects carry Android smartphones in their front pant pockets, and they are asked to walk, jog, ascend stairs, descend stairs, sit, and stand for a period of time.

- *UCI HAR* [3]. A group of 30 volunteers with ages ranging from 19 to 48 years old are selected. Each person is instructed to follow a protocol of activities while wearing a waist-mounted Samsung Galaxy S II smartphone. The six selected ADL are standing, sitting, lying down, walking, walking downstairs and upstairs.

- *Freedom RFID* [42]. The data includes 12 activities performed by 6 subjects. Each subject stands between RFID signal receiver and transmitter and the corresponding signal streaming variations are recorded while performing different activities.

Low-Level Feature Extraction

It is noted that for simplicity and fair comparison, we extract the *same* types of lightweight statistical features from these four datasets. This set of feature types are selected because they are lightweight and easy to compute. We first divide the continuous sequence of signal strength data stream into fixed length data windows. In our experiments, the sensor readings are divided into data segments of length $\Delta t = 10$ seconds. This time slice duration is long enough to be discriminative yet is short enough to provide high accuracy labeling results. The information is then transformed by designing 12 types of lightweight statistical features, which are widely used in the activity recognition community, from each segment. The 12 features are listed in Table 1. Because the scale factors and units of the features described above are different, all the datasets are normalized by standard deviation.

Validation Strategy

Specifically, we use the *person-independent* validation strategy, wherein the first to the $(n-1)^{th}$ subjects as training examples, and data from the left-out subject n^{th} subject are used for testing. This process iterates for every subject over each dataset. The final result is the averaged value over all the subjects. We measure the performance using a number of metrics such as accuracy, precision, recall, F1 score. Due to space constraints, we only show the accuracy performance.

Experimental Results

In this section, we report several empirical studies to compare the performance of our proposed approach with other methods, along with a parameter tuning study and an analysis in efficiency.

Parameter Study

There are several parameters to be tuned in the proposed approach, including the regularizations α, β, μ and k for constructing similarity graph where the graph Laplacian is computed, and *sd* for the dimension of the shared subspace. For simplicity, we only use one dataset for illustrative purpose. We use the Actitracker dataset for studying the parameters of our proposed approach. By fixing other parameters, we vary the dimension of the shared subspace (i.e., *sd*) from 1 to 10 with a step size of 1, along with different number of labeled training samples. Figure 3 (a) shows the results. From the figure we can observe that the recognition accuracy consistently decreases with larger shared subspace on different size of labeled training samples. Bigger shared subspace may capture more discriminant information, but also can possibly cause the over-fitting problem and lead to the deterioration of the approach. In our work, we set $sd = 5$ as the default value.

Next, we study the parameters α and β by tuning them on sets $\{10^i | i = -3,-2,-1,0,1,2,3\}$. Figure 3 (b) illustrates the performance variations with different combinations of α and β by fixing other parameters. We observe that the performance is not quite sensitive to β variations, but is easily affected by α. The best combination is $\alpha = 10$ and $\beta = 0.01$ for the Actitracker dataset. We also explore the effect of different μ and k, and the results show that both of them are not quite sensitive. It is worth noting that how to determine the optimal parameter setting is a non-trivial task and dependent on different datasets.

Overall Comparison

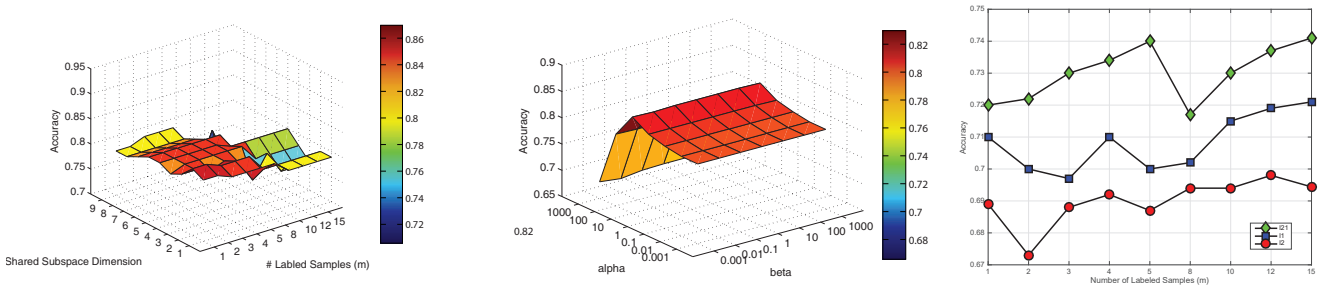
To evaluate the performance of the proposed approach (denoted as RSAR, Robust Semi-supervised Activity Recognition), we conduct extensive experiments to compare its performance with the state-of-the-art methods.

The detailed setting of comparison follows the one in conventional semi-supervised learning approaches. Specially, the training dataset contains both labeled and unlabeled data, and the testing dataset is not available during the training phase. We use c to denote the number of activities per subject in each dataset, i.e., c = 12 and 6, for the USC-HAD dataset and the Actitracker dataset, respectively. We randomly select m labeled samples ($m = 1,2,3,4,5,8,10,12,15$) per activity in the training dataset, thus resulting in $m \times c$ randomly labeled sensor samples. The remaining training samples are unlabeled. We conduct the experiments using a *person-independent* strategy, wherein we select the training samples from a certain number of persons to train our model, which is then tested by the left-out persons.

We compare our proposed method using settings $\alpha = 0.1$, β = 1, $k = 5$, $sd = 5$ and $\mu = 0.001$, with the following state-of-the-art supervised and semi-supervised methods:

- k-Nearest Neighbor (k**NN**) is a common classifier for a variety of classification problems. It predicts the class of a sample by a majority voting of the class labels of the k nearest training instances. We set $k = 3$ in this work.

- Linear Support Vector Machine (**LSVM**) aims at finding the best separation of binary-labeled instances by deter-



(a) (b) (c)

Figure 3. Performance variations with (a) different dimension (sd) of shared subspace and different number of labeled training samples (m); (b) different α and β; (c) comparison on different minimization of shared structure analysis over Actitracker dataset.

Table 1. Lightweight Statistical Features

No.	Feature	Description
1	Min	Minimum of signal over the segment
2	Max	Maximum of signal over the segment
3	Mean	Average value of signal over the segment
4	Root Mean Square	Quadratic mean value of signal over the segment
5	Variance	Square of standard deviation
6	Standard Deviation	Mean deviation of the signal compared to the average
7	Kurtosis	Degree of peakedness of the sensor signal distribution
8	Skewness	Degree of asymmetry of the sensor signal distribution
9	Entropy	Measure of the distribution of frequency components
10	Median	Median signal value
11	Zero Crossing Rate	Total number of times the signal changes from positive to negative or back, normalized by the segment length
12	Mean Cross Rate	Total number of times the signal changes from below average to above average, normalized by the segment length

mining a hyperplane which maximizes the margin between support vectors of different classes. In this work, the key setting is $C = 1$ with linear kernel.

- Random Forest (**RF**) builds a forest of decision trees that have the same distribution but independent output classes. It is based on a random selection of features for each tree and construction of a combination of the individual tree outputs. The number of of trees is 500 in this work.

- Naive Bayes (**NB**) classifier finds the most posterior probability $Pr(y_k|x^*)$ for a given testing sample x^* as its predicted label $y_k(x^*)$.

$$\begin{aligned} y_k &= \arg\max_{y_k} \frac{Pr(y_k)\prod_j^D Pr(x_j^*|y_k)}{\sum_j^k Pr(y_k)\prod_j^D Pr(x_j^*|y_k)} \\ &= \arg\max_{y_k} Pr(y_k)\prod_j^D Pr(x_j^*|y_k) \end{aligned} \tag{24}$$

In this work, we model the conditional probability $Pr(x_j^*|l_k)$ which follows the Gaussian mixture distribution. The number of Gaussian clusters is adaptively determined for each training dataset via adopting the Dirichlet Process Gaussian Mixture Model (DPGMM) [8].

- **Co-training**: This method [9] is developed for boosting the performance of learning algorithms given insufficient training data. It has been adopted in [35] for semi-supervised activity recognition. We adopt the two classifier strategies (using logistic regression and naive Bayes for two groups of features, e.g., features 1 - 6, features 7 - 12) in this work, which are initially trained with a small set of labeled data. Each classifier is trained on a different subset of the training data. At the beginning, only a small number of labeled training samples are available. When a classifier is trained, it is used to predict the labels of the unlabeled data and assign confidences to the predictions. The top several samples, which are predicted to be positive or negative with high-confidences, are then selected and removed from the unlabeled data set. These samples are then merged into the labeled training data set and used for training the other classifier. Recursively, the smaller set of labeled training samples is augmented by a larger and unlabeled samples.

- **GLSVM**. This method [34] is built on graph-based label propagation with SVM classifier, denoted in this paper as GLSVM. It is a graph-based semi-supervised learning framework and works in the following way. Firstly, a graph is constructed and used to estimate the underlying structure of data using both labeled and unlabeled data. The predicted labels of unlabeled data are then propagated on this graph by learning the manifold. In contrast to our method, wherein the graph manifold is unified into a robust joint framework with shared subspace learning, GLSVM uses both the initial labeled training dataset and the propagated labels to train the SVM classifier.

Figure 4 shows the recognition results using the four datasets w.r.t. different number of labeled training data. We can draw the following key observation: our method gains better performance comparing to the comparison methods when we only include a *very small* number of labeled data in training, although that difference is not statistically significant when more training samples are included. For example, when *only one* labeled training sample from each activity of subjects is used, our method achieves 71.34% accuracy for the Actitracker dataset, which is significantly better than other methods (see Figure 4 (b)). In the meanwhile, we can observe that our method consistently attains the best recognition performance on the four datasets, only slightly inferior to SVM over

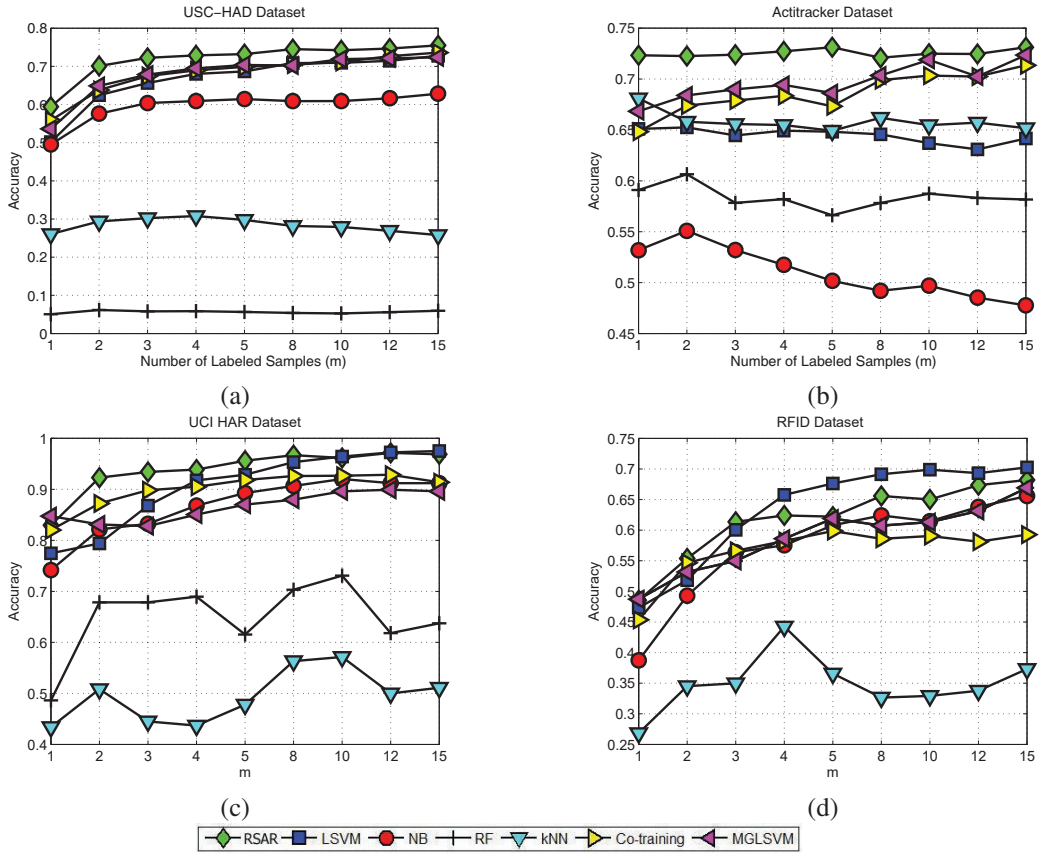


Figure 4. Overall recognition performance comparison using (a) the USC-HAD dataset, (b) the Actitracker dataset, (c) the UCI HAR dataset, and (d) the RFID dataset

RFID dataset with more labeled samples are included. In addition, the recognition accuracy of most methods is improved when the amount of labeled training samples increases. The results have indicated that our algorithm has robust performance by learning the analysis of the correlations between different features in a shared structure.

To examine the effectiveness of $\ell_{2,1}$ regularization on shared structure, we also compare $\ell_{2,1}$ minimization with ℓ_2 and ℓ_1 over the Actitracker dataset. From Figure 3(c), we can observe that ℓ_{21} minimization consistently outperforms the other methods. The experimental results on other datasets have the similar results.

We also take a closer look at the recognition accuracy on each specific activity. Figure 5 illustrates the confusion matrices of all datasets in terms of average accuracy per activity, with default settings of our proposed method, and $m = 2$ (only two labeled training samples from each activity). Taking the USC-HAD dataset as an example, we can see that for most activities, our approach works well (e.g., *Sleeping*, *Sitting* and *Jumping*). Only for similar activities such as *Walking Forward*, *Walking Left* and *Walking Downstairs*, there are some misclassification (Figure 5 (a)). The similar results can also be observed on other three datasets (Figure 5 (b) - (d)).

We present some brief analysis on the reasons why our proposed model performs better. Firstly, our approach takes the advantage of both labeled and unlabeled data during the training stage, which makes it outperform the supervised learning based methods when only a small number of labeled data are available. Secondly, our unified model can simultaneously learn from feature correlations. Specifically, compared with the traditional graph-based semi-supervised methods, we adopt the $\ell_{2,1}$ minimization on both loss function and regularization. The $\ell_{2,1}$ norm based loss function is robust to outliers of data points, and the $\ell_{2,1}$ norm regularization can select features across all data points with joint sparsity.

Discussion

In this section, we draw some discussions on efficiency concerns and implications of the proposed model. Since the method needs to calculate a few inverse matrices during the training stage, it may cause the concerns of efficiency as inverse matrix computation lead to more computational cost. We briefly analyze the computation complexity of the proposed approach in both the training stage (model learning) and the testing stage (activity recognition).

There are two main sources that contribute to the complexity of the training stage: i) computing Laplacian matrix L

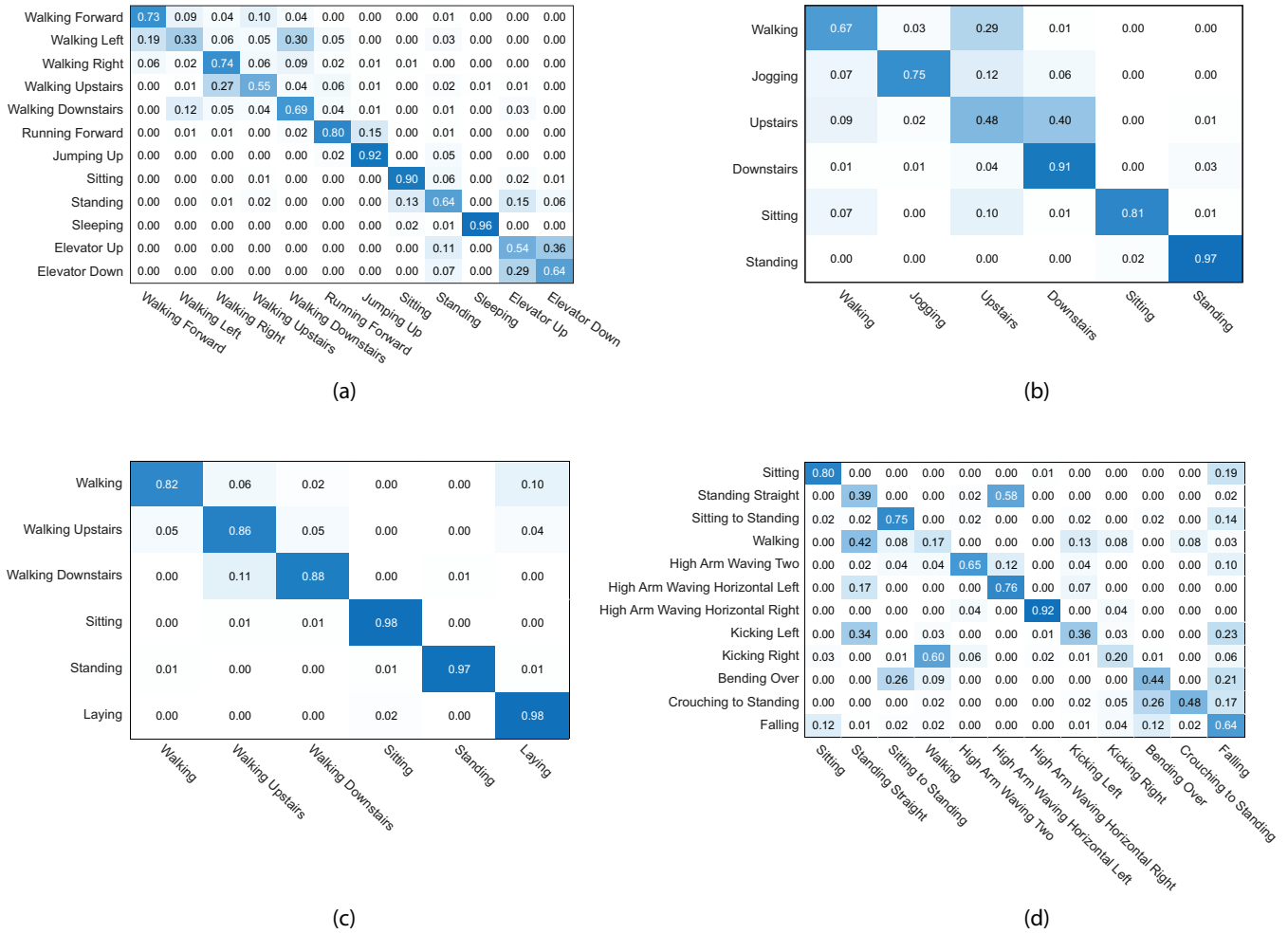


Figure 5. The confusion matrix of our proposed approach for activity recognition on (a) the USC-HAD dataset, (b) the Actitracker dataset, (c) UCI HAR dataset (d) RFID dataset

takes the complexity of $\mathcal{O}(d \times n^2)$, and ii) during the optimization process, we need to compute the inverse of several matrices and perform eigen-decomposition, and its complexity is $\mathcal{O}\{\max(d^3, n^3)\}$, where d is the feature dimension and n is the number of features. It is noted that n is usually bigger than d. Therefore, for the training stage, the approximate complexity is dominant by $\mathcal{O}(n^3)$. For the recognition stage, we need to perform $c \times d$ times multiplications to assign the predicted label for an input testing sample, where c is the number of activity class. Our experiments study the computation time of activity recognition with different number of labeled training samples. The results show that efficiency increases with larger size of training samples, but the overall recognition running time is generally feasible and tolerable for an activity recognition system.

CONCLUSION AND FUTURE WORK

In this paper, we propose an approach to recognize human activities via discerning commonalities between multi-type features. First, our approach can simultaneously uncover the intrinsic and invariant low-dimensional subspace to improve overall activity recognition performance against intra-class variability. Second, we adopt the $\ell_{2,1}$ minimization to make our approach more robust to sensor reading outliers. Third, we unify both solutions into a semi-supervised learning framework for reducing the tedious efforts on manual labeling. We conduct extensive experiments to compare our approach with the state-of-the-art methods including supervised and semi-supervised methods. The experimental results on four public datasets show the promising performance of our proposed approach. Our approach outperforms all of these methods, especially when the size of labeled data is small.

In the future, we would like to further explore our approach in a larger dataset, which will be collected from a real living environment and contain more types of common activities. We also plan to extend our model for detecting complex activities (e.g., human daily routines) by elaborating a set of reasoning strategies embedded into our current model. As our model can capture the activity patterns across different persons by learning the shared commonality, another interesting direction for us is to investigate the possibility of identifying a specific user via her/his unique activity patterns after eliminating the learned commonality.

REFERENCES

1. Ryan Aipperspach, Elliot Cohen, and John Canny. 2006. Modeling human behavior from simple sensors in the home. In *Pervasive Computing*. Springer, 337–348.
2. Rie Kubota Ando and Tong Zhang. 2005. A framework for learning predictive structures from multiple tasks and unlabeled data. *The Journal of Machine Learning Research* 6 (2005), 1817–1853.
3. Davide Anguita, Alessandro Ghio, Luca Oneto, Xavier Parra, and Jorge L Reyes-Ortiz. 2013. A public domain dataset for human activity recognition using smartphones. In *European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, ESANN*.
4. Ling Bao and Stephen S Intille. 2004. Activity recognition from user-annotated acceleration data. In *Pervasive computing*. Springer, 1–17.
5. Mikhail Belkin, Partha Niyogi, and Vikas Sindhwani. 2006. Manifold regularization: A geometric framework for learning from labeled and unlabeled examples. *The Journal of Machine Learning Research* 7 (2006), 2399–2434.
6. Sourav Bhattacharya, Petteri Nurmi, Nils Hammerla, and Thomas Plötz. 2014. Using unlabeled data in a sparse-coding framework for human activity recognition. *Pervasive and Mobile Computing* 15 (2014), 242–262.
7. Ulf Blanke, Bernt Schiele, Matthias Kreil, Paul Lukowicz, Bernhard Sick, and Thiemo Gruber. 2010. All for one or one for all? Combining Heterogeneous Features for Activity Spotting. In *Pervasive Computing and Communications Workshops (PERCOM Workshops), 2010 8th IEEE International Conference on*. IEEE, 18–24.
8. David M Blei, Michael I Jordan, and others. 2006. Variational inference for Dirichlet process mixtures. *Bayesian analysis* 1, 1 (2006), 121–143.
9. Avrim Blum and Tom Mitchell. 1998. Combining labeled and unlabeled data with co-training. In *Proceedings of the eleventh annual conference on Computational learning theory*. ACM, 92–100.
10. Andreas Bulling, Ulf Blanke, and Bernt Schiele. 2014. A tutorial on human activity recognition using body-worn inertial sensors. *ACM Computing Surveys (CSUR)* 46, 3 (2014), 33.
11. Liming Chen, Jesse Hoey, Chris D Nugent, Diane J Cook, and Zhiwen Yu. 2012. Sensor-based activity recognition. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on* 42, 6 (2012), 790–808.
12. B. Clarkson and A. Pentland. 1999. Unsupervised clustering of ambulatory audio and video. In *Acoustics, Speech, and Signal Processing, 1999. Proceedings., 1999 IEEE International Conference on*, Vol. 6. IEEE, 3037–3040.
13. Gene H Golub and Charles F Van Loan. 2012. *Matrix computations*. Vol. 3. JHU Press.
14. Enamul Hoque and John Stankovic. 2012. AALO: Activity recognition in smart homes using Active Learning in the presence of Overlapped activities. In *Pervasive Computing Technologies for Healthcare (PervasiveHealth), 2012 6th International Conference on*. IEEE, 139–146.
15. Harold Hotelling. 1936. Relations between two sets of variates. *Biometrika* (1936), 321–377.
16. Tâm Hunh, Ulf Blanke, and Bernt Schiele. 2007. Scalable recognition of daily activities with wearable sensors. In *Location-and context-awareness*. Springer, 50–67.
17. Nicky Kern, Bernt Schiele, Holger Junker, Paul Lukowicz, and Gerhard Tröster. 2003. Wearable sensing to annotate meeting recordings. *Personal and Ubiquitous Computing* 7, 5 (2003), 263–274.
18. Jennifer R Kwapisz, Gary M Weiss, and Samuel A Moore. 2011. Activity recognition using cell phone accelerometers. *ACM SigKDD Explorations Newsletter* 12, 2 (2011), 74–82.
19. Jason J Liu, Wenyao Xu, Ming-Chun Huang, Nabil Alshurafa, Majid Sarrafzadeh, Nitin Raut, and Behrooz Yadegar. 2013. A dense pressure sensitive bedsheet design for unobtrusive sleep posture monitoring. In *Pervasive Computing and Communications (PerCom), 2013 IEEE International Conference on*. IEEE, 207–215.
20. Beth Logan, Jennifer Healey, Matthai Philipose, Emmanuel Munguia Tapia, and Stephen Intille. 2007. *A long-term evaluation of sensing modalities for activity recognition*. Springer.
21. Brent Longstaff, Sasank Reddy, and Deborah Estrin. 2010. Improving activity classification for health applications on mobile devices using active and semi-supervised learning. In *2010 4th International Conference on Pervasive Computing Technologies for Healthcare*. IEEE, 1–7.
22. Paul Lukowicz, Jamie A Ward, Holger Junker, Mathias Stäger, Gerhard Tröster, Amin Atrash, and Thad Starner. 2004. Recognizing workshop activity using body worn microphones and accelerometers. In *Pervasive Computing*. Springer, 18–32.
23. Ryunosuke Matsushige, Koh Kakusho, and Takeshi Okadome. 2015. Semi-supervised learning based activity recognition from sensor data. In *2015 IEEE 4th Global Conference on Consumer Electronics (GCCE)*. IEEE, 106–107.
24. Uwe Maurer, Asim Smailagic, Daniel P Siewiorek, and Michael Deisher. 2006. Activity recognition and monitoring using multiple sensors on different body positions. In *Wearable and Implantable Body Sensor Networks, 2006. BSN 2006. International Workshop on*. IEEE, 4–pp.

25. Alfredo Nazabal, Pablo Garcia-Moreno, Antonio Artes-Rodriguez, and Zoubin Ghahramani. 2015. Human activity recognition by combining a small number of classifiers. (2015).
26. Long-Van Nguyen-Dinh, Mirco Rossi, Ulf Blanke, and Gerhard Tröster. 2013. Combining crowd-generated media and personal data: semi-supervised learning for context recognition. In *Proceedings of the 1st ACM international workshop on Personal data meets distributed multimedia*. ACM, 35–38.
27. Feiping Nie, Heng Huang, Xiao Cai, and Chris H Ding. 2010. Efficient and robust feature selection via joint *l*2, 1-norms minimization. In *Proc. of the 24th Annual Conference on Neural Information Processing Systems (NIPS)*. 1813–1821.
28. Kaare Brandt Petersen, Michael Syskind Pedersen, and others. 2008. The matrix cookbook. *Technical University of Denmark* 7 (2008), 15.
29. Nishkam Ravi, Nikhil Dandekar, Preetham Mysore, and Michael L Littman. 2005. Activity recognition from accelerometer data. In *AAAI*, Vol. 5. 1541–1546.
30. Daniele Riboni, Claudio Bettini, Gabriele Civitarese, Zaffar Haider Janjua, and Rim Helaoui. 2015. Fine-grained recognition of abnormal behaviors for early detection of mild cognitive impairment. In *Pervasive Computing and Communications (PerCom), 2015 IEEE International Conference on*. IEEE, 149–154.
31. Julia Seiter, Walon Wei-Chen Chiu, Mario Fritz, Oliver Amft, and Gerhard Tröster. 2015. Joint Segmentation and Activity Discovery Using Semantic and Temporal Priors. In *Proc. of the 2015 IEEE International Conference on Pervasive Computing and Communications (PerCom 2015)*. St. Louis, MO, USA.
32. Gina Sprint, Douglas Weeks, Vladimir Borisov, and Diane Cook. 2014. Wearable sensors in ecological rehabilitation environments. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*. ACM, 163–166.
33. Maja Stikic, Diane Larlus, Sandra Ebert, and Bernt Schiele. 2011. Weakly supervised recognition of daily life activities with wearable sensors. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 33, 12 (2011), 2521–2537.
34. Maja Stikic, Diane Larlus, and Bernt Schiele. 2009. Multi-graph based semi-supervised learning for activity recognition. In *Wearable Computers, 2009. ISWC'09. International Symposium on*. IEEE, 85–92.
35. Maja Stikic, Kristof Van Laerhoven, and Bernt Schiele. 2008. Exploring semi-supervised and active learning for activity recognition. In *Wearable computers, 2008. ISWC 2008. 12th IEEE international symposium on*. IEEE, 81–88.
36. Emmanuel Munguia Tapia, Stephen S Intille, and Kent Larson. 2004. *Activity recognition in the home using simple and ubiquitous sensors*. Springer.
37. TLM Van Kasteren, Gwenn Englebienne, and Ben JA Kröse. 2010. Transferring knowledge of activity recognition across sensor networks. In *Pervasive computing*. Springer, 283–300.
38. Tim Van Kasteren, Athanasios Noulas, Gwenn Englebienne, and Ben Kröse. 2008. Accurate activity recognition in a home setting. In *Proceedings of the 10th international conference on Ubiquitous computing*. ACM, 1–9.
39. Kristof Van Laerhoven, David Kilian, and Bernt Schiele. 2008. Using rhythm awareness in long-term activity recognition. In *Wearable Computers, 2008. ISWC 2008. 12th IEEE International Symposium on*. IEEE, 63–66.
40. Yi Yang, Heng Tao Shen, Zhigang Ma, Zi Huang, and Xiaofang Zhou. 2011. l2, 1-norm regularized discriminative feature selection for unsupervised learning. In *IJCAI Proceedings-International Joint Conference on Artificial Intelligence*, Vol. 22. Citeseer, 1589.
41. Yi Yang, Fei Wu, Feiping Nie, Heng Tao Shen, Yueting Zhuang, and Alexander G Hauptmann. 2012. Web and personal image annotation by mining label correlation with relaxed visual graph embedding. *Image Processing, IEEE Transactions on* 21, 3 (2012), 1339–1351.
42. Lina Yao, Quan Z Sheng, Xue Li, Sen Wang, Tao Gu, Wenjie Ruan, and Wan Zou. 2015a. Freedom: Online Activity Recognition via Dictionary-based Sparse Representation of RFID Sensing Data. In *Proceedings of the IEEE International Conference on Data Mining (ICDM)*. IEEE.
43. Lina Yao, Quan Z. Sheng, Wenjie Ruan, Tao Gu, Nickolas Falkner, Xue Li, and Zhi Yang. 2015b. RF-Care: Device-Free Posture Monitoring of Elderly People Using a Passive RFID Tag Array. In *Proc. of the 12th International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services (MobiQuitous 2015)*. Coimbra, Portugal.
44. Mi Zhang and Alexander A Sawchuk. 2012. Usc-had: a daily activity dataset for ubiquitous activity recognition using wearable sensors. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*. ACM, 1036–1043.
45. Dengyong Zhou, Olivier Bousquet, Thomas Navin Lal, Jason Weston, and Bernhard Schölkopf. 2004. Learning with local and global consistency. *Advances in neural information processing systems* 16, 16 (2004), 321–328.
46. Xiaojin Zhu, Zoubin Ghahramani, John Lafferty, and others. 2003. Semi-supervised learning using gaussian fields and harmonic functions. In *ICML*, Vol. 3. 912–919.