Paper Summary

This paper is a survey of the Human Activity Recognition (HAR) research done till 2014. There are some unique challenges to HAR compared to other pattern recognition problems like

1. Definition and Diversity of Physical Activities: Human activities are diverse and very limited usage of a uniform taxonomy.
2. Class Imbalance: Some activities are sparse (like sipping a drink whereas working, sleeping etc. is relatively common).
3. Ground Truth Annotation: Manual and tedious. May be in controlled lab-settings it can be done in a post-hoc manner.
4. Data Collection and Experiment Design : Not enough collaborative work to generative a rich pool of data since the task requirements can be very diverse for every research group.

Activity Recognition Chain (ARC) :

I. Sensor Data Acquisition and Preprocessing: Addresses the below challenges:
    a. Non-uniform sampling rate
    b. Unsynchronised multivariate inputs
    c. Corrupted sensor data

II. Segmentation
    a. Energy based : Every activity has its distinct energy signature.
    b. Windowing based (overlapping/non-overlapping)
    c. Additional context based segmentation : Using Video camera and so.

III. Feature Extraction and Selection
    a. Signal-based features : Statistical features or domain specific metrics like Mel-frequency cepstral coefficients for audio signals
    b. Positional estimates from 3-D reconstruction.
    c. Event based : Capturing specific limb, eye etc. Based features.
    d. Multilevel features : Sliding window based using encoded duration, frequency of occurrences etc.

    Higher dimensionality in feature space demands more training data. Many traditional techniques exist to do feature reduction/selection, although modern tools like AdaBoost and SVMs can automate this process implicitly in their functionality.

IV. Training and Classification
    a. Statistical methods : Hidden Markov Model, Support Vector Machine, naive Bayes, k-Nearest Neighbors
    b. Template based methods : Dynamic Time Warping, string matching
    c. Tree based Learners
    d. Decision Fusion : Ensemble Methods
        i. Bagging (N parallel learners and use a decision rule to combine their predictions)
        ii. Boosting (N sequential learners where the prediction of one learner influences the subsequent one).

V. Performance evaluation
    a. Confusion matrices
    b. Accuracy, precision, recall, and F-scores
    c. Decision-independent Precision-Recall (PR)- or Receiver Operating Characteristic (ROC)
    d. Time based evaluation
    e. Event based evaluation : Cross-validation techniques like leave-one-out

Some key finding in their controlled-setting experiment with simple gesture recognition tasks :

- Person dependent models outperform the generic models.
- Mean and Variance features yield the best accuracy.

- Choice of window size influence the performance of model.
- Accelerometer sensing modality outperforms gyroscope.
- SVM and kNN are the best performing classifiers for HAR tasks.

Things I liked about the paper

The paper presents a plausible sequence of steps to develop a predictor for HAR tasks particularly but can be broadly extended to other multi-sensor system tasks. It is a good guide to follow while studying biophysical and other human-centric data which share a lot of the challenges clearly addressed in this paper.

Areas that can use improvement

The paper did not evaluate the performance of any template based modelling techniques for HAR. It would have been interesting to see if the vanilla temporal characteristics play any role in activity recognition.